Development, theory and application of the reflection confocal scanning infra-red microscope

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

P. Török

Wolfson College
Oxford

Trinity Term, 1994

RADCLIFFE SCIENCE LIBRARY OXFORD
DEPOSITED
- 9 FEB 1995
THESIS

To my parents.

"...a good theory is characterized by the fact that it makes a number of predictions that could in principle be disproved or falsified by observation. Each time new experiments are observed to agree with the predicitions the theory survives, and our confidence in it is increased; but ever a new observation is found to disagree, we have to abandon or modify the theory. At least that is what is supposed to happen, but you can always question the competence of the person who carried out the observation."

S.W. Hawking: A Brief History of Time

Preface

This thesis is an account of the studies carried out by the author in the Department of Materials in the University of Oxford. The work described was performed solely by the author. Results of other researchers are duly acknowledged and a list of references is given at the end of the thesis. No part of this work has been submitted for any degree in this or any other universities.

I would like to thank Professors Sir P.B. Hirsh FRS, R.J. Brook and D.G. Pettifor FRS for the provision of laboratory and computing facilities. I am grateful to MEMC Electronic Materials SpA., Novara, Italy for financial support.

I am greatly indebted to my supervisors Dr G.R. Booker (Oxford) and Dr R. Falster (MEMC) for their support, encouragement and advice during the course of this work.

Thanks are due to my colleagues in Oxford and elsewhere for their help and friendship on many occasions. In particular I thank:

Dr P. Varga of Central Research Institute for Physics, Hungarian Academy of Sciences for introducing me to theoretical optics and teaching me with great wisdom and patience.

Dr A. Hámori of Central Research Institute for Physics, Hungarian Academy of Sciences for introducing me to experimental optics and friendly discussions when I visited Hungary.

Prof. E. Wolf of Department of Physics and Astronomy, University of Rochester for encouragement and discussions on some parts of the work and for his continuing interest.

Dr Z.J. Laczik of Department of Materials, University of Oxford for help and discussions.

Dr A. Konkol of Research Institute for Technical Physics, Hungarian Academy of Sciences and Department of Materials, University of Oxford for help and friendly, occasionally deeply philosophical, discussions.

Dr S. Dudarev of Department of Materials, University of Oxford for his help with mathematics.

Mr J.M. Stead, Mr J.M. Short, Mr D.W. Pinfold and Mr J.W. Bryant for their help and the construction of mechanical parts of the scanning infra-red microscope.

Mr R.M. Cripps for his help with electronics and friendly hours spent chatting.

Dr B. Pécz of Research Institute for Technical Physics, Hungarian Academy of Sciences for his friendship and for being the other pipe smoker at the Department on his occasional visits to Oxford and for obtaining TEM images on some of the specimens.

Dr Gy. Bencze of Central Research Institute for Physics, Hungarian Academy of Sciences and CERN for his friendship and for many personal and professional discussions.

Prof. J. Gyulai of Central Research Institute for Physics, Hungarian Academy of Sciences and Technical University of Budapest for his help and for his trust in me.

Dr J. Whitehurst, Dr M. Dellith, Dr S.A. Galloway, Dr T-Y. Seong and many other members of the Semiconductor Group and the Department for their help, friendship and discussions that led me to see the world better and differently.

All of my friends in Hungary who, despite our lives being separated by distance, always made me feel close to home.

My very special thanks are due to Éva. Her love, support, understanding and never-ending patience and tolerance helped me through the most difficult moments and cheered me up. Without her loving care this work would have been a far more difficult undertaking.

Parts of this work have been or are to be published as follows:

- P Török, Z Laczik and GR Booker

  The development and use of a new confocal SIRM incorporating reflection, double-pass and phase contrast modes

  Proc. of 8th Oxford Conference on Microscopy of Semiconducting Materials, April 1993, Oxford, UK; Inst. of Physics Conf. Series, No. 134, 1993, p771.

- Z Laczik, P Török, GR Booker and R Falster

  New applications of the scanning infra-red microscope (SIRM) to inhomogeneities in bulk GaAs and Si

  7th Oxford Conference on Microscopy of Semiconducting Materials, 25-28 March 1991, Oxford, UK; Inst. of Physics Conf. Series, No. 117, Sec. 10, 1991, p785.

- R Falster, Z Laczik, GR Booker and P Török

  Gettering of copper and nickel in Czochralski silicon by oxide particles: assessment of thermal stability

  4th International Autumn Meeting on Gettering and Defect Engineering in Semiconductor Technology, October 1991, Berlin, Germany; Solid State Phenomena, Vols. 19-20, 1991, p33.

- P Török, P Varga, Z Laczik and GR Booker

  Electromagnetic diffraction of light focussed through a planar interface between materials of mismatched refractive indices - An integral representation

  accepted for publication in J. Opt. Soc. Am. A

- P Török, P Varga and GR Booker

  Electromagnetic diffraction of light focussed through a planar interface between materials of mismatched refractive indices - The structure of the diffracted field

  to be submitted to J. Opt. Soc. Am. A

- P Török, Z Laczik and GR Booker

  Imaging of oxide particles within bulk Czochralski silicon wafers using a high- per-

formance reflection confocal scanning infra-red microscope

to be submitted to Scanning

- P Török, Z Laczik and GR Booker

  Imaging of oxide particles within bulk Czochralski silicon wafers using a high- performance reflection differential phase contrast confocal scanning infra-red microscope

  to be submitted to J. Microsc.

CONTENTS

1. Introduction **5**

2. Experimental techniques **9**
2.1 Etching and optical surface microscopy 9
2.2 Transmission electron microscopy 10
2.3 Infra-red microscopy 11
2.4 Infra-red laser scattering tomography 11
2.5 Optical precipitate profiler 12
2.6 Scanning infra red microscopy 13

3. Imaging modes in SIRM **16**
3.1 Introduction 16
3.2 Literature review 17
3.3 General design considerations 23
3.4 Surface reflection 26
3.5 Transmission non-confocal SIRM 29
3.6 Reflection confocal SIRM 31
3.7 Transmission confocal SIRM 33
3.8 Reflection confocal DPC SIRM 35
3.9 Transmission confocal DPC SIRM 37
3.10 Nomarski confocal DIC SIRM 38
3.11 Interference confocal SIRM 41
3.12 Other imaging modes 43

4. Focussing theories **44**
4.1 Introduction 44
4.2 High aperture focussing theories 44
4.2.1 The Kirchhoff theory 44
4.2.2 The Debye theory 47
4.2.3 Luneburg's interpretation of the Debye theory 48
4.2.4 Wolf's integral representation 51
4.3 Application of high aperture focussing in confocal microscopy 52

5. Extension of Wolf's diffraction theory **56**
5.1 Introduction 56
5.2 The integral representation 58
5.3 The electric and magnetic strength vectors 63
5.4 The electric and magnetic field vectors 66
5.5 Physical interpretation 69
5.6 Paraxial approximation 71

6. Analytical solution of the diffraction integrals **74**
6.1 Introduction 74
6.2 Evaluation of the diffraction integrals 77
6.3 Linearisation of the aberration function 83
6.4 Numerical results 85
6.5 Conclusions 94

7. Interpretation of the diffraction integrals **95**
7.1 Introduction 95
7.2 Circular polynomials of Zernike 96
7.3 Expansion of the aberration function 98
7.4 Numerical results 98

8. Light scattering by particles - Theory and application to the reflection SIRM **104**
8.1 Introduction 104
8.2 Formulation 106
8.3 Solution of the wave equation 111
8.4 Far field solution 113
8.5 Numerical results 116

9. Solution of the diffraction integrals in the presence of a scatterer in the focal region - Backward propagation **133**
9.1 Introduction 133
9.2 Formulation and solution 135

10. Experimental results and interpretation **141**
10.1 Introduction 141
10.2 System design 141
10.2.1 Mechanical design 142
10.2.2 Electronic design 144
10.2.3 Optical design 148
10.3 Reflection confocal measurements 149
10.3.1 Introduction 149
10.3.2 The specimens 152
10.3.3 Experimental results 153
10.3.4 Effect of the spherical aberration 162
10.3.5 Contrast 166
10.4 Other imaging modes in the SIRM 167
10.5 Transmission confocal SIRM 175
10.6 Interpretation of results 177

11. Summary, suggestions for further work and conclusions **180**

Appendix A. Special functions and orthogonal polynomials **A-1**
A.1 Bessel and associated functions A-1
A.1.1 The Bessel function A-1
A.1.2 The Neumann function A-1
A.1.3 The Hankel functions A-2
A.2 The spherical Bessel and associated functions A-2
A.2.1 The spherical Bessel function A-2
A.2.2 The spherical Neumann function A-3
A.2.3 The spherical Hankel functions A-3
A.2.4 The Riccati-Bessel functions A-4
A.3 The Legendre and associated functions A-5
A.3.1 The Legendre functions A-5
A.3.2 The associated Legendre functions A-6
A.3.3 The τ_n and π_n functions A-6
A.4 The Gegenbauer polynomials A-7
A.5 Orthogonal polynomials A-7

Appendix B. Proof of two solutions **B-1**
B.1 The solution of Bessel's differential equation B-1
B.2 The associated Legendre equation B-2

References **Ref - 1**

Preface

This thesis is an account of the studies carried out by the author in the Department of Materials in the University of Oxford. The work described was performed solely by the author. Results of other researchers are duly acknowledged and a list of references is given at the end of the thesis. No part of this work has been submitted for any degree in this or any other universities.

I would like to thank Professors Sir P.B. Hirsh FRS, R.J. Brook and D.G. Pettifor FRS for the provision of laboratory and computing facilities. I am grateful to MEMC Electronic Materials SpA., Novara, Italy for financial support.

I am greatly indebted to my supervisors Dr G.R. Booker (Oxford) and Dr R. Falster (MEMC) for their support, encouragement and advice during the course of this work.

Thanks are due to my colleagues in Oxford and elsewhere for their help and friendship on many occasions. In particular I thank:

Dr P. Varga of Central Research Institute for Physics, Hungarian Academy of Sciences for introducing me to theoretical optics and teaching me with great wisdom and patience.

Dr A. Hámori of Central Research Institute for Physics, Hungarian Academy of Sciences for introducing me to experimental optics and friendly discussions when I visited Hungary.

Prof. E. Wolf of Department of Physics and Astronomy, University of Rochester for encouragement and discussions on some parts of the work and for his continuing interest.

Dr Z.J. Laczik of Department of Materials, University of Oxford for help and discussions.

Dr A. Konkol of Research Institute for Technical Physics, Hungarian Academy of Sciences and Department of Materials, University of Oxford for help and friendly, occasionally deeply philosophical, discussions.

Dr S. Dudarev of Department of Materials, University of Oxford for his help with mathematics.

Mr J.M. Stead, Mr J.M. Short, Mr D.W. Pinfold and Mr J.W. Bryant for their help and the construction of mechanical parts of the scanning infra-red microscope.

Mr R.M. Cripps for his help with electronics and friendly hours spent chatting.

Dr B. Pécz of Research Institute for Technical Physics, Hungarian Academy of Sciences for his friendship and for being the other pipe smoker at the Department on his occasional visits to Oxford and for obtaining TEM images on some of the specimens.

Dr Gy. Bencze of Central Research Institute for Physics, Hungarian Academy of Sciences and CERN for his friendship and for many personal and professional discussions.

Dr J. Whitehurst, Dr M. Dellith, Dr S.A. Galloway, Dr T-Y. Seong and many other members of the Semiconductor Group and the Department for their help, friendship and discussions that led me to see the world better and differently.

All of my friends in Hungary who, despite our lives being separated by distance, always made me feel close to home.

My very special thanks are due to Éva. Her love, support, understanding and never-ending patience and tolerance helped me through the most difficult moments and cheered me up. Without her loving care this work would have been a far more difficult undertaking.

Parts of this work have been or are to be published as follows:

- P Török, Z Laczik and GR Booker

  The development and use of a new confocal SIRM incorporating reflection, double-pass and phase contrast modes

  Proc. of 8th Oxford Conference on Microscopy of Semiconducting Materials, April 1993, Oxford, UK; Inst. of Physics Conf. Series, No. 134, 1993, p771.

- Z Laczik, P Török, GR Booker and R Falster

  New applications of the scanning infra-red microscope (SIRM) to inhomogeneities in bulk GaAs and Si

  7th Oxford Conference on Microscopy of Semiconducting Materials, 25-28 March 1991, Oxford, UK; Inst. of Physics Conf. Series, No. 117, Sec. 10, 1991, p785.

- R Falster, Z Laczik, GR Booker and P Török

  Gettering of copper and nickel in Czochralski silicon by oxide particles: assessment of thermal stability

  4th International Autumn Meeting on Gettering and Defect Engineering in Semiconductor Technology, October 1991, Berlin, Germany; Solid State Phenomena, Vols. 19-20, 1991, p33.

- P Török, P Varga, Z Laczik and GR Booker

  Electromagnetic diffraction of light focussed through a planar interface between materials of mismatched refractive indices - An integral representation

  accepted for publication in J. Opt. Soc. Am. A

- P Török, P Varga and GR Booker

  Electromagnetic diffraction of light focussed through a planar interface between materials of mismatched refractive indices - The structure of the diffracted field

  to be submitted to J. Opt. Soc. Am. A

- P Török, Z Laczik and GR Booker

  Imaging of oxide particles within bulk Czochralski silicon wafers using a high- per-

formance reflection confocal scanning infra-red microscope

to be submitted to Scanning

- P Török, Z Laczik and GR Booker

  Imaging of oxide particles within bulk Czochralski silicon wafers using a high- performance reflection differential phase contrast confocal scanning infra-red microscope

  to be submitted to J. Microsc.

CHAPTER 1

Introduction

"Times are bad. Children no longer obey their parents, and everyone is writing a book."

Cicero

Czochralski (Cz) silicon wafers are used almost exclusively for the fabrication of VLSI devices. Such silicon contains excess oxygen which precipitates as oxide particles either when the initial ingot is grown or subsequently during the wafer device fabrication. Such oxide particles can produce reduced device performance or failure if they occur within the active device regions. However, they can be used to improve the device performance by a process known as internal oxide gettering. The wafers are given a series of preanneal treatments to produce controlled precipitation in which a surface zone of the wafer to a depth of typically in the range of 10 to 50 μm is denuded of oxide particles, while the remainder of the wafer contains large numbers of well formed particles. The devices are fabricated in the surface denuded zone and harmful contaminating metal impurities are attracted during the heat treatment stages away from the device regions to precipitate at the underlying oxide particles or their associated dislocations. In this way device yields can be significantly increased.

Because of the importance of these oxide particles and the oxygen precipitation process for VLSI fabrication, considerable efforts have been made to develop methods to assess the numbers and distributions of such particles within the wafers. The number density range of most interest is 10^7 to 10^{10} cm^{-3}, and the particle size range is typically 30 to 300 nm.

The method that has mostly been used is surface etching followed by optical microscopy to obtain etch pit densities. Transmission electron microscopy is a research method used

for obtaining detailed information concerning a small number of individual particles. However, because these methods are destructive, much attention has been given during the last few years to the development of infra-red microscopy methods to directly image the particles within the silicon wafers. Although the particles are smaller than the resolution of these methods, individual particles can nevertheless be imaged. This is because the particles are mostly further apart than the resolution limit, and the sensitivity can be sufficient high that adequate contrast occurs. The contrast arises from scattering or absorption of the light by the particle.

Infra-red imaging methods developed include infra-red microscopy (IRM), laser scattering tomography (LST), optical precipitate profiler (OPP) and scanning infra-red microscopy (SIRM), all described more fully in Chapter 2. The SIRM has been developed and used to investigate a variety of semiconductor specimens in the Materials Department, Oxford University, during the last ten years. The SIRM has a good performance and flexibility making it especially suitable as a research instrument.

Although all of these infra-red imaging methods have been successful to different degrees in assessing oxide particles in Cz silicon wafers, their performance has at least initially been assessed by comparing the number densities and distributions thus obtained with the corresponding results produced by etch pit studies. Furthermore, no serious attempt has yet been made to develop a rigorous theory of the imaging process and to compare the predictions with the experimental images. One of the main objectives of the present work is to do this or at least to make a significant start to such a project based on the SIRM.

The outline of an ideal project which aims at a full understanding of the imaging process and the contrast mechanisms is as follows. The performance of the present Oxford SIRM should be improved and the number of imaging modes increased. The improved performance, i.e. better lateral and depth resolutions and higher sensitivity, would enable smaller particles and higher number densities to be imaged, and hence better quantitative data obtained. The larger number of imaging modes would enable the optimum method to be used to image different types of particle.

1. Introduction

A rigorous theory should be developed that can describe the imaging process and the contrast mechanisms. First, the illumination system should be studied, and in particular the structure of the focussed probe within the specimen and how the structure changes on focussing deeper into the specimen. Second, the interaction of the light with the specimen should be investigated and especially how light is scattered by individual oxide particles in silicon for the case of the particle size being smaller than the light wavelength. Third, the detection system should be considered. For example, for the reflection confocal SIRM, how the light back scattered by the particles is collected by the probe forming lens and imaged at a pin-hole aperture placed in the front of the detector. Well designed experiments are required to determine the imaging properties of the different modes and comparisons should be made between the experimental and theoretical data.

The successful conclusion of such a project would enable SIRM images of the particles to be more fully interpreted and hence more detailed information obtained concerning the particles. Furthermore, the images expected from different types of particle could be more closely predicted, e.g. whether they are detectable or not, and hence materials projects could be better planned at the outset.

In this thesis we describe the methods that are presently being used to assess oxide particles in bulk silicon (Chapter 2). We review the literature on scanning optical microscopy covering both visible and infra-red light, present some considerations regarding the design of a high performance and versatile SIRM, and describe the various microscope modes that have been or could be used to image particles in semiconductors with infra-red light (Chapter 3).

We give a detailed rigorous theoretical analysis of the energy distributions in the probe for the case when the light is focussed by a high numerical aperture lens from air into silicon (Chapters 4, 5). Numerically computed distributions are obtained to illustrate how the probe changes under different conditions, e.g. different focussing depths (Chapter 6). The relationship between the penetration depth of the probe and the spherical aberration coefficient arising from the silicon specimen is determined (Chapter 7).

The classical theory of light scattering is applied to individual spherical silicon dioxide

particles embedded in silicon. Numerical results are presented and a contrast mechanism is proposed to describe how the scattered intensity depends on particle size (Chapter 8). A formal solution relevant to the reflection confocal SIRM is given to treat the backward propagation of light using a model which takes into account the polarisation state of the incident light, the spherical aberration introduced by the silicon wafer, the polarisation state of the scattered light and the size of the pin-hole (Chapter 9). Experimental results are obtained for most of the imaging modes described in Chapter 3, specimens being selected so that the wide range of the imaging capabilities of the SIRM is shown, and experimental contrast values are compared with theoretical values (Chapter 10). Finally, overall conclusions are drawn and suggestions are made for completing the work started here (Chapter 11).

CHAPTER 2

Experimental techniques

"The fundamental principle of science, the definition almost, is this: the sole test of validity of any idea is experiment."
R.P. Feynman

The aim of this chapter is to review the experimental techniques that are presently being used to investigate precipitate particles in bulk silicon. The most important methods are summarised giving also the type of information that can be obtained.

The ideal method should enable the particle number densities and distributions, and the depth of the surface denuded zone (SDZ), to be determined. For commercial silicon wafers as supplied to the electronics industry for VLSI processing, usually the 'front' surface on which the devices are to be fabricated is well- polished while the 'back' surface is less-well polished. For a particle assessment method to be acceptable to the electronics industry for routine wafer inspection, the method should be non-destructive. This means that the wafer should not have to be broken, the 'front' surface modified in any way or the 'back' surface given additional polishing to make it well-polished. If the silicon wafers are to be used for research assessment purposes, such non-destructive considerations generally do not apply.

2.1 Etching and optical surface microscopy

Defect etching is used to study precipitate particle number densities and distributions. This can either be done by producing plan-view or cross-sectional specimens, though etching is mostly used on cleaved cross-sections. The etching can be done by agitating the specimen in the etchant (commonly used Secco etchant) for typically 1 or 2 minutes. The etching rate should be calibrated prior to microscopy examination for the particular

crystallographic orientation used.

The etched specimen is examined under conventional reflection mode optical microscopy where, to obtain higher contrast on the each-pits, dark field or Nomarski microscopy may be used. The etch-pits are counted over a certain area (field of view) and this number is divided by the thickness of the layer removed by etching, typically 1 or 2 μm, to obtain the volume number density. Although this method is of a subjective nature, it can be used with good reliability in the particle number density range of $\sim 10^7$ to $\sim 10^9 \mathrm{cm}^{-3}$. The main disadvantages of this method are that the etchants are highly toxic, it is destructive and the particle number density range which can be examined is limited. This method is easily available and is frequently used.

2.2 Transmission electron microscopy

The transmission electron microscopy (TEM) is a method, unlike etching where macroscopic type information is available, which is capable of providing information on the fine structure of precipitate particles. For the TEM examination the specimen must be prepared carefully. For plan-view observation a disc is cut from the specimen which is then ground and polished from both sides down to a thickness of ~ 25 μm, then it is placed in an ion beam mill where the centre of the specimen is thinned until perforated. Alternatively, the thinning can be made by chemical polishing on plan-view specimens. The thickness of the specimen increases from the centre of the disk which is caused by the ion guns being directed at a small angle with respect to the surface of the specimen. The disadvantages of the method are that it requires a long and a laborious specimen preparation and it is not appropriate when low particle number densities are present since the TEM foil sample is very thin. The TEM is a research tool that is rarely used by manufacturers. The microscope itself is available commercially although it is very expensive.

2.3 Infra-red microscopy

The infra-red microscopy (IRM) method is generally taken to mean the transmission non-scanning method as developed and used by, for example, Brozel and Skolnick and their colleagues[18,89,90]. It utilises the fact that most of the semiconductor materials are transparent for the near infra-red region. This method usually requires no specimen preparation, except that both the front and back surfaces of the specimen must be polished to obtain good image quality. The contrast occurs either by absorption[18] or by scattering[64,20]. The simplest method[18] utilised a light bulb with an infra-red filter which illuminated the whole area of, say, a GaAs wafer. A PbS, PbO or silicon vidicon camera was placed just after the specimen so converting the transmitted infra-red light to visible light. The magnification used was typically 2 to 5× and the resolution was poor. However, at the time the method was developed, the main interest was to image the macroscopic structure of the cell pattern in LEC GaAs wafers[89,90] where the contrast arises from the absorbing EL2 levels at the cell walls, hence the optical set-up described above was satisfactory. Later the method was used to determine EL2 concentrations at the cell walls[18].

2.4 Infra-red laser scattering tomography

Infra-red laser scattering tomography (LST)[65,71,28,97,29] is a method which can assess semiconductor specimens for precipitate particle number density and particle distribution with good sensitivity. The technique relies on a 90° scattering dark field microscope. A portion of the specimen is usually cleaved and the cleaved surface is polished at 90° with respect to the front surface. A low numerical aperture lens is used to form an elongated probe inside the specimen which illumination usually occurs from the cleaved and polished surface and runs parallel with the front surface. A second, high numerical aperture lens is placed at the front surface of the specimen so that the optical axis of the second lens is at 90° with respect to that of the illuminating lens. The high numerical aperture lens images the illuminated volume of the specimen either to a line CCD or a vidicon camera. A two-dimensional image is obtained by slowly moving the specimen or the illumination

perpendicular to the plane determined by the optical axes of the two lenses. A digital image is built up usually in the memory of a computer line-by-line.

This method can provide information on particles situated at different depths in the specimen. However, shallow surface denuded zones (SDZ) cannot be measured at large distances from the cleaved surface, since the low numerical aperture lens has a long working distance and the beam waist diameter is large. If the waist is, say, 40 μm and the specimen is aligned so that the illumination is running parallel and close to the front surface, then the depth from which information is obtained is 0 to 40 μm, which corresponds to a poor depth resolution. If a higher numerical aperture lens is used to form a smaller waist, the working distance of the lens does not usually enable the illumination of material at large distances from the cleaved surface. Consequently, if the better depth resolution is required, only a small region of the specimen can be examined or the specimen has to be cut into several pieces. The method can give particle number densities and distributions and SDZ depths, but is destructive. The LST has been available on the market in the last eight years. The equipment is inflexible and very expensive.

Recently a new method has been developed which is similar to the LST[98] and which is called Brewster angle LST. In this method a fibre optics is used to illuminate the front polished surface of the specimen at the Brewster angle and the back-scattered light is received by a high numerical aperture (NA) lens. The main advantages of the method are that it is non-destructive and use of the Brewster angle enables much of the light reflected from the polished surface to be eliminated from the image when it is used to investigate near surface regions. The illumination system is of poor quality and requires a powerful laser. Little information is available about its performance.

2.5 Optical precipitate profiler

The optical precipitate profiler (OPP) was developed as a commercial equipment and the available information is limited[70]. The method utilises a transmission, common path, differential interference contrast interferometer. The light of an infra red laser is passed through a Wollaston prism and is focussed by a microscope objective lens to give two

closely separated probes inside the semiconductor specimen. The two probes pass through the specimen and are recombined and analysed. Two detectors are applied, one detecting the amplitude signal, the other detecting the phase signal. A major advantage of this method is that by using the amplitude signal rather than the more usual intensity signal, a better detection dynamics throughout the whole particle size range occurs. The specimen is raster scanned and the image is stored in the memory of a computer. The images are mostly obtained by cross-section scans and are visually difficult to interpret. An automated image processing algorithm calculates the particle number density from the images. The equipment is especially sensitive to phase changes occurring in the focal region of the probes, therefore is able to detect small inhomogeneities. It can provide information on precipitate particle number density and distribution and depth of SDZ. It requires, however, the specimen to be polished on both surfaces and is thus destructive. As it follows from the construction, it does not possess the advantages offered by the confocal detection and the equipment is inflexible. The OPP is commercially available and is expensive.

2.6 Scanning infra red microscopy

The scanning infra red microscopy (SIRM) is a versatile method to provide information on precipitate particle number density and distribution and SDZ depths. The method evolved from the combination of the principles of the infra-red microscope (IRM) and the visible laser scanning microscope (LSM)[112]. The main development and application of the transmission SIRM to semiconductors were made in the Materials Department, Oxford University, by Kidd working with LEC GaAs specimens[52–54] and Laczik working with Si, GaAs and CdTe specimens[55–57]. The principle of the transmission SIRM is shown in Fig. 2.1.

The light sources used were initially a He-Ne laser operating at 1.15 μm, 1 mW and a semiconductor laser operating at 1.3 μm, 3 mW (Kidd), and later a semiconductor laser operating at 1.3 μm, 30 mW (Laczik). In the SIRM, in general, a localised volume of the specimen is illuminated by imaging and demagnifying the laser source. This volume

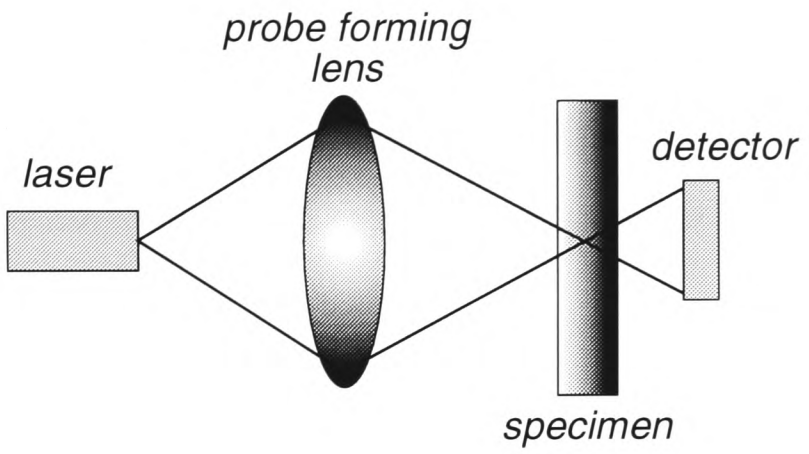


Figure 2.1: The transmission SIRM.

is referred to as the probe. Local inhomogeneities present within the illuminated volume interact with the light and scattering/absorption occurs. Scattered light can directly be detected by an off-axis detector whilst absorbed light can be detected against the total intensity incident on the specimen by an on-axis detector. The signal of the detectors is amplified and digitised. The specimen is mechanically raster scanned and the image is recorded as the detector signal corresponding to each raster point. The magnification of the microscope can be altered, in the typical range of 30 to 5000×, by changing the length of the scan. The specimen can be scanned either in the X-Y plane (corresponding to a plan-view image) or in the X-Z plane (corresponding to a cross-section image), where X, Y and Z indicate directions perpendicular and parallel to the optical axis respectively.

The specimen is scanned and changes in the light intensity are detected by a Ge diode. Contrast arises from scattering and/or absorption inside the specimen. The lateral and axial resolutions are $\sim$ 2 and $\sim$ 30 μm respectively (NA=0.6) and individual particles with diameters down to $\sim$ 50 nm can be imaged. With the SIRM quantitative information can be obtained on SDZ and particle number density and distribution.

It is emphasised that the SIRM can also be operated in the confocal mode, when the axial and lateral resolutions should be significantly improved (a main topic of the present thesis). The possible imaging modes of the confocal SIRM are discussed in Chapter 3. From the practical point of view the SIRM, either in confocal or non-confocal mode, is a conventional scanning microscope and therefore the different set-ups hitherto proposed or

used for the visible light scanning microscope (LSM) could, in principle, also be applied to the SIRM. A main difference between the SIRM and the LSM is the contrast mechanism and therefore when adopting a particular LSM set-up for the SIRM, one must consider how this affects the imaging capabilities of the SIRM.

The SIRM is commercially not available, although it could be marketed at a modest price.

CHAPTER 3

Imaging modes in SIRM

"The real voyage of discovery consists not in seeking new landscapes but in having new eyes."

M. Proust

3.1 Introduction

This thesis is concerned with the SIRM and especially the development and understanding of the various reflection confocal modes of the SIRM in order to image oxide particles in silicon. The optical principles of the reflection confocal scanning microscopes using infra-red light are similar to those using visible light. The main differences are that for infra-red light the wavelength is 2 or 3× longer and so the resolution is poorer, and when applied to semiconductor specimens the absorption is so much less that imaging can be performed deep into the specimen and particles can be observed with the contrast arising from light scattering. The combination of a high numerical aperture to obtain the best resolution and deep focussing into the specimen of high refractive index means that the aberrations are extremely high, much higher than have been previously considered for visible light applications. Little work has been performed on the reflection confocal SIRM prior to that of the present thesis, but considerable work has been performed over many years, both experimental and theoretical, on reflection confocal scanning microscopy using visible light.

In this chapter we initially review the extensive literature on scanning optical microscopy using visible light, and then describe the small amount of work that has been done using infra-red light. We continue with a general account of the SIRMs and the considerations that are important when designing a high performance instrument. We

conclude with a description of the different types of transmission and reflection SIRMs that either have been used or could be used for materials studies.

3.2 Literature review

The principle of confocal microscopy was proposed by Lukosz[59] in 1966 who recognised the fact that the lateral resolution of an imaging system can be improved if one is willing to sacrifice the field of view. In his paper he suggested the use of a small aperture placed very close to an evenly illuminated specimen and by scanning the aperture the confocal image is built up pixel by pixel. The principle is exactly the same today, except for convenience the small aperture is *imaged* onto or into the specimen and then this localised illumination is imaged back to a 'pin-hole' aperture.

Sheppard and Choudhury[79] in 1977 described a paraxial theory for confocal microscopy using Fourier optics. They introduced the notation of Type 1 and Type 2 microscopes, where Type 1 indicates scanning microscopes with imaging properties similar to those of conventional microscopes, whilst Type 2 indicates scanning microscopes with coherent imaging properties. They also established the use of the frequency transfer function $C(m, n; p, q)$ as a measure of the frequency transmitting capability of a scanning microscope. They showed that a Type 2 microscope could transmit spatial frequencies twice as high as a Type 1 microscope.

Brakenhoff *et* al.[15] in 1979 described an experimental transmission confocal scanning microscope which used high aperture immersion lenses to study biological specimens. Measured lateral point response curves for the main peak agreed well with those calculated from paraxial theory, even for numerical apertures (NA) of 1.3 (paraxial theory is generally considered not to be accurate once NA>0.5). Brakenhoff[16] in 1980 introduced a combined reflection/transmission confocal microscope and proposed that a transmission confocal interference microscope based on the Mach-Zender interferometer technique could be used to measure optical anisotropy in specimens. Brakenhoff[16] pointed out that several other imaging modes were also available for confocal microscopy such as the stereoscopic, differential imaging, transmission differential interference contrast (DIC), fluorescent and

multiple wavelength modes. These were pioneering ideas in 1980.

Material science applications of scanning optical microscopes appeared relatively early. Wilson *et* al.[106] in 1979 obtained images from dislocations in silicon by using the laser beam to inject electrical carriers which were subsequently collected at a p-n junction. This was termed the optical beam induced current (OBIC) method by analogy with the electron beam induced current (EBIC) method.

Sheppard and Wilson showed[80] in 1979 how primary (Seidel) aberrations affected the imaging performances of Type 1 and Type 2 scanning microscopes. The spherical aberration worsened the frequency transmittance function but this could be partly compensated by introducing a lens defocus.

Brakenhoff *et* al.[17] in 1980 considered the phase error on a geometrical optical basis as light was focussed by high aperture lenses through different refractive index materials. Sheppard[81] in 1980 gave a comprehensive overview of the paraxial theory of scanning microscopes. He introduced the Wolf integrals to the theory of confocal microscopy and suggested that the angular spectrum representation of plane waves could be used to calculate the behaviour of high aperture optical systems. Sheppard and Wilson[82] developed the theory of the direct-view confocal microscope originally proposed by Egger and Petrañ[27].

Sommagren[91] described a reflection heterodyne interference microscope used to measure relative specimen heights with a sensitivity of 1 Å. Hamilton and Sheppard[35] proposed a possible reflection confocal interferometer scanning microscope based on a Michelson interferometer, the reference beam being obtained by splitting the beam and reflecting it back from a mirror. Two detectors were placed behind two pin-holes so that the sum of the two signals gave the conventional confocal image, whilst the difference gave the interference image. They presented several images and described a simple theory for the formation of the interference fringes. Hamilton and Wilson[36] discussed the possibility of applying the reflection confocal microscope to measure surface profiles, the method being based on the depth discrimination property of the confocal microscope. The specimen was scanned along the optical axis and one of the lateral axes and the signal of the detector was used to reconstruct the surface profile.

Based on the idea of Dekkers and de Lang[26], Hamilton and Wilson[39] constructed a transmission differential phase contrast (DPC) scanning microscope where a large area split detector was used to detect the transmitted light. Integration of the ratio of the sum and the difference of the signals of the two detectors produced a phase image, whilst the difference signal of the two detectors produced the differential phase contrast image. A similar optical set-up was used by Hamilton and Wilson[38] to achieve edge enhancement by differential detection. The inherent property of DPC imaging was used, namely, that the DPC mode differentiates phase objects and leaves unmodified amplitude objects. If the gradient of the DPC image is added to the conventional image then edge enhancement occurs.

Wilson and Sheppard[112] published in 1984 the first comprehensive book on confocal microscopy describing the theory and giving experimental examples and this is now the standard reference book on the subject. Wijnaendts van Restandt *et* al.[105] in 1984 described the possibility of the three dimensional reconstruction of images obtained by confocal microscopy and successfully demonstrated this for a biological specimen. Wilson and Hamilton[110] considered the imaging properties of a dark field scanning optical microscope and showed that it gave worse depth discrimination than the corresponding bright field microscope. Benschop[7] proposed a set-up for a differential phase contrast (DPC) microscope combined with confocal detection. He gave a theoretical treatment and derived the frequency transmittance function. He pointed out that the real part of the transfer function gave the modulation of the detected signal due to a phase object and the imaginary part gave the modulation due to an amplitude object.

Carlini and Wilson[19] described the effect of the detector size and its position on the imaging properties of a confocal scanning microscope. Up to a normalised detector radius of $v_D \sim 4$, the half width of the measured signal was strongly influenced by the size of the detector, whilst beyond this detector radius the measured half width of the resolution of a confocal microscope did not change appreciably. They also studied the effect of pin-hole misalignment on the image and found that for extended focus images, the misalignment of an ideal pin-hole was indicated by increased infringement at the edges of phase objects.

Theoretical predictions for the lateral resolution dependence on the pin-hole size were later experimentally confirmed for a reflection confocal microscope by Glass and Dabbs[31], although the theory gave smaller values. Wilson and Carlini[111] showed that by laterally offsetting the pin-hole, an imaging mode similar to dark field detection could be produced, and that the lateral pin-hole displacement increased the infringement in the image of either a phase or an amplitude edge.

Sheppard and Matthews[83] treated imaging by a high aperture lens. The scalar high aperture theory yielded the same formula for an axial scan as the paraxial theory with an apodised pupil. No attempt was made to use the full electromagnetic theory. In a subsequent article Sheppard[84] treated the problem of high aperture focussing in the presence of aberrations. He used the scalar form of the Wolf integrals to describe the field in the focal region and expanded the aberration function in terms of spherical harmonics (associated Legendre functions). He showed the correspondence between his expansion of the aberration function and the Seidel aberration coefficients and presented numerical results for different degrees of aberration. Hellmuth *et* al.[44] presented experimental results showing the effect on spherical aberration of high numerical aperture confocal systems, the aberrations being introduced by focussing through a glass plate. The high aperture system was extremely sensitive to both over- and under-correction for spherical aberration.

Keller[51] reviewed the objective lenses available for confocal microscopy. He considered experimentally the effect of different aberrations and presented images similar to those previously shown by Born and Wolf[14]. van der Voort and Brakenhoff[102] presented the first high aperture full electromagnetic treatment for a non-aberrated optical system. They used formulae given by Jackson[46] to derive the equation for the electric field vectors which were obtained previously by Richards and Wolf[75]. From these formulae they analysed a fluorescence system by considering the backward propagation as an incoherent radiation and by multiplying the two point spread functions. A problem here is that the validity of the point spread function approach is strictly limited to paraxial optics. Furthermore, the treatment was not rigorous in the mathematical sense. Visser and Wiersma[100] also made a full electromagnetic treatment of high numerical aperture confocal microscopes, but the

results were incorrect. The authors applied Kirchhoff's integral to an open surface whilst the original application of the integral was valid only for a closed surface. Furthermore, the application of Kirchhoff's integral requires the continuity of Green's function, which is not the case for this problem. Sheppard and Gu[86] showed that the spherical aberration introduced by a glass slab can be compensated in a confocal microscope by altering the tube length of the microscope objective lens. They used a geometrical optics approach to obtain the phase error introduced by the focussing and included this result in a scalar integral.

Seger and Lenz[77] and Seger[78] used an operator model (similar to that applied in quantum mechanics) to describe high aperture confocal microscope systems. The mathematics of the operator model was far from trivial, but once it was adopted it simplified the computation of confocal systems. Sheppard *et* al.[87] described the effect of the signal-to-noise ratio on confocal imaging including signal rejection due to the pin-hole, detector performance and three dimensional imaging methods. Cogswell and Sheppard[23] made a comprehensive treatment of the confocal differential interference contrast (DIC) system. They considered the effect of bias retardation on signal strength, linearity of response and signal-to-noise ratio. Transfer functions of the optical system were given for different signal detections (weak object and phase gradient) with various shears. Atkinson and Dixon[4] suggested a new optical set-up to achieve DPC imaging. They replaced the detector in the reflection confocal set-up by a lens so that it formed a beam expander with the detector lens in their infinity corrected optical system. A split detector was placed behind the new lens and DPC imaging occurred by subtracting the signal of the two detectors. They analysed their system by the paraxial theory and showed that it can only form images when the detector is not infinitely small.

The extensive use and development of the theory of confocal microscopy triggered the publication of a new book edited by Wilson[113]. The book was contributed from a group of experts from different fields of confocal microscopy and therefore gave an excellent collection of confocal microscopy techniques.

With regard to the theory of confocal scanning microscopy, the paraxial scalar theory

has been fully developed for the transmission and reflection confocal microscopes and significantly developed for the transmission and reflection confocal DPC microscopes. Some work has also been done for the reflection confocal DIC microscope. With regard to the high numerical aperture scalar theory, some work has been done for the reflection confocal microscope. With regard to the high numerical aperture vectorial theory, some work has been done for the reflection confocal microscope, embracing the cases of no aberrations, aberrations introduced by the imaging lens and aberrations introduced by focussing into a specimen. The advantage of the vectorial theory over the scalar theory is that it is more rigorous and accurate, and it must be used if a theoretical treatment with high numerical apertures is to be made of polarisation behaviours, light scattering by particles, fluorescent excitation, etc.

For confocal microscopes using light of wavelength 632.8 nm (He-Ne laser), the lateral and depth resolutions for NA=0.3 are typically 0.81 μm and 8.74 μm respectively, and the corresponding values for NA=0.9 are typically 0.27 μm and 0.71 μm respectively. (These numbers are given by the paraxial theory considering a point object and a point detector.)

All of the above experimental work has been for confocal scanning microscopy using visible light and the corresponding theoretical work was mainly intended for visible light. We now turn our attention to similar microscopy using infra-red light. Wilson *et* al.[107] and Hamilton and Wilson[41] used a scanning optical microscope in the OBIC mode to examine a GaP light emitting diode. When a laser of wavelength 1.15 μm was used to inject the electrical carriers, sub-surface defects were imaged because of the large penetration depth of the infra-red light. Hamilton and Wilson[40] used a reflection confocal scanning optical microscope with infra-red light to examine a silicon micro-circuit. By focussing from the front surface down through the silicon wafer they imaged the metal bonding at the back surface.

Considerable work has been performed at Oxford using the transmission non-confocal SIRM to investigate semiconductor specimens. Kidd *et* al.[52,53] examined As-rich precipitate particles in LEC GaAs wafers, Laczik *et* al.[55,56] examined oxide particles in Cz silicon wafers and Te-rich particles in CdTe wafers, and Jin *et* al.[47] examined In-rich particles

in InP wafers. Some of this work was recently reviewed[13]. For light of wavelength 1.3 μm and a lens of NA=0.6, lateral and depth resolutions of typically 2 μm and 30 μm were obtained. Particle number densities and distributions were determined. Individual dislocations were observed either because they were decorated with precipitate particles, or on using polarised light because of their strain fields.

Apart from the above SIRM work on semiconductor specimens, no other SIRM experimental work seems to have been reported, and no theoretical treatments of confocal microscopy specifically for infra-red light and semiconductor applications seem to have been reported, apart from the work of the present thesis. As the present work is concerned with infra-red light scattering by particles, the vectorial theory is required.

3.3 General design considerations

The design of the SIRM optics should satisfy two major requirements. First, the microscope should be easily alignable and such that the specimen can be changed without the need for realignment. Second, the microscope should be non-contacting and non-contaminating and such that successive measurements can be performed fast and effectively. Hitherto these two requirements have rarely been achieved on the same instrument. It is therefore the aim of the design to meet both of these requirements.

The main difference between a visible light scanning microscope and the SIRM follows from the material of the specimen into which the focussing occurs. For the SIRM the wavelength of the light needs to be sufficiently large that the semiconductor is non-absorbing, but sufficiently small that the resolution is not significantly degraded and that suitable optical components such as lenses are available with good performance. For silicon a wavelength of $\sim$ 1.3 μm is optimum. It is also an advantageous property of monocrystalline silicon that it is isotropic, i.e. the propagation properties of the light do not depend on its direction in the crystal.

As it is well known from the literature, the imaging properties of a confocal scanning microscope depend significantly upon the probe-forming and collector lenses. Thus the next step in the design is to consider what lenses should be used. It will be shown

later in this work that on considering a well aligned on-axis confocal microscope utilising a monochromatic light source and high quality optical components, the only aberration that exists in the system is spherical aberration. Different amounts of spherical aberration are introduced by different focussing depths into the specimen, which depth can typically be from some microns up to 500 μm. Another problem is that in the case of a confocal microscope both of the lenses equally contribute to the image formation, therefore in both the transmission and reflection modes the spherical aberration affects the point spread function of both lenses. Thus, it is required that the system should not have large spherical aberration or serious loss of resolution could occur.

It is shown in Chapter 8 that for oxide particles in silicon, the backward scattered light intensity can be more than two orders of magnitude less than the forward scattered light intensity. It is essential to use a high numerical aperture lens to collect as much light as possible. This effect becomes more clear with the consideration of the law of refraction, i.e. the ray pencil at the surface of the specimen (possessing relatively high refractive index) refracts so that the effective numerical aperture is substantially decreased. The use of a high numerical aperture lens, on the other hand, ensures that both the lateral and the axial resolutions are kept at the optimum level. There are, however, two disadvantages of the use of high numerical apertures. With increasing numerical aperture, the free working distance of the lens decreases and the spherical aberration from the specimen for a given penetration depth increases.

The free working distance of the objective lens is important in the design. In the reflection confocal mode it should be long enough to be able to examine the whole depth of the given specimen (wafer thicknesses are in the range of 500 to 750 μm). In the confocal transmission mode the requirement is more stringent: the working distances of the two lenses should be long enough to put a specimen of standard thickness between the lenses.

Objective lenses have long been designed with relatively high numerical aperture and long working distances, such as heating stage objectives where the lens should be kept far from the hot specimen, but the numerical aperture of these lenses did not reach the

value of 0.7. The highest numerical aperture lenses available on the market possess a numerical aperture of 1.4 but this value is calculated for oil immersion and for a glass substrate. As the SIRM must be neither a contaminating nor a contacting method, the highest numerical aperture that can be used is in the region of $0.8 - 0.95$. Unfortunately the free working distances of lenses possessing higher numerical apertures than 0.85 are very short (usually up to some 100 μm) which would not be long enough to measure the full depth of a standard semiconductor wafer. Recently new microscope objective lenses appeared on the market which possess numerical apertures as high as 0.85 with working distances of up to a millimetre and, additionally, they have been designed for infra-red light. This latter is important since usually surfaces of microscope objective lenses are coated with an antireflection layer that is effective in the visible range but is not effective in the infra-red. In fact, it can be so highly reflective for infra-red light that it reflects almost all of the incident light so preventing the use of such lenses in the SIRM.

Having discussed the microscope objective lenses, the problem of spherical aberration is still not solved. There are several papers which dealt with the problem of spherical aberration correction in confocal microscopy (see Section 3.2). In particular, Sheppard showed[88] that by inserting weak negative lenses before the detector lens, one can correct spherical aberration introduced by focussing into the specimen. The physical background of this solution is to introduce artificially a wave front distortion (through a radial phase modification) that compensates the spherical aberration. In our case when, in principle, an infinite number of penetration depths are required, this solution is unattractive as it would require a continuously adjustable correction lens.

There are two possibilities to overcome the problem of spherical aberration. If the resolution conditions of the microscope are not stringent or the detectable particle number density is not high, then one could calculate or measure the three-dimensional optical transfer function of the microscope to perform a deconvolution on the images obtained. We recognise that the best result is given by obtaining the optical transfer function of the actual system for all possible penetration depths. As this method is not practical an alternative way is to calculate the three-dimensional optical transfer function of the

system leaving the penetration depth in the solution as parameter. After obtaining an image, this transfer function could be deconvolved from the raw images yielding an image with improved resolution. If, on the other hand, the required resolution of the system or the detectable particle number density are high then the inevitable solution is to purchase a special infra-red lens possessing a continuously adjustable correction collar for spherical aberration.

There are several other aspects of the design of a SIRM and an account on these problems is given in Chapter 10.

3.4 Surface reflection

When the SIRM is operating in the reflection mode an important concern is caused by the specular reflection of light from the polished front surface of the specimen. This is an optical quality surface and its smoothness depends on the manufacturing processes. When the specimen is aligned so that the lens focusses the light at the surface, specularly reflected light reaches the detector through the pin-hole. When defocussing occurs the depth sectioning property of the confocal microscope causes a fall in intensity of the detected signal which can be described by $I \sim |\sin(u/2)/(u/2)|^{2*}$, where u is the normalised optical co-ordinate in the z direction (defined later). The intensity of light scattered from a particle which is situated close to the front surface, however, can be several magnitudes lower than that of the specular reflection. Thus, even providing the best possible optical sectioning and assuming a confocal microscope with a lossless point-like pin-hole in front of the detector, the minimum depth at which particles are visible for a three orders of magnitude intensity difference between the scattered and reflected light (the scattered light intensity being weaker) and a numerical aperture of 0.85 is calculated to be $\sim$ 73 μm (in Si). If the particles to be measured are situated deep in the bulk then specular reflection from the front surface does not play a significant role. In the semiconductor industry, however, the upper 0 to 50 μm of the specimen is a region (SDZ) of particular

*This is an approximate formula which essentially involves the paraxial approximation and assumes a perfect plane reflector and a point-like detector.

interest since most of the electronic devices are constructed in the top few microns of this depth. Furthermore the merit of oxide gettering is to keep away harmful impurities from the crystal in the SDZ. These arguments indicate an important direction of the reflection confocal SIRM design: the aim is to eliminate the specular reflection but maintain the detection of scattered light from the particles in such a way that neither resolution nor light intensity loss should occur.

These requirements imply the first solution of the problem, namely, the application of different dark field methods, which were the subject of an Oxford Patent Application[12] and were applied briefly to material science problems and oxide precipitates[57]. However, dark field microscope modes usually cause loss in resolution and/or detected light intensity. In one of the methods suggested in these two references, the illumination occurs so that the laser illuminates only a fraction of the aperture of the lens. The detection occurs so that scattered light is only collected within an annulus determined by the radius of the illumination (inner radius) and the radius of the aperture of the lens (outer radius). This set-up possesses the disadvantage that the illumination reduces the numerical aperture of the lens, and this worsens the lateral and depth resolutions and causes loss of light. Another method suggested in these references utilises a half stop to achieve the dark field imaging and this has similar disadvantages.

Considering further the possibilities offered by conventional optics, the second solution could be to apply an arrangement where the resolution and the collected light are maintained but specular reflection is filtered out. Fourier filtering appears to be a straightforward solution for the problem whereby application of appropriate lens combinations provides the possibility to separate the spatial frequency components of the reflected/scattered light. In such an arrangement the specular reflection would be represented in the spatial frequency domain by low frequency components and scattered light as higher frequency components. Based on numerical results not presented here, the approximate value of surface reflection suppression against scattered light that can be achieved in this way is in the order of $\sim 10-20\times$. Comparing this number to the assumed difference between the intensity of the specular reflection and the scattered light (three

orders of magnitude) it becomes obvious that the Fourier filtering method can decrease the effect of specular reflection but not suppress it. The Fourier filtering results in loss of axial resolution and some loss of light.

The third solution to suppress specular reflection could be a method which can take out the DC components from the measured signal. In principle, this could be done with electronic differentiation (for example by applying an AC amplifier) assuming that the detector possesses an infinite dynamic range and the noise of the system is much smaller than that of the signal strength from the scatterer†. Unfortunately real detectors have limited dynamics and so do the associated amplifiers, thus this method has little practical significance. An alternative way to approach the problem is to consider such optical set-ups where the detection either does not require infinite detector dynamics or the differentiation (or DC level subtraction) occurs by optical means. The author in general is in favour of optical processing, and electronic manipulation of signals should only be used when optics do not offer a simple solution. A nice example for the mixed optical and electronic signal manipulation is differential phase contrast (DPC) microscopy. An elegant solution for full optical signal manipulation is provided by the differential interference contrast (DIC) technique. The use of the DPC or DIC techniques does not result in loss of either light or resolution.

Finally it must be emphasised again, with reference to Chapter 1, that the simplest method to investigate semiconductor specimens is the transmission non-confocal mode provided that both surfaces of the specimen are polished and the highest resolution is not required.

In the following we overview the various imaging modes of the SIRM. Of these modes all have been used previously with visible light, except the interference mode (Section 3.11) which is newly proposed here. For infra-red light only the transmission non-confocal mode (Section 3.5) and some initial studies with the reflection confocal mode[107] have been used previously.

†Various aspects of system noises and their effect on the instrumentation are considered in Chapter 10.

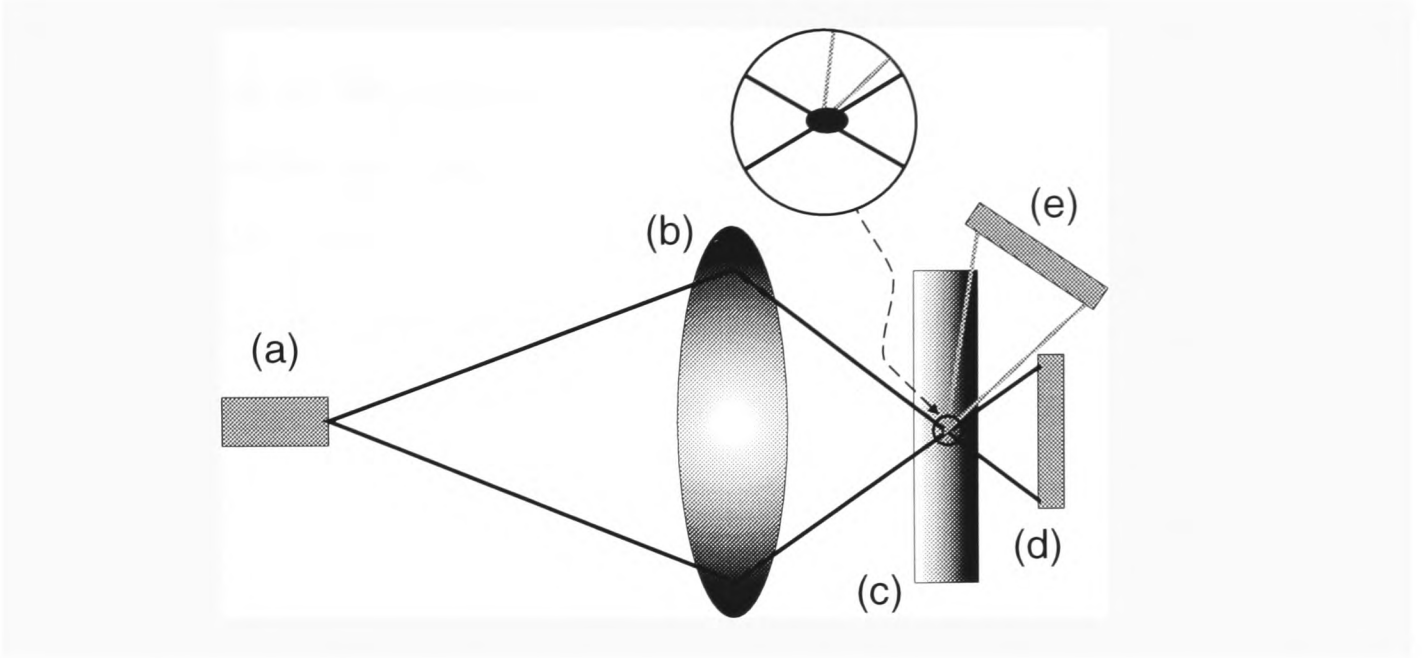


Figure 3.1: The transmission non-confocal SIRM. For notation see text.

3.5 Transmission non-confocal SIRM

A configuration of the SIRM is shown in Fig. 3.1 which embodies the simplest example for a transmission non-confocal scanning microscope.

Light emerges from the laser (a). The diagram shows an optical system where the lens is corrected for finite tube length. The beam of the semiconductor laser should have a divergence angle that matches the input aperture of the lens when it is placed in the object plane. Since the divergence angle of semiconductor lasers is usually higher than required for this application, box (a) needs to contain additional optical components. This could be a lens that images the exit aperture of the laser to the object plane of the objective lens (and at the same time magnifies or demagnifies the diameter of the exit aperture of the laser to the required diameter of the image in the object plane of the objective lens), or if the objective lens is corrected for an infinite tube length it could be a lens in front of the semiconductor laser that collimates the light from it. An alternative way could be the use of a gas or solid state laser. In this case, depending on the tube length correction of the objective lens, the box could contain, for example, a spatial filter with a beam expander.

Light is incident on the input aperture of the lens (b) which images/focusses the light inside the specimen (c). Light traverses the specimen and is collected by either detector

(d) or detector (e). The specimen is scanned with respect to the optical beam[‡] and the image is built up in the memory of a computer where position data and an amplified signal of the detector are stored. As the magnified area of the specimen in Fig. 3.1 shows, if a particle is situated in the focus region then light is scattered from this particle. We consider first the effect of a detector (e) displaced from the optical axis.

As a particle approaches the focus, light is scattered by the particle. The strength of the scattered light depends on the distance between the particle and the diffraction focus. It is advantageous (see Chapter 8) to detect the scattered light close to the optical axis, as it gives the strongest scattered intensity. If the detector is displaced from the optical axis and detects no light propagating within the solid angle determined by geometrical optics, then the detection is said to be *dark field.* There are some advantages to detect scattered light in this way. For example, a commercial instrument (LST) is available that uses a detector displaced by 90° with respect to the optical axis. Dark field images of large oxide particles in a silicon specimen were successfully obtained using a displaced detector and the transmission non-confocal SIRM[55].

In the usual transmission non-confocal set-up, the detector (d) is placed on the optical axis close to the specimen. In such a set-up the detection is said to be *bright field.* The particle contrast mechanism in this case is substantially different from that of dark field as here the scattering is not responsible primarily for the contrast but the blocking effect (Chapter 10), at least for the particle size range of present interest. Applications of this SIRM mode were the subject of several papers[52,53] and a thesis[54].

The advantages of this method are as follows. This set-up is the simplest and easiest to align. It does not require high precision fine-mechanics or optics. Disadvantages are that the lateral and depth resolutions are essentially similar to those of a conventional microscope. It follows that the particle number density range that can be examined is limited, typically $10^5 - 10^8 \mathrm{cm}^{-3}$. Furthermore, the specimen must be polished on both sides. The size of the detector is a dominant parameter since either it must be placed close to the specimen to collect sufficient light or must possess a large area. This SIRM

[‡]The advantages and disadvantages of on- and off-axis optical systems are considered in Chapter 10.

mode is sensitive to background light or heat radiation.

3.6 Reflection confocal SIRM

When the collection of the scattered light occurs with the same lens which illuminates the specimen, the operation of the SIRM occurs in the reflection mode. This operational mode can be classified as dark field imaging. Due to the need to reduce the surface reflection, as discussed in Section 3.4, the application of a pin-hole in front of the detector, i.e. confocal mode, is essential. This SIRM can be considered as a Type 2 scanning microscope.

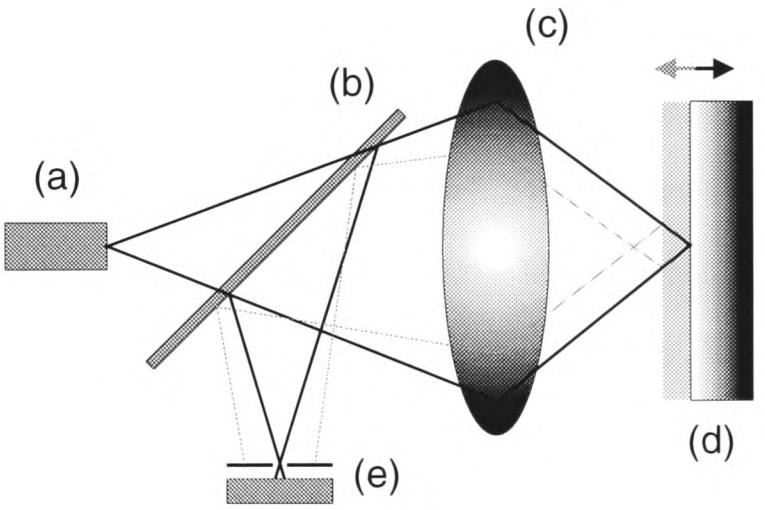


Figure 3.2: The reflection confocal SIRM. For notation see text.

The schematic diagram of the microscope is shown in Fig. 3.2 where box (a) represents the laser to which similar considerations apply as those described in Section 3.5. Light emerging from the laser passes through a beam splitter (b) and is then imaged and demagnified by the probe-forming objective lens (c) into the specimen (d). If the specimen surface is in focus, then light specularly reflected from the surface is collected and imaged back to the rear image plane of the objective lens (c) through the beam splitter (b) and light passes through the pin-hole (e) and reaches the detector. When defocussing occurs (indicated by a brighter specimen brought forward) the image of the specimen surface does not coincide any more with the position of the pin-hole, therefore less surface intensity is detected. When the surface is out-of-focus, particles at a certain depth within

the specimen are in-focus. The scattered light from these particles is focussed onto the pin-hole and reaches the detector. These particles can then be imaged without being 'swamped' by light reflected from the specimen surface. Although this description of the principle of a confocal microscope is simple, it may be slightly misleading as the imaging of a real confocal microscope cannot be described entirely by geometrical optics.

The optical path consisting of the laser, the objective lens and the specimen is usually referred to as the *illumination path*, and the optical path consisting of the specimen, the objective lens, the beam splitter, the pin-hole and the detector is usually referred to as the *detector path*. The specimen volume illuminated by the objective lens in the focal region is referred to as the *probe*. The distance between the Gaussian focus within the specimen and the front surface of the specimen is termed the *penetration or focussing depth*.

An important difference between the transmission non-confocal and reflection confocal modes is that if an objective lens with finite tube length correction is used (as depicted in Fig. 3.2) then the detector can be directly placed at the rear image plane of the lens. If the lens possesses an infinite tube length correction, however, then the specularly reflected light from the surface of the specimen is collected and re-collimated by the objective lens, thus an additional lens is needed to focus light in the detector path to the pin-hole. The reflection confocal SIRM is the most important operational mode and is further described in Chapter 10.

The advantages of this mode are as follows. The initial alignment of the optical system is not too difficult since it can be done by scanning a plane reflector along the optical axis through the focal plane. By this method the microscope can also be checked for the presence of spherical aberration. The proper alignment, however, involves more work as the proper lateral position of the pin-hole cannot be determined by the plane-reflector scanning method. For this purpose one can use a knife-edge scanned laterally through the probe in the focal plane. By observing the infringement on the image of the edge, some conclusions can be drawn for the correct position of the pin-hole[§]. A further ad-

[§]This method essentially refers to the results obtained by Carlini and Wilson[19] and Wilson and Carlini[111].

vantage of the reflection confocal SIRM is that once the optical system is aligned, there is no need to realign the system, except perhaps for correction of spherical aberration when the focussing depth is changed. The typical range of particle number density that can be observed by this operational mode is 10^6 to 10^{10}cm^{-3} which follows from the improved lateral and depth resolution with respect to the transmission non-confocal SIRM. The main disadvantages are as follows. The surface reflection prevents the detection of particles from 0 to 50 μm depth. The direction of detection coincides with the direction of minimum scattering of light from the particles (Chapter 8) which makes the method slightly inefficient. As this set-up is more sensitive to mechanical vibration than the transmission non-confocal mode, extra care needs to be taken with the fine mechanical design and construction.

3.7 Transmission confocal SIRM

With some modification the transmission non-confocal SIRM can achieve better lateral and depth resolution. When a lens is placed in front of the detector to collect the transmitted light and image the probe to the detector through a pin-hole, the optical arrangement is called the transmission confocal SIRM. This mode is a bright field, Type 2 microscope and its block diagram is shown in Fig. 3.3.

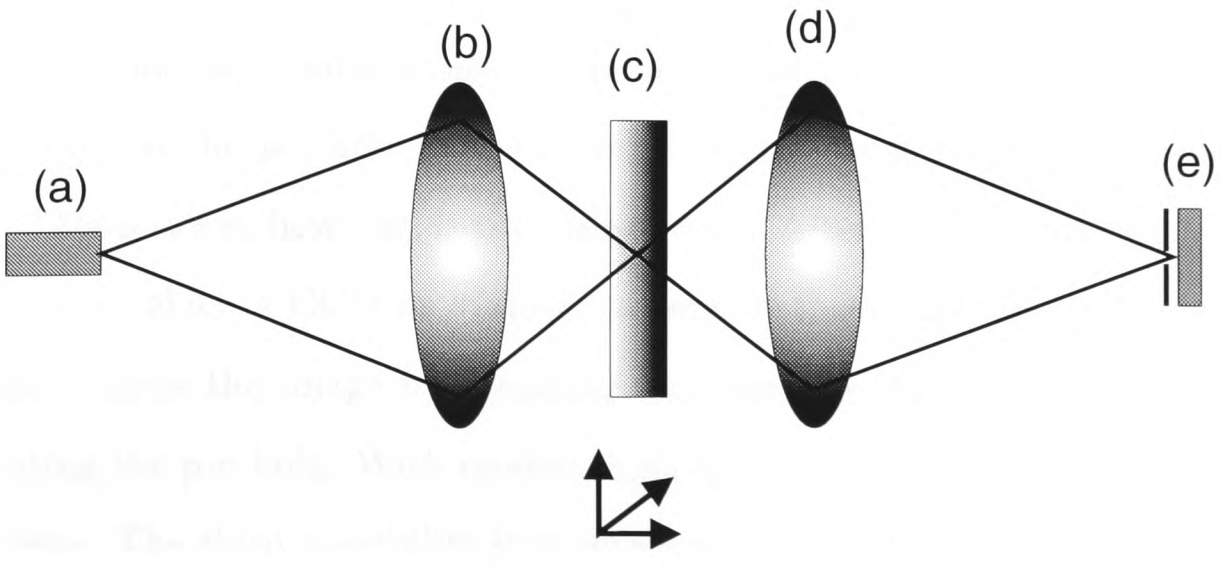


Figure 3.3: The transmission confocal SIRM. For notation see text.

Light emerging from the laser (a) is imaged into the specimen (c) by the probe-forming lens (b). Light traversing the specimen is collected by the collector lens (d) and imaged to the detector through a pin-hole (e). The specimen is raster scanned in a similar manner described above for the reflection confocal SIRM. For the box schematically shown and denoted as (a) the same considerations apply as for the reflection confocal SIRM. When the collector lens is corrected for infinite tube length then an additional lens should be applied to focus the light beam collimated by the collector lens to the detector. The application of the transmission confocal microscope for visible and ultra-violet light is well known, mainly for biological specimens. The transmission confocal SIRM can be applied in material science to detect precipitates in semiconductors especially when the power of the illumination laser is not high enough to use it in reflection confocal mode.

This set-up possesses however a big disadvantage in that changes in the position and/or thickness of the specimen modify the lateral position of the imaged probe on the pin-hole (e). In particular, when the specimen surface is not aligned perpendicular to the optical axis and is moved in the lateral direction (e.g. scanned), then the direction of each ray pencil traversing the specimen will be different depending on the focussing depth. The above process results in a laterally misplaced image at the detector (e). A specimen of non-uniform thickness gives the same result.

There are many ways to overcome this disadvantage. It is possible to mount a mirror onto a two-axis piezo rotator and to reflect the light to the pin-hole. If this arrangement is applied, then for each raster point the maximum intensity is found by scanning the image of the probe on the pin-hole and stopping the movement when the maximum intensity occurs. This process, however, is time consuming. Alternatively, an expensive but elegant solution is to place a CCD or vidicon camera in the image plane of the collector lens and then process the image by summing the measured light within an imaginary circle representing the pin-hole. With modern high speed image processor ICs this can be done in real time. The third possibility is to apply a collector lens which is infinite tube length corrected and to collimate the light beam traversing the specimen. Then the collimated beam is reflected back to the lens, traverses the specimen again and a detector path similar

to that of the reflection confocal SIRM can be applied to detect the light. This optical arrangement is called double-pass SIRM and is described in the literature. In spite of all possible solutions to the problem, the transmission confocal SIRM is not commonly used in material science.

The main advantages of this imaging mode are as follows. The surface reflection does not affect the imaging and therefore investigations can be performed for particles close to the surfaces. The lateral and axial resolutions are high which makes possible to examine specimens in the particle number density range of 10^6 to 10^{10} cm^{-3}. The main disadvantage is misalignment arising from the specimen thickness variations.

3.8 Reflection confocal DPC SIRM

Differential phase contrast (DPC) microscopy is a possibility to suppress the effect of specular reflection from the surface of the specimen in the reflection mode and further advantages are described in Chapter 10. This imaging mode by the usual classification is a dark field, Type 2 microscope.

The schematic diagram of the reflection mode DPC SIRM is shown in Fig. 3.4. The light of the laser (a) is incident on a collimator lens (b) and traverses a beam splitter (c). Light is focussed by the probe forming lens (d) into the specimen (e). Reflected/scattered light is collected by the probe forming lens and split by the beam splitter (c). In the detector path another beam splitter halves the light beam in such a way that one half is split towards one detector lens (h) and the other half after traversing the beam splitter is incident on the other detector lens (h). Before the light would reach the detectors through the pin-holes (g), two half-stops are placed into the detector path in such a way that each of them stops a different half of the incident beam. The alignment of this system is difficult, moreover the light loss due to the half stops can be significant. To overcome this problem in practice, instead of using the second beam splitter a half mirror can be placed so that it transmits one half of the light and mirrors the other half towards the other detector. There is also a possibility to use two closely spaced detectors placed behind a pin-hole. This latter arrangement provides the best solution against misalignment in the

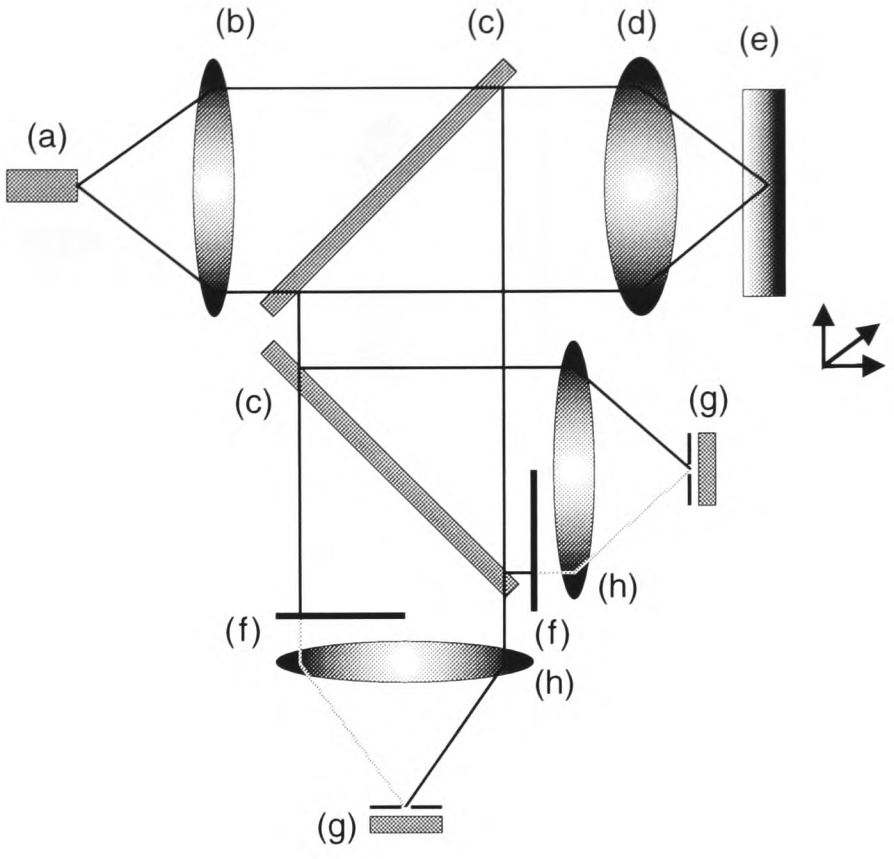


Figure 3.4: The reflection confocal DPC SIRM. For notation see text.

system.

The operation principle of this mode is as follows. When the specimen surface is in focus, both detectors give the same intensity. If the signals of the two detectors are added, the resulting signal is identical to that of a reflection confocal microscope. If however the signals are electrically subtracted, then no signal is detected. If a lateral scan is performed then no signal is detected as long as the surface of the specimen is evenly smooth. When a line scan occurs through an imperfection on the surface, the light is scattered resulting in a direction change in the collimated beam in the detector path. This causes the image of the probe at the detector plane to be laterally displaced with respect to the optical axis. It follows that one detector has an increased signal whilst the other has a decreased signal and the difference is no longer zero. The same principle applies when the probe is scanned through a particle in the bulk specimen, the only difference being the absence of specularly reflected light on the detectors. It is found experimentally (see Chapter 10) that for oxide particles in silicon, the difference signal gives a bright/dark image with both halves equally pronounced when the particle is in focus.

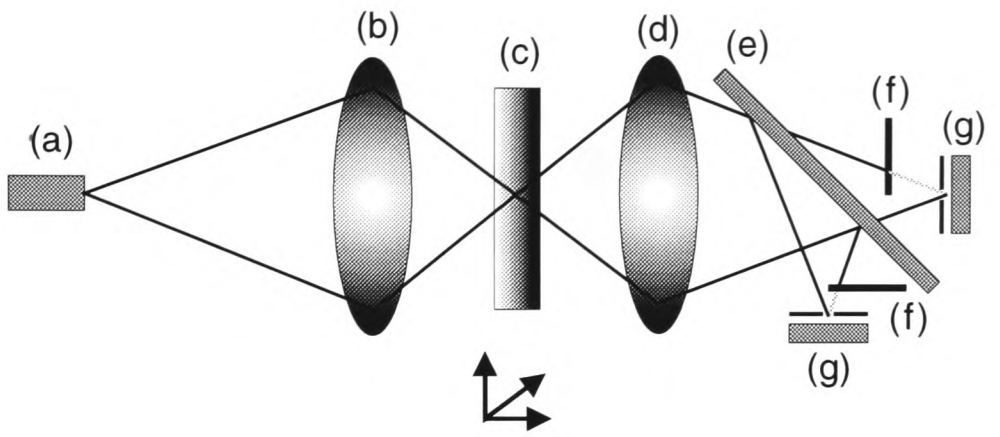


Figure 3.5: The transmission confocal DPC SIRM. For notation see text.

The advantages of the reflection confocal DPC SIRM are as follows. It suppresses the effect of specularly reflected light from the surface of the specimen. Based on experimental data it is found that this imaging mode offers the possibility of distinguishing between in-focus and out-of-focus particles from a single measurement. By utilising two closely spaced detectors, e.g. a single chip detector divided into two halves, the alignment of the system should be similar to that of the reflection confocal microscope. The main disadvantage of the method is that when no split detector is available, the alignment is difficult and the system rarely gives optimum performance.

3.9 Transmission confocal DPC SIRM

When the differential contrast principle is applied in the transmission mode then the microscope is called a transmission confocal differential phase contrast SIRM. Its schematic diagram is shown in Fig. 3.5

The operational principle is the same as that of the reflection confocal DPC SIRM. Light emerging from the laser (a) is incident on the probe forming lens (b) which images the exit aperture of the laser into the specimen (c) so forming the probe. Transmitted light is collected by the collector lens (d) which images the probe through a beam splitter (e) onto two detectors (g). In front of each detector a pin-hole is placed. The half stops

(f) are arranged so that when the probe is not blocked or scattered, both detectors give the same intensity of the light. For this set-up the use of a half mirror and/or two closely placed (or split) detectors behind a single pinhole can also be successful. By the usual classification this imaging mode embodies a bright field, Type 2 microscope.

In the transmission mode the specular reflection from the surface of the specimen does not play any role. Therefore the significance of the application of the transmission confocal DPC SIRM is different from that of the reflection confocal DPC SIRM. As discussed in Chapter 8 the contrast mechanism of a transmission microscope is different from that of the reflection SIRM. This difference is caused by the blocking effect of the particles as they block or absorb some of the incident light. There are however other features in the semiconductor specimens that affect the propagation of the light. These mainly are strain fields surrounding particles. By scanning through a small particle surrounded by a strain field, the effect of both imperfections changes the direction of propagation of light close the focus. This direction changing results in a laterally displaced image of the probe on the pin-holes and therefore in some change in the difference signal. It is shown experimentally (see Chapter 10) that some particles could be revealed by the transmission confocal DPC SIRM which were invisible for conventional transmission confocal SIRM. The advantages of the above imaging mode are as follows. It is simple to align especially when using two closely spaced detectors or a split detector. It provides additional information with respect to the previous imaging modes as it images some particles which would not have revealed. The disadvantages of this method are the same as those of a conventional transmission confocal SIRM.

3.10 Nomarski confocal DIC SIRM

Probably one of the most elegant ways to eliminate the specularly reflected light can be achieved by using differential interference contrast techniques in the reflection mode. In particular, unlike the reflection DPC imaging mode, the subtraction of the signals can be performed by optical means which makes the dynamics of the system infinite. Not all of the interference contrast techniques are advantageous for SIRM purposes but only those

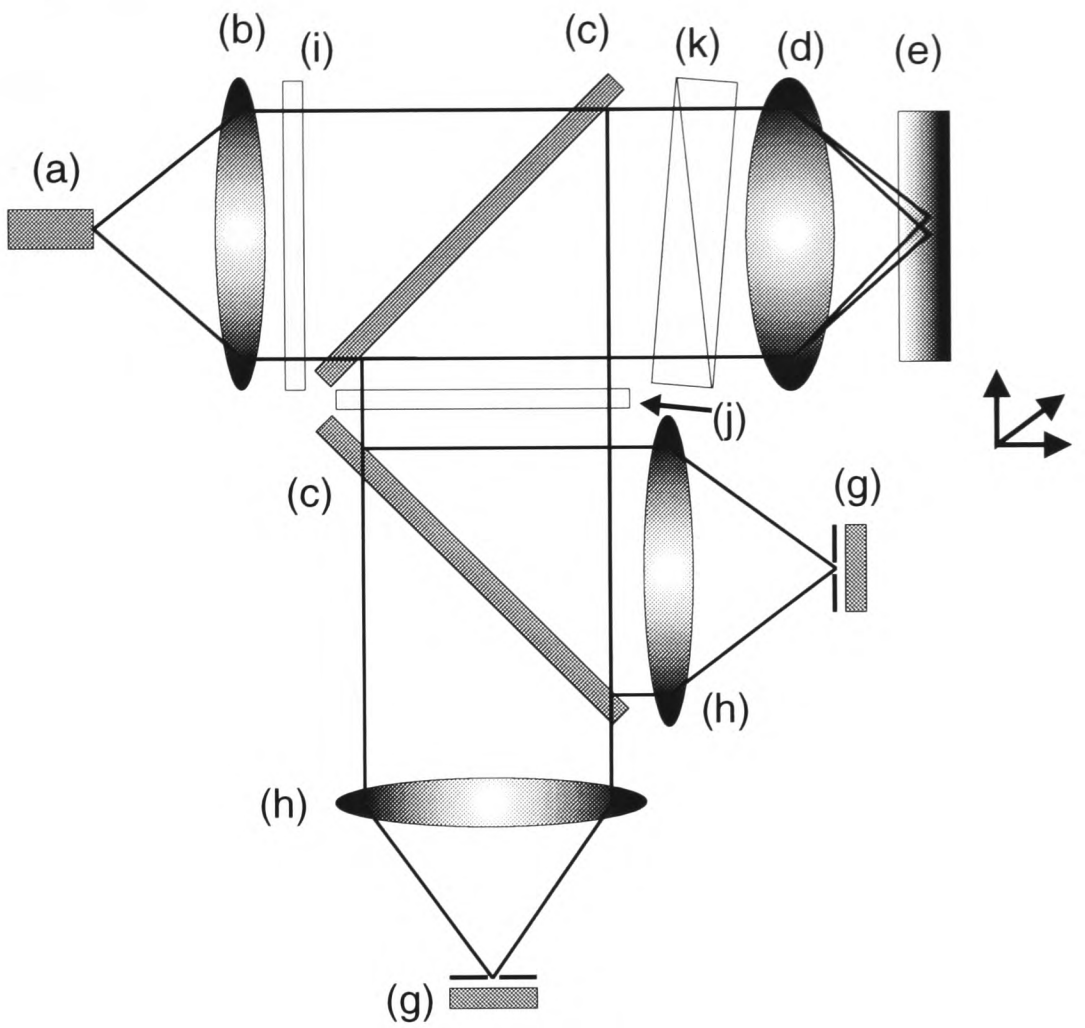


Figure 3.6: The reflection confocal DIC (Nomarski) SIRM. For notation see text.

where the two arms of the interferometer traverse the same optical path, which is primarily due to stability reasons. Such interferometers are called common path interferometers. The first interference technique described here is the Nomarski differential interference contrast (DIC) microscope whose schematic diagram is shown in Fig. 3.6.

This set-up is a dark-field, differential interference contrast, confocal SIRM embodying the Nomarski principle. The light is provided by a linearly polarised laser (a) and is collimated by a lens (b) and passed through a quarter-wave plate (i) which produces circularly polarised light. The light passes through a non-polarising beam-splitter (c) to a Wollaston prism (k). The Wollaston prism splits the incident light into two orthogonal, linearly polarised and equal intensity rays. A probe forming lens (d) focusses the rays into the specimen (e), so two closely spaced focussed probes are produced. The displacement between the two spots is referred to as the shear.

Light scattered back from particles present in the specimen is collected by the probe-forming lens (d) and recombined by the Wollaston prism (k). Light is then passed to a polarising beam-splitter (l). This beam-splitter is arranged so that its principal axis is at 45° with respect to the principal axis of the Wollaston prism. It acts as an analyser since light having the same polarisation direction as the light after the half wave plate will be transmitted completely to one of the detectors (g), while that having other polarisation components will be split to the other detector. If a particle is near the front surface of the wafer, then strong reflected light will result from the surface compared with weak back-scattered light from the particle, and so the reflected light signal needs to be eliminated. Backward propagating light is recombined by the Wollaston prism, and in the case of a perfect planar reflector object, the recombined beam possesses circular polarisation. Light is split by the non- polarising beam splitter. A quarter wave plate (j) is inserted to the optical path which converts the circular polarisation of the incoming beam to linear polarisation. If one of the rays from the two probes undergoes a phase shift, then the recombined ray will no longer be circularly but elliptically polarised. An appropriately aligned quarter-wave plate in the case of a circularly polarised incoming light, produces a linearly polarised light whose polarisation is at 45° with respect to the principal axis of the wave-plate. In this case, it could be either vertically or horizontally oriented. Let us assume that the direction of polarisation is horizontal if the incoming light possesses circular polarisation. If the light becomes elliptically polarised, the wave plate produces a linearly polarised light, whose direction of polarisation will be different from the horizontal. The linearly polarised light is analysed by the polarising beam splitter. If the direction of the polarised light after the wave plate is horizontal, then the beam splitter transmits the light entirely and no light will be split towards one detector. If the direction of polarisation changes, the beam splitter splits the vertical component of the light so that some signal will be received by the other detector. Thus the latter detector gives the optically subtracted phase contrast image, while the former detector gives the optically added amplitude contrast image. If the Wollaston prism introduces some phase difference between the two rays, then the quarter-wave plate has to be aligned accordingly.

The alignment of the Nomarski DIC SIRM can be considered by adopting the visible light results of Cogswell and Sheppard[23]. The desired imaging mode to suppress surface reflection is a zero bias retardation. In this setting the contribution of amplitude signal is minimal, whilst phase signal is maximal. The amount of the shear is chosen so that the maximum frequency transmittance occurs for high spatial frequencies (corresponding to small particles). This is when $m_0\Delta = 1/4$, where $m_0 = 2(\sin\alpha)\lambda$ is the cut-off frequency and Δ is the lateral shear.

The advantages of the reflection confocal Nomarski DIC SIRM are as follows. It can measure those regions near the specimen surface which were previously observable only by transmission methods. The sensitivity due to the DIC detection is increased and some particles can be imaged (see Chapter 10) that is not possible by other methods. By rotating the wave plates and the Wollaston prism in the system, all the DIC and the conventional imaging modes can progressively be interchanged. The main disadvantage of this system is that no comprehensive theoretical treatment is available on the imaging mode which makes it increasingly difficult to align the system to the optimum performance. The reason for the lack of theoretical treatment is mainly the extremely complex image formation of this microscope.

3.11 Interference confocal SIRM

The Nomarski DIC technique, as its name implies, can only provide *differential* interference contrast images. In order to obtain *real* interference images from the specimens a new method is here suggested that utilises a common path interferometer. The main concern is providing the reference beam for the interference. This is the reason why, for example, the Michaelson interferometer is not practically available for SIRM purposes. Instead, the reference path is now provided from the surface of the specimen as shown in Fig. 3.7, which needs to be of sufficiently good quality to reflect back the plane waves incident upon it.

The *coherent* light is provided by a linearly polarised laser (a) and is collimated by a lens (b) and passed through a half-wave plate (i). From there the light passes through

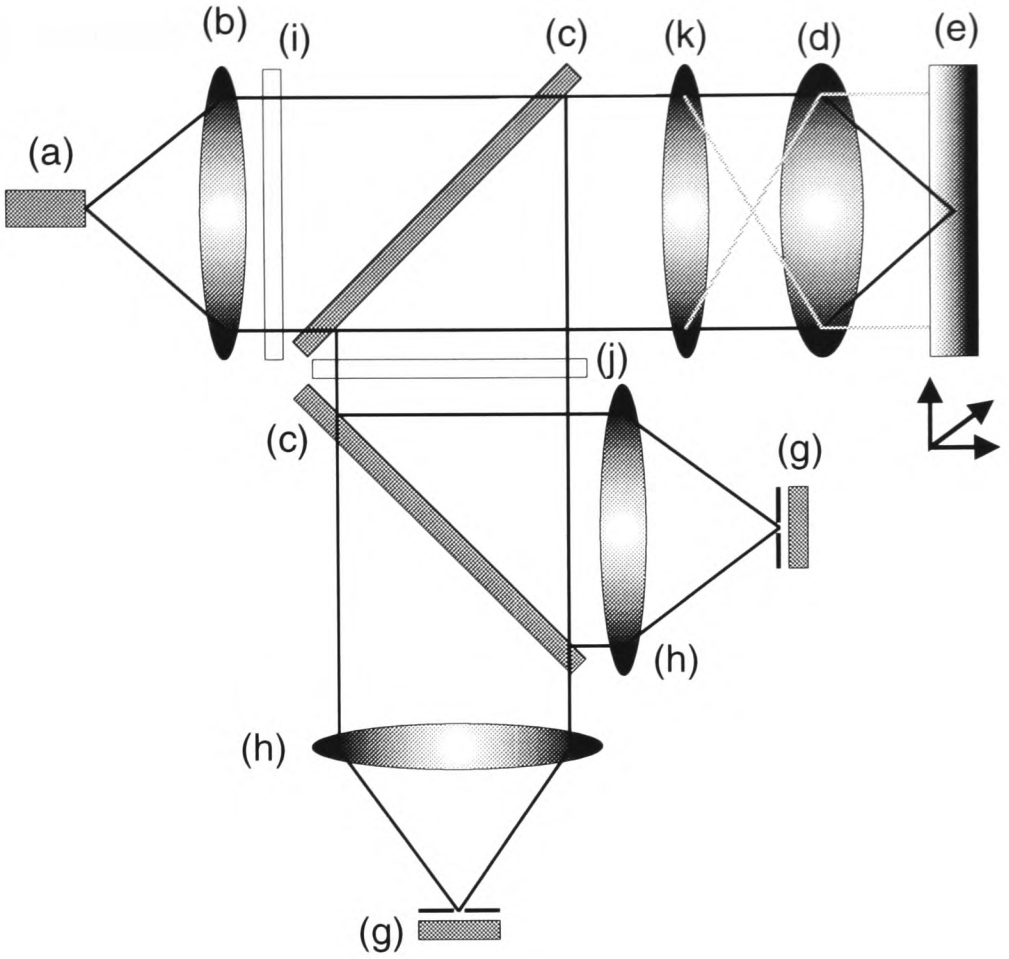


Figure 3.7: The reflection confocal interference SIRM. For notation see text.

a non-polarising beam-splitter (c) to a low numerical aperture birefringent lens (k), with the polarisation direction of the incident beam being at 45° from the principal axis of the birefringent lens. The birefringent lens splits the incoming light into two linearly polarised and equal intensity rays. The front focus of this lens is set to be in common axial position with the back focus of the probe forming lens (d) which focusses one of the rays (ray 1) into the specimen (e), so one focal spot is produced. The ray with the other polarisation direction (ray 2) traverses the lens without being focussed. Reflection of this latter ray from the surface (which is optically smooth and planar) provides the constant phase reference beam. Light of ray 1 back-scattered from particles is collected by the probe-forming lens (d). Light of ray 2 is reflected back from the surface of the specimen and traverses the probe forming and birefringent lenses. Light is then passed to a polarising beam-splitter (l). This beam-splitter is arranged so that its principal axis is at 45° with respect to the principal axis of the birefringent lens. It acts as an analyser since light having the same polarisation direction as the light after the half wave plate will

be transmitted completely to one detector (g) while that having the other polarisation direction will be split to the other detector. If the object is a perfect planar reflector, both of the rays have the same phase and so the recombined ray has the same polarisation direction as the illuminating ray. If a particle is the object, ray 1 suffers a phase shift with respect to ray 2, so the recombined ray no longer maintains the original polarisation direction. If the microscope is aligned so that the first detector measures zero intensity with a perfect planar reflector object, its signal is increased due to the polarisation change and the signal of the second detector is decreased.

The advantage of this method is that it provides an interferometer quality image of the specimen, especially from the critical surface region, and at the same time the set-up optically filters out the specularly reflected light. The sensitivity of this imaging mode is the best of all of the described set-ups. A disadvantage is that a coherent laser must be used and so relatively cheap and small semiconductor lasers cannot be applied.

3.12 Other imaging modes

The above described imaging modes are the most important optical arrangements for the SIRM. There are however several other operating modes (such as double reflection, multi-pass, folded detector path, etc.) that could be used. It must be emphasised again that the main difference between the applications of the various SIRMs and other microscopes is the material into which the focussing occurs as the usual semiconductors possess high refractive indices. A further difference is that the abrupt change in refractive index on going from air into the semiconductor cannot be matched by using immersion fluids as this would be highly contaminating to the specimen and because such high refractive index immersion fluids are not available.

CHAPTER 4

Focussing theories

"God runs electromagnetics by wave theory on Monday, Wednesday, and Friday, and the Devil runs them by quantum theory on Tuesday, Thursday, and Saturday."
W. Bragg

4.1 Introduction

In this work the focussing theories play a key role. It is the aim of the present chapter to show how different focussing theories can be developed and also to establish some important correspondences between different results. The differences between the focussing theories are revealed and considerations are given to the application of these theories to confocal microscopy and SIRM. This we do since the high aperture theory in confocal microscopy and therefore in SIRM is not yet fully developed, although authors of some papers proposed different treatments.

4.2 High aperture focussing theories

In the following we give a brief account on the two main interpretations of the focussing problem. These two are different from one another in the physical sense as the Kirchhoff theory expresses the field as a sum of secondary spherical waves, whilst the Debye theory represents the field as a sum of plane waves.

4.2.1 The Kirchhoff theory

Let us consider a monochromatic scalar wave

$$u = v\,\mathrm{e}^{-\mathrm{i}\omega t}$$

where v is a solution of the Helmholtz equation (time-independent wave equation):

$$(\nabla^2 + k^2)\, v = 0 \tag{4.1}$$

and k is the wave number. To obtain the Kirchhoff integral we use Green's identity:

$$\iiint_V (v\nabla^2 w - w\nabla^2 v)\mathrm{d}x\mathrm{d}y\mathrm{d}z = \iint_S \left(v\frac{\partial w}{\partial \nu} - w\frac{\partial v}{\partial \nu}\right)\mathrm{d}S \tag{4.2}$$

where V is a volume bounded by a closed surface S and $\partial/\partial\nu$ denotes the differentiation along the outward normal to S. The validity conditions for Green's identity are that *both* v and w and their first and second partial derivatives are continuous within and on S and that w satisfies the Helmholtz equation. When these continuity conditions are satisfied then it follows from Helmholtz's equation

$$\iint_S \left(v\frac{\partial w}{\partial \nu} - w\frac{\partial v}{\partial \nu}\right)\mathrm{d}S = 0.$$

If we now substitute $w = \mathrm{e}^{\mathrm{i}kr}/r$ into the above equation where r denotes distance from a fixed point P we obtain:

$$\iint_S \left\{v\frac{\partial}{\partial \nu}\left(\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\right) - \frac{\mathrm{e}^{\mathrm{i}kr}}{r}\frac{\partial v}{\partial \nu}\right\}\mathrm{d}S = 0 \tag{4.3}$$

provided that P and the singularities of v are outside S. Imposing the latter condition on P would prevent the successful use of the Green's identity. Instead, we bound the singularity at $r = 0$ by a sphere σ with a radius of ϵ, centered at P, and perform the integration between S and the surface of the sphere. Thus we obtain:

$$\begin{aligned}\iint_S \left\{v\frac{\partial}{\partial \nu}\left(\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\right) - \frac{\mathrm{e}^{\mathrm{i}kr}}{r}\frac{\partial v}{\partial \nu}\right\}\mathrm{d}S &= -\iint_\sigma \left\{v\frac{\partial}{\partial \nu}\left(\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\right) - \frac{\mathrm{e}^{\mathrm{i}kr}}{r}\frac{\partial v}{\partial \nu}\right\}\mathrm{d}S \\ &= \lim_{\epsilon\to 0}\iint_\sigma \left\{v\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\left(\mathrm{i}k - \frac{1}{r}\right) - \mathrm{e}^{\mathrm{i}kr}r\frac{\partial v}{\partial \nu}\right\}\mathrm{d}S\end{aligned}$$

This integral can be evaluated and the surface element on σ is $\epsilon^2 \sin\theta \mathrm{d}\theta\mathrm{d}\phi$. The value of the limit is $-4\pi v(P)$, where $v(P)$ indicates the value of v at point P. Thus the field is given by:

$$v(P) = \frac{1}{4\pi}\iint_S \left\{\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\right)\right\}\mathrm{d}S \tag{4.4}$$

which is the integral theorem of Helmholtz and Kirchhoff. From this the Fresnel-Kirchhoff integral can readily be derived. For this we consider a wavefront propagating from a point source and incident on an opaque plane screen containing an aperture. The usual assumptions for this formulation that the aperture is large compared to the wavelength and the distance of the point source from the screen and the distance of the screen from the point in which the field is to be determined are large compared to the aperture. In principle, from Eq. (4.4) the field can be determined in any point of the space so only the surfaces of the integration must be set. Let us denote the aperture by $\aleph$, a portion of the opaque plane screen by ℓ and a large sphere centered at P by $\wp$. The field at P is then expressed by:

$$v(P) = \frac{1}{4\pi}\left[\iint_{\aleph} + \iint_{\ell} + \iint_{\wp}\right]\left\{\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\frac{\partial v}{\partial \nu} - v\frac{\partial}{\partial \nu}\left(\frac{\mathrm{e}^{\mathrm{i}kr}}{r}\right)\right\}\mathrm{d}S \tag{4.5}$$

To solve this equation we have to know the *exact* values for both v and $\partial v/\partial \nu$ which are never known. It is reasonable to impose, however, the following boundary conditions:

$$\left.\begin{array}{lll} v = v^{(i)}, & \dfrac{\partial v}{\partial \nu} = \dfrac{\partial v^{(i)}}{\partial \nu} & \text{on } \aleph, \\ v = 0, & \dfrac{\partial v}{\partial \nu} = 0 & \text{on } \ell, \end{array}\right\} \tag{4.6}$$

where

$$v^{(i)} = \frac{A\mathrm{e}^{\mathrm{i}kr}}{r}, \qquad \frac{\partial v^{(i)}}{\partial \nu} = \frac{A\mathrm{e}^{\mathrm{i}kr}}{r}\left(\mathrm{i}k - \frac{1}{r}\right)\cos(\nu, r) \tag{4.7}$$

Equations (4.6) and (4.7) are the famous Kirchhoff's boundary conditions. It is worth pointing out that these boundary conditions are mathematically inadmissible. The reason for this is the Riemann's theory of functions which states that if a two-dimensional function together with its first derivative vanish at any finite section of a closed surface, then the function must vanish everywhere on the surface. Many researchers tried to impose different boundary conditions[61,92] but this problem is still the subject of extensive investigations in diffraction optics*. With the above boundary conditions the only specification not yet made is on $\wp$. If we consider that the radius of that sphere is infinitely large

*Sommerfeld (*loc. cit.*) showed that by imposing his radiation condition, Kirchhoff's boundary conditions become consistent.

and the irradiation from the source is started at a time differing from infinity, then we can argue, following Born and Wolf[14], that the contribution from the sphere to the field is arbitrarily small because it could not reach point P from infinity to contribute to the disturbance. Thus Eq. (4.5) takes the form:

$$v(P) = \frac{iA}{2\lambda} \iint_{\aleph} \frac{e^{ik(r+s)}}{r\,s} [\cos(\nu, r) - \cos(\nu, s)] dS \tag{4.8}$$

which is the Fresnel-Kirchhoff integral†.

4.2.2 THE DEBYE THEORY

To obtain the diffraction formula of Debye we start with Eq. (4.4) noting that for the derivation of this expression we have not used the boundary conditions of Kirchhoff. The geometry used for the solution is simple. Let us fix the origin of a co-ordinate system at some distance from a plane screen with a circular aperture, in such a way that the circular aperture is centered on the z axis of the co-ordinate system whose direction is pointing away from the screen. Consider a perfect sphere with its centre at the origin of the co-ordinate system which fills the aperture. We require the expression for the field at P which point is at the vicinity of the origin of the co-ordinate system. Let us fix a point on the surface of the sphere and denote this point by Q and let us characterise this point by its co-ordinate components in a spherical polar co-ordinate system $r_0, \vartheta_0, \omega_0$. The co-ordinates of a unit vector pointing from P towards Q is denoted by r, ϑ, ω. If we set now:

$$R = r_0 - r\cos\gamma \tag{4.9}$$

where

$$\cos\gamma = \cos\vartheta\,\cos\vartheta_0 + \sin\vartheta\,\sin\vartheta_0\,\cos(\omega - \omega_0)$$

then it is evident that $r\cos\gamma$ is the scalar product of the unit vector of $\overline{Q\,P}$ and the unit vector pointing from the origin of the co-ordinate system to P. We require the solution in

†Eq. (4.8) differs from the corresponding equation of Born and Wolf (*loc. cit*) by the sign. The reason for this is that we took the outward normal for the derivation of the Helmholtz and Kirchhoff theorem instead of the inward normal.

the form:

$$v(P) = \frac{e^{ikr_0}}{r_0} \qquad \text{and} \qquad \frac{\partial v(P)}{\partial \nu} = ik\frac{e^{ikr_0}}{r_0},$$

and we also have from Eq. (4.9):

$$\left.\begin{aligned} \frac{e^{-ikR}}{R} &= \frac{e^{-ikr_0}}{r_0}e^{ikr\cos\gamma} \\ \frac{\partial}{\partial\nu}\left(\frac{e^{-ikR}}{R}\right) &= -ik\frac{e^{-ikr_0}}{r_0}e^{ikr\cos\gamma} \end{aligned}\right\}. \tag{4.10}$$

On substituting $v(P)$ into Eq. (4.4) we obtain:

$$v(P) = \frac{ik}{2\pi r_0^2}\iint_S e^{ikr\cos\gamma}\mathrm{d}S$$

or on expressing the unit surface element as $\mathrm{d}S = r_0^2\mathrm{d}\Omega = r_0^2\sin\vartheta_0\mathrm{d}\vartheta_0\mathrm{d}\omega_0$:

$$v(P) = \frac{ik}{2\pi}\int_0^{\alpha}\int_0^{2\pi} e^{ikr[\cos\vartheta\cos\vartheta_0+\sin\vartheta\sin\vartheta_0\cos(\omega-\omega_0)]}\sin\vartheta_0\mathrm{d}\vartheta_0\mathrm{d}\omega_0, \tag{4.11}$$

which formula is the diffraction integral of Debye.

4.2.3 LUNEBURG'S INTERPRETATION OF THE DEBYE THEORY

It is seen that Kirchhoff's diffraction integral overspecifies the field, i.e. it requires knowledge of both v and its derivative. Sommerfeld[92] suggested a method of how this problem can be avoided, namely the use of a different Green's function. For this, the point P is mirrored with respect to the plane screen and the mirrored point is denoted by S. We now require the solution at P in the form:

$$w = \frac{e^{ikr}}{r} - \frac{e^{ikr_1}}{r_1} \tag{4.12a}$$

$$\frac{\partial w}{\partial z} = -\frac{2x_0}{r}\frac{\partial}{\partial r}\left(\frac{e^{ikr}}{r}\right) \tag{4.12b}$$

where r is the distance of an arbitrary point in the image plane from $P = (x_0, y_0, z_0)$ and r_1 is the distance of the mirror image $S = (x_0, y_0, -z_0)$ from P. It can be shown that, on substituting into Eq. (4.2), the solution in this case takes the form:

$$v(P) = -\frac{1}{4\pi}\iint_S f(x,y)\frac{\partial w(x,y,z)}{\partial z}\mathrm{d}x\mathrm{d}y \tag{4.13}$$

where $f(x,y)$ is the boundary value set in such a way that $v(x,y,0) = f(x,y)$ and which expression is essentially identical to the diffraction integral of Sommerfeld. If we now substitute Eq. (4.12b) into the above expression we obtain:

$$v(P) = \frac{z_0}{2\pi} \iint_{-\infty}^{\infty} f(x,y) \left(\frac{\mathrm{i}k}{r^2} - \frac{1}{r^3} \right) \mathrm{e}^{\mathrm{i}kr} \mathrm{d}x\mathrm{d}y \tag{4.14}$$

where $r = [(x-w_0)^2 + (y-y_0)^2 + z_0^2]^{1/2}$. To obtain the solution for the field we set an imaging system where points P_0 and P_1 are perfect conjugate points on the axis of rotation. Point P_1 is situated in the image space and is the point where spherical waves are converging to. The equation of the spherical wavefronts is given by Hamilton's point characteristic (eikonal):

$$\varphi(x,y,z) = C - R\,n_1$$

where $R = (x^2 + y^2 + z^2)^{1/2}$ is the distance from P_1, C is the optical distance between P_0 and P_1 and n_1 is the refractive index. If the electromagnetic vectors of the geometrical optics are known on a particular wavefront and these are denoted by $\mathbf{V}_0$ and $\mathbf{W}_0$, then on any other spherical wavefront the corresponding vectors ($\mathbf{V}_0^*$ and $\mathbf{W}_0^*$) are given by:

$$\mathbf{V}_0^* = \frac{1}{R}\mathbf{V}_0 \tag{4.15a}$$

$$\mathbf{W}_0^* = \frac{1}{R}\mathbf{W}_0. \tag{4.15b}$$

We set P_1 to be the origin of a spherical polar co-ordinate system of (R,ϑ,φ). It was suggested by Luneburg[60] that the solutions $(\mathbf{v},\mathbf{w})$ of Maxwell's equations could be found in a form where this solution has the same boundary values in infinity as the solution of geometrical optics:

$$\mathbf{v_0} = \frac{1}{R}\mathbf{V}_0(\vartheta,\varphi)\mathrm{e}^{\mathrm{i}k(C-n_1R)} \tag{4.16a}$$

$$\mathbf{w_0} = \frac{1}{R}\mathbf{W}_0(\vartheta,\varphi)\mathrm{e}^{\mathrm{i}k(C-n_1R)} \tag{4.16b}$$

thus the boundary conditions are:

$$\lim_{R\to\infty} R(\mathbf{v} - \mathbf{v}_0) = 0 \tag{4.17a}$$

$$\lim_{R\to\infty} R(\mathbf{w} - \mathbf{w}_0) = 0 \tag{4.17b}$$

Consider now a plane in the image space normal to the z axis, at $z = -D$. For the geometrical optical field at $z = -D$ we can write:

$$\mathbf{v}_0(\xi, \eta) = \frac{1}{R}\mathbf{V}_0(\vartheta, \varphi)\mathrm{e}^{\mathrm{i}k(C - n_1 R)} \tag{4.18a}$$

$$\mathbf{w}_0(\xi, \eta) = \frac{1}{R}\mathbf{W}_0(\vartheta, \varphi)\mathrm{e}^{\mathrm{i}k(C - n_1 R)} \tag{4.18b}$$

where $R = (\xi^2 + \eta^2 + D^2)^{1/2}$. The vector $\mathbf{v}$ must satisfy the wave equation (Eq. (4.1)). The solution is considered as follows. First, the solution is determined from wave optics (involving vectors $\mathbf{v}$ and $\mathbf{w}$) so that it has boundary values Eq. (4.18) at $z = -D$. Then the limit is calculated as shown by Eq. (4.17) which operation should result in the desired field.

It follows from Eq. (4.14) that:

$$\mathbf{v}_D(x, y, z) = \frac{D + z}{2\pi}\,\mathrm{e}^{\mathrm{i}kC} \int\int \mathbf{V}_0 \frac{\mathrm{e}^{-\mathrm{i}kn_1(R - r)}}{R}\left(\frac{\mathrm{i}kn_1}{r^2} - \frac{1}{r^3}\right)\mathrm{d}\xi\mathrm{d}\eta \tag{4.19}$$

where $r = [(\xi - x)^2 + (\eta - y)^2 + (z + D)^2]^{1/2}$, and

$$\xi = D\tan\vartheta\,\cos\varphi \tag{4.20a}$$

$$\eta = D\tan\vartheta\,\sin\varphi \tag{4.20b}$$

with which we obtain

$$\mathrm{d}\xi\mathrm{d}\eta = D^2\frac{\sin\vartheta\,\mathrm{d}\vartheta\mathrm{d}\varphi}{\cos^3\vartheta} \tag{4.21a}$$

$$R = \frac{D}{\cos\vartheta} \tag{4.21b}$$

$$r = \sqrt{(z + D)^2 + D^2\tan^2\vartheta - 2D\tan\vartheta(x\cos\varphi + y\sin\varphi)}. \tag{4.21c}$$

The solution is given if we let $D \to \infty$:

$$\mathbf{v}(x, y, z) = \frac{\mathrm{i}kn_1}{2\pi}\mathrm{e}^{\mathrm{i}kC}\iint \mathbf{V}_0(\vartheta, \varphi)\mathrm{e}^{-\mathrm{i}kn_1[\sin\vartheta(x\cos\varphi + y\sin\varphi) - z\cos\vartheta]}\sin\vartheta\mathrm{d}\vartheta\mathrm{d}\varphi, \tag{4.22}$$

which formula is identical to that of Eq. (4.11). It must be emphasised that this is an exact solution as it satisfies the homogeneous wave equation. This elegant solution was given by Luneburg[60] and this derivation plays an important role in Chapter 5.

4.2.4 WOLF'S INTEGRAL REPRESENTATION

Whilst the solution given by Luneburg is based on an idealised problem, another approach of the problem was given by Wolf[115]. The mathematical elegance which was present in Luneburg's formulation can also be found here, together with physical interpretations. Following Wolf's formulation, let us consider an optical set-up, similar to that treated by Luneburg, where for the image space the electric field can be written:

$$\mathbf{e}(P) = \iint_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}k(s_x x + s_y y)} \left[\mathbf{U}(s_x, s_y)\mathrm{e}^{\mathrm{i}ks_z z} + \mathbf{V}(s_x, s_y)\mathrm{e}^{-\mathrm{i}ks_z z}\right] \mathrm{d}s_x \mathrm{d}s_y, \tag{4.23}$$

where $k = k_0 n$ and (s_x, s_y, s_z) are components of the unit wave vector satisfying the relation:

$$\left.\begin{aligned} s_z &= \sqrt{1 - s_x^2 - s_y^2} \quad \text{when} \quad s_x^2 + s_y^2 \leq 1 \\ &= \mathrm{i}\sqrt{s_x^2 + s_y^2 - 1} \quad \text{when} \quad s_x^2 + s_y^2 > 1 \end{aligned}\right\}. \tag{4.24}$$

When $s_x^2 + s_y^2 \leq 1$ the waves are homogeneous waves, when $s_x^2 + s_y^2 > 1$ the waves are inhomogeneous waves. For the present considerations the inhomogeneous waves are ignored as for $\mathbf{V}$ they are divergent and for $\mathbf{U}$ are negligible. The second approximation to be made is imposing Kirchhoff's boundary conditions for the aperture which are valid as long as the field is examined far enough from the aperture. It was shown by Wolf that Eq. (4.23) can be approximated, under the assumption that $-k_z z' \gg 1$, by its asymptotic approximation:

$$\left.\begin{aligned} \frac{2\pi\mathrm{i}}{kr'} u_z \left\{-\mathbf{U}(-u_x, -u_y)\mathrm{e}^{-\mathrm{i}kr'} + \mathbf{V}(u_x, u_y)\mathrm{e}^{\mathrm{i}kr'}\right\} &= \mathbf{e}^{(i)}(P') \quad \text{if} \quad P' \in \aleph \\ &= 0 \qquad\quad \text{if} \quad P' \notin \aleph \end{aligned}\right\}, \tag{4.25}$$

where $P' = (x', y', z')$ is a point in the exit aperture $\aleph$ and

$$r' = \sqrt{x'^2 + y'^2 + z'^2} \tag{4.26a}$$

and

$$u_x = \frac{x'}{r'}, \qquad u_y = \frac{y'}{r'}, \qquad u_z = \frac{z'}{r'}. \tag{4.26b}$$

Wolf also shows that by using the eikonal function, the incident field can be expressed as:

$$\mathbf{e}^{(i)}(P') = \frac{\mathbf{a}(s_x, s_y)}{r'} \mathrm{e}^{\mathrm{i}k[C - r']}$$

where C is a constant and $\mathbf{a}(s_x, s_y)$ is a strength vector. On setting $s_x = -u_x$, $s_y = -u_y$ and $s_z = -u_z$ we obtain:

$$\begin{aligned} \mathbf{U}(s_x, s_y) &= -\frac{ik}{2\pi}\frac{\mathbf{a}(s_x, s_y)}{s_z}e^{ikC} \quad \text{when} \quad \mathbf{s} \in \Omega \\ &= 0 \quad \text{when} \quad \mathbf{s} \notin \Omega \end{aligned} \tag{4.27a}$$

and

$$\mathbf{V}(s_x, s_y) = 0 \tag{4.27b}$$

From this Wolf obtains on substituting into Eq. (4.23):

$$\mathbf{e}(x, y, z) = -\frac{ik}{2\pi}\iint_{\Omega}\frac{\mathbf{a}(s_x, s_y)}{s_z}e^{ik(s_x x + s_y y + s_z z)}ds_x ds_y \tag{4.28}$$

where Ω is the solid angle formed by all $\mathbf{s} = (s_x, s_y, s_z)$ vectors.

4.3 Application of high aperture focussing in confocal microscopy

The high aperture rigorous treatment only needs to be applied for special cases because the usual description of the confocal and non-confocal microscopes works surprisingly well. There are, however, some weak points in the paraxial treatment of confocal microscopy and because of these the use of the electromagnetic treatment for some problems seems to be inevitable.

In the previous Chapter we gave an overview on the literature of confocal microscopes. There we referenced two articles[102,100] and briefly mentioned that their treatments might have been somewhat oversimplified. The authors of the first paper, van der Voort and Brakenhoff, used a rigorous theory[‡] to obtain the electric field in the focus region. In order to simplify the mathematics they ignored the rotational asymmetry of the electric and magnetic field and considered the field to be rotationally symmetrical (corresponding to the 45° meridional plane). The emitted light from the specimen, being randomly polarised, was taken into account by the same formulation. They obtained the emission

[‡]They used a solution given by Jackson[46] of the Maxwell's equations and imposing Kirchhoff's boundary conditions.

spectrum due to a fluorescent specimen by approximating it with a Gaussian function and considering the emitted light from the specimen to be randomly polarised, incoherent and polychromatic. Then the point spread function of the emission was convolved with the geometric image of the pin-hole and the result was multiplied by the emission spectrum which expression was integrated for the different wave numbers. They obtained the overall point spread function of the microscope by multiplying the result of the above integration by the strength of the electric field corresponding to the illumination.

There are however some problems one can encounter with the above approach. It was shown by Wilson and Sheppard[112] that this representation *presumes* the paraxial approximation is valid. Let us suppose that the paraxial approximation is valid. For this case it was shown by Richards and Wolf[75] that only one vector component is enough to fully characterise the field in the focal region. It follows therefore that there is no need to treat the problem by the vectorial theory. In another article Hell *et* al.[43] calculate the intensity response of a confocal microscope by squaring the absolute value of the overall point spread function of their microscope. They do this after obtaining the electromagnetic field in the focal region by a full vectorial treatment. When this intensity calculation was performed by Wilson and Sheppard[112] it was stated clearly that the approximation was only valid for ideal point-like objects. It should also be noted that the within the accuracy of the paraxial approximation the assumption of an *ideal* point object is plausible, whilst when a rigorous treatment is applied the simplest equivalent to the point object is the radiating dipole.

There is no doubt that these simplifications of the theory could still lead to an approximation of high accuracy, but the conditions for the use of the above approximations were not stated clearly.

The SIRM can be treated in two different ways. First, by adopting the well developed theory of confocal microscopy. In principle, the description of three-dimensional objects is not a great problem because it has already been developed[113]. Also, the spherical aberration could have been included in the theory by means of the quadratic phase aberration component.

There was however a need to derive the theoretical treatment of perturbation of the field by precipitate particles embedded in semiconductor materials. For the solution of the problem it was found to be more comprehensive to consider a treatment which was different from those developed previously, having had the additional benefit and beauty of a rigorous solution. It was obvious that the problem of scattering could not be solved without a proper and rigorous treatment on the interface diffraction problem. Thus the solution of the problem has been considered as follows. The wavefront of the focussed field formed by the objective lens is taken to be perfect which step is justified in the next chapter. The interface diffraction is solved and the electromagnetic field is given in its vectorial form. Then the scattering is treated by considering the field incident on the particle as a sum of plane waves subtending the solid angle determined by the effective numerical aperture in silicon. The scattering is solved for each of these plane waves and the resulting field is given as the superposition of individual plane waves. Light propagating backward after the scattering by the particles traverses the interface once again where the solid angle within which light can be detected in the backward direction is well defined by the total internal reflection of the silicon/air interface. The light at the input aperture of the collector (probe-forming) lens is then given and a low aperture projection is performed to describe the light propagating towards the pin-hole. The polarisation of the light loses importance after traversing the interface since the polarisation dependent transmittance on the lens is not considered here. Therefore, after obtaining the field at the lens in the backward direction, the treatment is confined to expressing field at the pin-hole, back projected by the objective lens, with a scalar, low aperture treatment.

For general application, methods involving the sum of planar waves (Debye theory, Wolf's integrals) are favourable over the Kirchhoff approximation. This is because, in the most cases, manipulation of plane waves are far easier than that of spherical waves. It also follows that the perturbation caused by the object cannot be uniformised. There is a possibility, however, to use the method commonly referenced as angular spectrum representation, where the integral for the electric and magnetic field consists of the Fourier transform of a function. This function expresses the direction cosines of plane waves

subtending a solid angle over which the integration is performed. If the perturbation of the field by the object can be expressed as spectrum modification then the two spectra can simply be multiplied and the resulting field is given by an inverse Fourier transform. If the perturbed field cannot be expressed in this way, then coherent superposition can be applied to the individual plane wave components. This can be done by letting each of plane wave components to interact with the object before summing up the perturbed plane wave components.

It is emphasised that for a high aperture confocal microscope, there are no universal theories developed yet with which the calculation of these systems could be performed so easily as by Fourier optics.

CHAPTER 5

Extension of Wolf's diffraction theory

"Truth, in science, can be defined as the working hypothesis best fitted to open the way to the next better one."

K. Lorenz

5.1 Introduction

Focussing through materials of mismatched refractive indices has attracted particular interest because of microscopy applications in the biological and material sciences. Microscope objective lenses have long been manufactured with aberration correction for certain but fixed penetration depths, e.g. lenses to be used with a cover glass for biological applications. Microscope objective lenses have recently been produced which provide continuously variable correction for spherical aberration introduced by off-axis propagation of light through materials of different refractive indices. Nevertheless, the importance of investigations directed at a better understanding of high aperture focussing is still great, and this demand is increasing especially with new material science applications.

The first article which dealt with the focussing of electromagnetic waves into mismatched refractive index materials was by Gasper et *al.*[30], who considered an arbitrary electromagnetic wave as it traversed a planar interface. They used the representation of angular spectrum of plane waves and considered the focussing problem for isotropic homogeneous media*. They derived a rigorous solution of the problem and its asymptotic approximation and gave expressions for the electric and magnetic fields. However, due to the complexity of their theory, the use of their formulae for the computation of the electromagnetic field near focus was not practical in most cases. Ling and Lee[58] treated

*For a treatment of diffraction in anisotropic materials the reader is referred to Stamnes and Sherman[93,94]

the focussing of electromagnetic waves through an interface. A boundary condition in the form of a current distribution was used as the starting point and the representation of angular spectrum of plane waves was applied. In a semi-geometrical approach, using the stationary phase method, expressions were obtained for the electric and magnetic fields. However, Ling and Lee used approximations to obtain their integral formulae. A comprehensive treatment of different focussing theories was later given by Stamnes[95].

A recent article by Hell et *al.*[43] considered the focussing problem for mismatched refractive index materials using the Fresnel-Kirchhoff integral. They decomposed the incident electric vector into *s*– and *p*–polarised parts and also calculated the effect of spherical aberration on the image formation for a confocal fluorescence microscope. However, the integral formula that they used is derived from Green's theorem, which requires the continuity of the field and its first and second derivatives within and on the boundary of the area of integration, and this is not the case for the normal component of the electric field and the tangential component of the magnetic field, and so the final integral formulae obtained may not be rigorously correct.

The purpose of this Chapter is to extend Wolf's treatment of the diffraction problem (see Chapter 4) for the case when an electromagnetic wave is focussed in a single medium of propagation to the case when light is focussed by a high aperture system into a bulk specimen which has a refractive index different from the medium of propagation and introduces a considerable amount of spherical aberration. The starting point of the present treatment is Wolf's integral formulae with which the electromagnetic field is derived just before the interface between the two media. The field is then traversed across the interface by applying the Fresnel refraction law to the individual plane waves incident on the interface. The field so derived just after the interface is used as the boundary condition for a second set of integral formulae corresponding to a superposition of plane waves, which represent the field inside the second medium. In this way the diffraction problem is solved in a rigorous mathematical manner and the solution satisfies the homogeneous wave equation. Formulae are also obtained for the strength vectors in the second material in a form more generally applicable than those previously published.

In the present calculations the spherical aberration is introduced by the specimen rather than an aberration function related to the lens. This is because modern high performance lenses are generally well corrected for spherical aberration. Therefore, in considering a real optical system, the assumption of a perfect spherical wave emerging from the lens and converging towards the Gaussian focus point (with no spherical aberration) seems to be justified. Although strong spherical aberrations can be treated by geometrical optics successfully describing the main features of the focal region (e.g. caustic, circle of least confusion), the fine structure and the electromagnetic properties (e.g. polarisation, energy flow) of the field can only be described by a full electromagnetic treatment. This becomes increasingly important when the field so calculated is to be applied to wave scattering by small objects situated close to the focus and whose size is comparable to the wavelength, e.g. in biological and material science applications.

The presentation of this Chapter is as follows: in section 5.2 the integral representation of the electromagnetic field in the image space is derived. In section 5.3 the electric and magnetic vectors are decomposed as each traverses the optical system. In section 5.4 the above results are used and simplified formulae are given for the electric and magnetic fields in the second material. Finally, in sections 5.5 and 5.6 the physical interpretation of the formulae is considered and a paraxial approximation of the solution is given.

5.2 The integral representation

Consider an optical system of revolution with an optical axis z as shown in Fig. 5.1. This system images a point source, which is situated in the object space at $z = -\infty$ and radiates a linearly polarised monochromatic and coherent electromagnetic wave. This wave is incident on a lens of aperture Σ which produces a convergent spherical wave in the image space. The origin O of the (x, y, z) co-ordinate system is positioned at the Gaussian focus. The electric and magnetic fields are determined at the arbitrary point P in the focal region. The aperture size and the distance of P from the aperture are taken to be large compared with the wavelength. In Fig. 5.1 and the following, $\mathbf{s} = (s_x, s_y, s_z)$ is the unit vector along a typical ray in the image space and $\mathbf{r}_p = (x, y, z)$ is the position

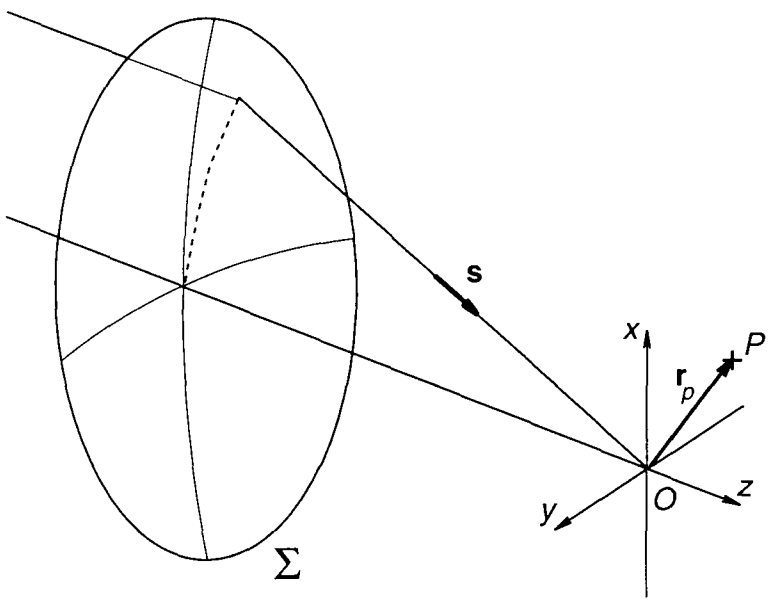


Figure 5.1: Diagram showing light focussed by a lens into a single medium.

vector pointing from O to P.

Let $\tilde{\mathbf{E}}(P,t)$ indicate the time-dependent electric field and $\tilde{\mathbf{H}}(P,t)$ the time-dependent magnetic field at P at time t, so that:

$$\left.\begin{aligned}\tilde{\mathbf{E}}(P,t) &= \Re\{\mathbf{E}(P)\exp(-\mathrm{i}\omega t)\}\\ \tilde{\mathbf{H}}(P,t) &= \Re\{\mathbf{H}(P)\exp(-\mathrm{i}\omega t)\}\end{aligned}\right\} \tag{5.1}$$

where $\Re$ indicates the real part.

In a homogeneous image space, the time-independent electric and magnetic fields can be represented as a superposition of plane waves[25] and we use the form developed by Wolf[115]:

$$\mathbf{E}(P) = -\frac{\mathrm{i}k}{2\pi}\int\int_{\Omega}\frac{\mathbf{a}(s_x,s_y)}{s_z}\exp\{\mathrm{i}k[\Phi(s_x,s_y)+\mathbf{s}\cdot\mathbf{r}_p]\}\mathrm{d}s_x\mathrm{d}s_y \tag{5.2}$$

and

$$\mathbf{H}(P) = -\frac{\mathrm{i}k}{2\pi}\int\int_{\Omega}\frac{\mathbf{b}(s_x,s_y)}{s_z}\exp\{\mathrm{i}k[\Phi(s_x,s_y)+\mathbf{s}\cdot\mathbf{r}_p]\}\mathrm{d}s_x\mathrm{d}s_y \tag{5.3}$$

where $\Phi(s_x,s_y)$ is the aberration function (describing the optical path difference between the aberrated and the spherical wavefront along $\mathbf{s}$), $\mathbf{a}$ and $\mathbf{b}$ are the electric and magnetic strength vectors respectively of the unperturbed electric and magnetic fields in the exit aperture Σ, k is the wavenumber, and Ω is the solid angle formed by all the geometrical optics rays (and which is therefore a limit for all the $\mathbf{s}$ unit ray vectors).

Before we proceed with the formulation it is worth mentioning that Eqs. (5.2) and (5.3) represent a summation of the plane waves (unlike the Fresnel-Kirchhoff integral)

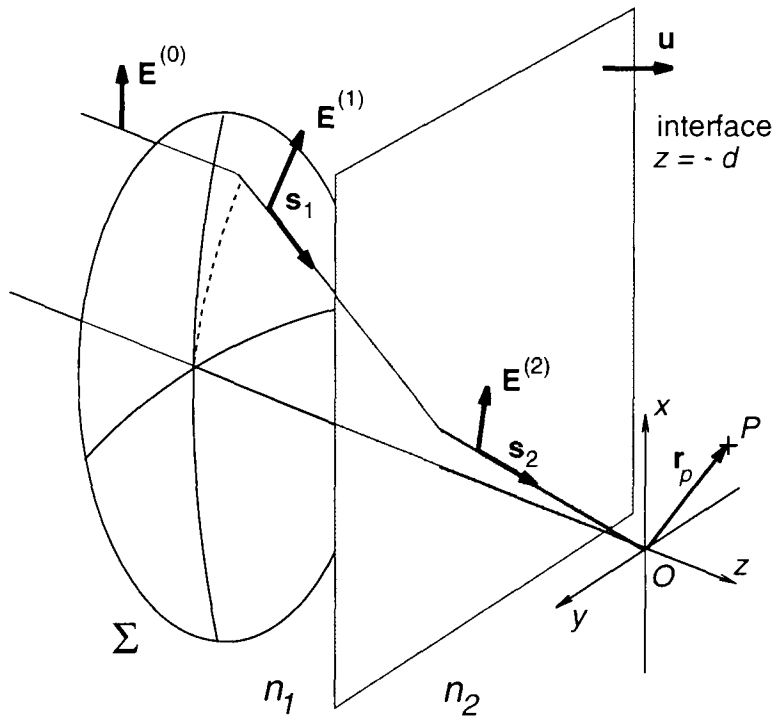


Figure 5.2: Diagram showing light focussed by a lens into two media separated by a planar interface.

that are leaving the aperture. It is also evident that the electric and magnetic fields, as represented by Eqs. (5.2) and (5.3), do not depend upon the particular wavefront within the solid angle Ω over which the integration is performed. This statement can be proved in a rigorous mathematical manner[60]. The above equations also show that the phase factor (apart from the aberration function) is the scalar product of vectors $\mathbf{s}$ and $\mathbf{r}_p$. It follows from the above that the phase factor expresses the optical path difference between wavefronts going through the point P and the Gaussian focus O, unlike the Fresnel-Kirchhoff integral where the phase factor is proportional to the full optical path from the aperture to P.

In this section from now on we shall not present the derivation of the formulae corresponding to the magnetic field because, apart from the strength vectors, Eqs. (5.2) and (5.3) are in the same form.

The case corresponding to Fig. 5.1, but for an image space consisting of materials 1 and 2 with refractive indices n_1 and n_2 respectively separated by a planar interface perpendicular to the optical axis, is shown in Fig. 5.2. The origin O is again positioned at the Gaussian focus. We reformulate Eq. (5.2) as follows.

In material 1 and at the interface ($z = -d$), the incident electric field is given by:

$$\mathbf{E}_1(x, y, -d) = -\frac{ik_1}{2\pi} \int\int_{\Omega_1} \mathbf{W}^{(e)}(\mathbf{s}_1) \exp[ik_1(s_{1x}x + s_{1y}y - s_{1z}d)] \mathrm{d}s_{1x}\mathrm{d}s_{1y} \qquad (5.4)$$

where subscripts 1 and 2 denote values corresponding to regions with materials 1 and 2 respectively, the objective lens is taken to be aberration free ($\Phi(s_{1x}, s_{1y}) = 0$) and

$$\mathbf{W}^{(e)}(\mathbf{s}_1) = \frac{\mathbf{a}(s_{1x}, s_{1y})}{s_{1z}}. \tag{5.5}$$

To describe the field in the second material we assume that each plane wave component refracting at the interface obeys Fresnel's refraction law and the resulting field is constructed as a superposition of refracted plane waves. If the amplitude of the plane waves incident on the interface is described by $\mathbf{W}^{(e)}$, then the amplitude of the transmitted plane waves in the second material is a linear function of $\mathbf{W}^{(e)}$, i.e.

$$\underline{\underline{T}}^{(e)}\,\mathbf{W}^{(e)}, \tag{5.6}$$

where the tensor operator $\underline{\underline{T}}^{(e)}$ is a function of the angle of incidence and n_1 and n_2. The transmitted field in the second material, at the close vicinity ($z = -d + \delta$) of the interface is given by

$$\mathbf{E}_2(x, y, -d) = -\frac{ik_1}{2\pi} \int\int_{\Omega_1} \underline{\underline{T}}^{(e)}\,\mathbf{W}^{(e)}(\mathbf{s}_1) \exp[ik_1(s_{1x}x + s_{1y}y - s_{1z}d)] \mathrm{d}s_{1x}\mathrm{d}s_{1y} \tag{5.7}$$

when $\delta \to 0$. We represent the field inside the second material again as a superposition of plane waves. This representation is a solution of the time-independent wave equation and can be written as:

$$\mathbf{E}_2(\mathbf{r}_p) = -\frac{ik_2}{2\pi} \int\int_{\Omega_2} \mathbf{F}^{(e)}(\mathbf{s}_2) \exp[ik_2\mathbf{s}_2 \cdot \mathbf{r}_p] \mathrm{d}s_{2x}\mathrm{d}s_{2y}. \tag{5.8}$$

We have to determine the function $\mathbf{F}^{(e)}(\mathbf{s}_2)$ and for this we shall make use of Eq. (5.7), which represents a boundary condition for Eq. (5.8). Furthermore we shall establish the relationship between $\mathbf{s}_1$ and $\mathbf{s}_2$.

It is evident from the vectorial law of refraction:

$$k_2\mathbf{s}_2 - k_1\mathbf{s}_1 = (k_2 \cos\varphi_2 - k_1 \cos\varphi_1)\mathbf{u} \tag{5.9}$$

where $\mathbf{u}$ represents the normal of the interface, $\varphi_1 = (\mathbf{s}_1, \mathbf{u})$ and $\varphi_2 = (\mathbf{s}_2, \mathbf{u})$, that:

$$k_2 s_{2x} = k_1 s_{1x} \qquad \text{and} \qquad k_2 s_{2y} = k_1 s_{1y}. \tag{5.10}$$

In the following we shall give a solution for Eq. (5.8) with boundary condition Eq. (5.7). As a result of a co-ordinate transformation and setting $\mathbf{s}_2 = f(\mathbf{s}_1)$, Eq. (5.8) yields:

$$\mathbf{E}_2(\mathbf{r}_p) = -\frac{\mathrm{i}k_2}{2\pi}\int\int_{\Omega_1}\mathbf{F}^{(e)}(\mathbf{s}_2)\exp[\mathrm{i}k_2\mathbf{s}_2\cdot\mathbf{r}_p] \times \mathrm{J}_0(s_{1x},s_{1y};s_{2x},s_{2y})\mathrm{d}s_{1x}\mathrm{d}s_{1y} \tag{5.11}$$

where J_0 is the Jacobian of the co-ordinate transformation:

$$\mathrm{J}_0 = \left(\frac{k_1}{k_2}\right)^2$$

obtained by using Eq. (5.10). Equation (5.11) satisfies the boundary condition represented by Eq. (5.7) when

$$\mathbf{F}^{(e)}(\mathbf{s}_2)\exp[\mathrm{i}k_2(s_{2x}x+s_{2y}y-s_{2z}d)]\left(\frac{k_1}{k_2}\right) = \underline{\underline{\mathrm{T}}}^{(e)}\ \frac{\mathbf{a}(s_{1x},s_{1y})}{s_{1z}}\exp[\mathrm{i}k_1(s_{1x}x+s_{1y}y-s_{1z}d)].$$

The above expression with the help of Eq. (5.10) yields:

$$\mathbf{F}^{(e)}(\mathbf{s}_1,\mathbf{s}_2) = \left(\frac{k_2}{k_1}\right)\underline{\underline{\mathrm{T}}}^{(e)}\ \frac{\mathbf{a}(s_{1x},s_{1y})}{s_{1z}}\exp[-\mathrm{i}d\ (k_1s_{1z}-k_2s_{2z})]. \tag{5.12}$$

On substituting from Eq. (5.12) into Eq. (5.11) we obtain the electric field in the second material:

$$\mathbf{E}_2(x,y,z) = -\frac{\mathrm{i}k_2^2}{2\pi k_1}\int\int_{\Omega_1}\underline{\underline{\mathrm{T}}}^{(e)}\ \frac{\mathbf{a}(s_{1x},s_{1y})}{s_{1z}}\exp[-\mathrm{i}d\ (k_1s_{1z}-k_2s_{2z})] \times \exp(\mathrm{i}k_2s_{2z}z)\exp[\mathrm{i}k_1(s_{1x}x+s_{1y}y)]\mathrm{d}s_{1x}\mathrm{d}s_{1y}. \tag{5.13}$$

The same procedure yields a similar expression for the magnetic vector:

$$\mathbf{H}_2(x,y,z) = -\frac{\mathrm{i}k_2^2}{2\pi k_1}\int\int_{\Omega_1}\underline{\underline{\mathrm{T}}}^{(m)}\ \frac{\mathbf{b}(s_{1x},s_{1y})}{s_{1z}}\exp[-\mathrm{i}d\ (k_1s_{1z}-k_2s_{2z})] \times \exp(\mathrm{i}k_2s_{2z}z)\exp[\mathrm{i}k_1(s_{1x}x+s_{1y}y)]\mathrm{d}s_{1x}\mathrm{d}s_{1y}. \tag{5.14}$$

From the above it also follows that in Eqs. (5.13) and (5.14)

$$s_{2z} = \left(1-\frac{n_1^2}{n_2^2}+\frac{n_1^2}{n_2^2}s_{1z}^2\right)^{1/2} = \left[1-\frac{n_1^2}{n_2^2}(s_{1x}^2+s_{1y}^2)\right]^{1/2}. \tag{5.15}$$

It is important to emphasize that since both the boundary condition represented by Eq. (5.7) and the integral representation Eq. (5.8) are exact solutions of the homogeneous wave equation, our formulae for the electric vector Eq. (5.13) and magnetic vector Eq.

(5.14) in the second material also satisfy the homogeneous wave equation. Therefore we have successfully obtained a consistent extension of Wolf's solution in the second material. We shall show in the last section that the paraxial approximation of Eq. (5.13) yields the well known axial distribution[14] for one material and it also correctly predicts the axial position of the Gaussian focus for two materials.

5.3 The electric and magnetic strength vectors

Determination of the electric and magnetic strength vectors for a single medium of propagation was described previously[75]. Now we obtain these vectors for a plane polarised wave incident on the lens and a material whose refractive index is different from the medium of propagation is placed into the image space. The material is taken to be isotropic and homogeneous and to have an optically smooth planar surface which is perpendicular to the optical axis. For our decomposition, the usual assumptions are made, namely, that the electric vector maintains its direction with respect to a meridional plane and the electric vector remains on the same side of a meridional plane on passing through the system.

For the optical system under consideration, the angle of incidence at the interface is denoted by φ_1 and the angle of refraction by φ_2. The unit vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ and the vector $\mathbf{r}_p$ (Fig. 5.2) are given in spherical polar co-ordinates by:

$$\mathbf{s}_1 = \sin\varphi_1 \cos\vartheta\ \mathbf{i} + \sin\varphi_1 \sin\vartheta\ \mathbf{j} + \cos\varphi_1\ \mathbf{k}, \tag{5.16}$$

$$\mathbf{s}_2 = \sin\varphi_2 \cos\vartheta\ \mathbf{i} + \sin\varphi_2 \sin\vartheta\ \mathbf{j} + \cos\varphi_2\ \mathbf{k} \tag{5.17}$$

and

$$\mathbf{r}_p = r_p(\sin\varphi_p \cos\vartheta_p\ \mathbf{i} + \sin\varphi_p \sin\vartheta_p\ \mathbf{j} + \cos\varphi_p\ \mathbf{k}). \tag{5.18}$$

where $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ are the unit base vectors of the (x, y, z) orthogonal system and the spherical polar co-ordinates r, φ and ϑ are defined so that $r > 0$, $0 \leq \varphi < \pi$ and $0 \leq \vartheta < 2\pi$. The co-ordinate system is chosen so that the y component of the incident electric vector is zero. For the incident electric vector $\mathbf{E}^{(0)}$ in front of the lens:

$$\mathbf{E}^{(0)} = \begin{pmatrix} E_0 \\ 0 \\ 0 \end{pmatrix}.$$

For a treatment of the refraction which occurs at the interface of the mismatched materials, it is convenient to decompose the electric vector into s– and p–polarised vector components, E_s and E_p respectively, and to rotate the co-ordinate system so that the new co-ordinate system will contain components in the (p, s, ζ) system. This co-ordinate system is defined in such a way that $E_\zeta = 0$. The electric vector components $\mathbf{E}^{(1)}_{(p,s,\zeta)}$ after the lens are then in the following form:

$$\mathbf{E}^{(1)}_{(p,s,\zeta)} = A(\varphi_1)\underline{\underline{\mathrm{P}}}^{(1)}\underline{\underline{\mathrm{L}}}\ \underline{\underline{\mathrm{R}}}\ \mathbf{E}^{(0)}_{(x,y,z)} \tag{5.19}$$

where $A(\varphi_1)$ is an amplitude function (defined below), matrix $\underline{\underline{\mathrm{R}}}$ describes the co-ordinate transformation for rotation around the z axis:

$$\underline{\underline{\mathrm{R}}} = \begin{pmatrix} \cos\vartheta & \sin\vartheta & 0 \\ -\sin\vartheta & \cos\vartheta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{5.20}$$

matrix $\underline{\underline{\mathrm{L}}}$ describes the changes in the electric field as it traverses the lens:

$$\underline{\underline{\mathrm{L}}} = \begin{pmatrix} \cos\varphi_1 & 0 & \sin\varphi_1 \\ 0 & 1 & 0 \\ -\sin\varphi_1 & 0 & \cos\varphi_1 \end{pmatrix}, \tag{5.21}$$

and matrix $\underline{\underline{\mathrm{P}}}^{(1)}$ describes the co-ordinate system rotation which generates E_s and E_p components with $E_\zeta = 0$:

$$\underline{\underline{\mathrm{P}}}^{(1)} = \begin{pmatrix} \cos\varphi_1 & 0 & -\sin\varphi_1 \\ 0 & 1 & 0 \\ \sin\varphi_1 & 0 & \cos\varphi_1 \end{pmatrix}. \tag{5.22}$$

It is interesting to note that in our case, when $E_z^{(0)} = 0$, the operations by $\underline{\underline{\mathrm{L}}}$ and $\underline{\underline{\mathrm{P}}}^{(1)}$ cancel out.

The electric field in the second material $\mathbf{E}^{(2)}$ is given by:

$$\mathbf{E}^{(2)}_{(x,y,z)} = [\underline{\underline{\mathrm{R}}}]^{-1}[\underline{\underline{\mathrm{P}}}^{(2)}]^{-1}\underline{\underline{\mathrm{I}}}\ \mathbf{E}^{(1)}_{(p,s,\zeta)} \tag{5.23}$$

where matrix $\underline{\underline{\mathrm{I}}}$ describes the effect of the interface:

$$\underline{\underline{\mathrm{I}}} = \begin{pmatrix} \tau_p & 0 & 0 \\ 0 & \tau_s & 0 \\ 0 & 0 & \tau_p \end{pmatrix}, \tag{5.24}$$

and τ_p and τ_s are the Fresnel coefficients:

$$\tau_s = \frac{2\sin\varphi_2\cos\varphi_1}{\sin(\varphi_1+\varphi_2)} \qquad \tau_p = \frac{2\sin\varphi_2\cos\varphi_1}{\sin(\varphi_1+\varphi_2)\cos(\varphi_1-\varphi_2)}, \tag{5.25}$$

matrix $[\underline{\underline{\mathrm{P}}}^{(2)}]^{-1}$ describes the rotation of the co-ordinate system required to return it to the (p,s,z) system,

$$[\underline{\underline{\mathrm{P}}}^{(2)}]^{-1} = \begin{pmatrix} \cos\varphi_2 & 0 & \sin\varphi_2 \\ 0 & 1 & 0 \\ -\sin\varphi_2 & 0 & \cos\varphi_2 \end{pmatrix}, \tag{5.26}$$

and matrix $\underline{\underline{\mathrm{R}}}^{-1}$ describes the inverse rotation around the z axis.

From Eqs. (5.19), (5.20), (5.21) and (5.23) we obtain the components of the electric vector after the lens in the (p,s,ζ) system:

$$\mathbf{E}^{(1)}_{(p,s,\zeta)} = E_0 \begin{pmatrix} \cos\vartheta \\ -\sin\vartheta \\ 0 \end{pmatrix}. \tag{5.27}$$

Finally Eq. (5.23) yields in the second material on setting $E_0 = 1$:

$$\mathbf{E}^{(2)}_{(x,y,z)} = A(\varphi_1) \begin{pmatrix} \tau_p\cos\varphi_2\cos^2\vartheta + \tau_s\sin^2\vartheta \\ \tau_p\cos\varphi_2\sin\vartheta\cos\vartheta - \tau_s\sin\vartheta\cos\vartheta \\ -\tau_p\sin\varphi_2\cos\vartheta \end{pmatrix} \tag{5.28}$$

Function $A(\varphi_1)$ can be regarded as an apodization function which depends on the lens used in imaging. Richards and Wolf[75] showed that when the system obeys Abbe's sine condition, i.e. is aplanatic, then

$$A(\varphi_1) = f l_0 \cos^{1/2}\varphi_1, \tag{5.29}$$

where f is the focal length of the lens and l_0 is an amplitude factor. Otherwise this function can take different forms†.

Having derived the electric vector $\mathbf{E}^{(2)}_{(x,y,z)}$, the electric strength vector $\mathbf{c}$ can be written:

$$\mathbf{c} = \underline{\underline{\mathrm{T}}}^{(e)}\,\mathbf{a} = \mathbf{E}^{(2)} \tag{5.30}$$

The magnetic strength vector $\mathbf{d}$ is then given by:

$$\mathbf{d} = \sqrt{\frac{\varepsilon_2}{\mu_2}}\,\mathbf{s}_2 \times \mathbf{c} \tag{5.31}$$

†For the expression of this function when the lens is not aplanatic, the reader is referred to Stamnes[95].

where μ_2 is the permeability and ε_2 is the permittivity of the second material. The magnetic strength vector from Eqs. (5.28), (5.30), (5.31) and (5.17) is given by:

$$\mathbf{d} = \sqrt{\frac{\varepsilon_2}{\mu_2}} A(\varphi_1) \begin{pmatrix} -\tau_p \cos\vartheta \sin\vartheta + \tau_s \sin\vartheta \cos\vartheta \cos\varphi_2 \\ \tau_p \cos^2\vartheta + \tau_s \sin^2\vartheta \cos\varphi_2 \\ -\tau_s \sin\vartheta \cos\varphi_2 \end{pmatrix} \tag{5.32}$$

We note that for the special case when $\varphi_1 = \varphi_2$, $\varepsilon_1 = \varepsilon_2 = 1$ and $\mu_1 = \mu_2 = 1$[‡], Eqs. (5.28) and (5.32) reduce to the electric strength vector given by Richards and Wolf[75], as they should.

5.4 The electric and magnetic field vectors

First we formulate the expressions needed to simplify Eqs. (5.13) and (5.14). We transform the integral variables $\mathrm{d}s_{1x}, \mathrm{d}s_{1y}$ to the spherical polar co-ordinate system. The Jacobian of the transformation is:

$$\mathsf{J}_p = \sin\varphi_1 \cos\varphi_1$$

thus

$$\mathrm{d}s_{1x}\mathrm{d}s_{1y} = \sin\varphi_1 \cos\varphi_1 \mathrm{d}\varphi_1 \mathrm{d}\vartheta. \tag{5.33}$$

We define

$$\kappa = n_1 \sin\varphi_1 \sin\varphi_p \cos(\vartheta - \vartheta_p) + n_2 \cos\varphi_2 \cos\varphi_p \tag{5.34}$$

and

$$\Psi(\varphi_1, \varphi_2, -d) = -d(n_1 \cos\varphi_1 - n_2 \cos\varphi_2). \tag{5.35}$$

So far we have constructed the quantities required to determine the electric and magnetic vectors. With the help of the above results Eqs. (5.13) and (5.14) can be rewritten to express the electric $\mathbf{E}_2$ and magnetic $\mathbf{H}_2$ vectors inside the second material:

$$\mathbf{E}_2(\mathbf{r}_p, -d) = -\frac{\mathrm{i}k_2^2}{2\pi k_1} \int\int_{\Omega_1} \mathbf{c}(\varphi_1, \varphi_2, \vartheta) \exp\{\mathrm{i}k_0[r_p\kappa + \Psi(\varphi_1, \varphi_2, -d)]\} \sin\varphi_1 \mathrm{d}\varphi_1 \mathrm{d}\vartheta \tag{5.36}$$

and

$$\mathbf{H}_2(\mathbf{r}_p, -d) = -\frac{\mathrm{i}k_2^2}{2\pi k_1} \int\int_{\Omega_1} \mathbf{d}(\varphi_1, \varphi_2, \vartheta) \exp\{\mathrm{i}k_0[r_p\kappa + \Psi(\varphi_1, \varphi_2, -d)]\} \sin\varphi_1 \mathrm{d}\varphi_1 \mathrm{d}\vartheta \tag{5.37}$$

[‡]In the C.G.S. unit system

where k_0 indicates the wavenumber *in vacuo*. On substituting from Eqs. (5.32), (5.30), (5.28) and (5.33) into Eqs. (5.36) and (5.37), changing the integration limits and assuming that the system obeys the sine condition Eq. (5.29), we obtain the following expressions for the electric field components:

$$e_{2x} = +\frac{\mathrm{i}K}{2\pi}\int_0^{\alpha}\int_0^{2\pi}\sqrt{\cos\varphi_1}\sin\varphi_1[(\tau_p\cos\varphi_2+\tau_s)+\cos 2\vartheta(\tau_p\cos\varphi_2-\tau_s)] \times\exp\{\mathrm{i}k_0[r_p\kappa+\Psi(\varphi_1,\varphi_2,-d)]\}\mathrm{d}\varphi_1\mathrm{d}\vartheta \tag{5.38a}$$

$$e_{2y} = +\frac{\mathrm{i}K}{2\pi}\int_0^{\alpha}\int_0^{2\pi}\sqrt{\cos\varphi_1}\sin\varphi_1\sin 2\vartheta(\tau_p\cos\varphi_2-\tau_s) \times\exp\{\mathrm{i}k_0[r_p\kappa+\Psi(\varphi_1,\varphi_2,-d)]\}\mathrm{d}\varphi_1\mathrm{d}\vartheta \tag{5.38b}$$

$$e_{2z} = -\frac{\mathrm{i}K}{\pi}\int_0^{\alpha}\int_0^{2\pi}\sqrt{\cos\varphi_1}\sin\varphi_1\tau_p\sin\varphi_2\cos\vartheta \times\exp\{\mathrm{i}k_0[r_p\kappa+\Psi(\varphi_1,\varphi_2,-d)]\}\mathrm{d}\varphi_1\mathrm{d}\vartheta \tag{5.38c}$$

where

$$K=\frac{k_2^2 f l_0}{2k_1}=\frac{\pi n_2^2 f l_0}{\lambda n_1}.$$

We also obtain the following expressions for the magnetic field components:

$$h_{2x} = +\frac{\mathrm{i}Kn_2}{2\pi}\int_0^{\alpha}\int_0^{2\pi}\sqrt{\cos\varphi_1}\sin\varphi_1\sin 2\vartheta(\tau_s\cos\varphi_2-\tau_p) \times\exp\{\mathrm{i}k_0[r_p\kappa+\Psi(\varphi_1,\varphi_2,-d)]\}\mathrm{d}\varphi_1\mathrm{d}\vartheta \tag{5.39a}$$

$$h_{2y} = +\frac{\mathrm{i}Kn_2}{2\pi}\int_0^{\alpha}\int_0^{2\pi}\sqrt{\cos\varphi_1}\sin\varphi_1[(\tau_p+\tau_s\cos\varphi_2)+(\tau_p+\tau_s\cos\varphi_2)\cos 2\vartheta)] \times\exp\{\mathrm{i}k_0[r_p\kappa+\Psi(\varphi_1,\varphi_2,-d)]\}\mathrm{d}\varphi_1\mathrm{d}\vartheta \tag{5.39b}$$

$$h_{2z} = -\frac{\mathrm{i}Kn_2}{\pi}\int_0^{\alpha}\int_0^{2\pi}\sqrt{\cos\varphi_1}\sin\varphi_1\tau_s\sin\varphi_2\sin\vartheta \times\exp\{\mathrm{i}k_0[r_p\kappa+\Psi(\varphi_1,\varphi_2,-d)]\}\mathrm{d}\varphi_1\mathrm{d}\vartheta \tag{5.39c}$$

where we assumed that $\mu_2 = 1$ and set α to be the angular semi-aperture of the lens. The integration in Eqs. (5.38) and (5.39) can be carried out[75] with respect to ϑ and from the

result the electric and magnetic field components can be expressed as the combination of two sets of three integrals I_0, I_1 and I_2:

$$e_{2x} = -\mathrm{i}K(I_0^{(e)} + I_2^{(e)} \cos 2\vartheta_p) \tag{5.40a}$$

$$e_{2y} = -\mathrm{i}K I_2^{(e)} \sin 2\vartheta_p \tag{5.40b}$$

$$e_{2z} = -2K I_1^{(e)} \cos \vartheta_p \tag{5.40c}$$

and

$$h_{2x} = -\mathrm{i}K n_2 I_2^{(h)} \sin 2\vartheta_p \tag{5.41a}$$

$$h_{2y} = -\mathrm{i}K n_2 (I_0^{(h)} - I_2^{(h)} \cos 2\vartheta_p) \tag{5.41b}$$

$$h_{2z} = -2K n_2 I_1^{(h)} \sin \vartheta_p \tag{5.41c}$$

where ϑ_p is defined in Eq. (5.18) and we have put $n_2 = \sqrt{\varepsilon_2}$. After substituting the normalised optical co-ordinates

$$\left.\begin{aligned} v &= k_1 (x^2 + y^2)^{\frac{1}{2}} \sin \alpha = k_1 r_p \sin \varphi_p \sin \alpha \\ u &= k_2 z \sin^2 \alpha = k_2 r_p \cos \varphi_p \sin^2 \alpha, \end{aligned}\right\} \tag{5.42}$$

into Eqs. (5.38) and (5.39), the integrals I_0, I_1 and I_2 are given by:

$$\left.\begin{aligned} I_0^{(e)} &= \int_0^\alpha \sqrt{\cos \varphi_1} \sin \varphi_1 \exp\left[\mathrm{i}k_0 \Psi(\varphi_1, \varphi_2, -d)\right](\tau_s + \tau_p \cos \varphi_2) \\ &\qquad \times J_0\left(\frac{v \sin \varphi_1}{\sin \alpha}\right) \exp\left(\frac{\mathrm{i}u \cos \varphi_2}{\sin^2 \alpha}\right) \mathrm{d}\varphi_1 \\ I_1^{(e)} &= \int_0^\alpha \sqrt{\cos \varphi_1} \sin \varphi_1 \exp\left[\mathrm{i}k_0 \Psi(\varphi_1, \varphi_2, -d)\right] \tau_p \sin \varphi_2 \\ &\qquad \times J_1\left(\frac{v \sin \varphi_1}{\sin \alpha}\right) \exp\left(\frac{\mathrm{i}u \cos \varphi_2}{\sin^2 \alpha}\right) \mathrm{d}\varphi_1 \\ I_2^{(e)} &= \int_0^\alpha \sqrt{\cos \varphi_1} \sin \varphi_1 \exp\left[\mathrm{i}k_0 \Psi(\varphi_1, \varphi_2, -d)\right](\tau_s - \tau_p \cos \varphi_2) \\ &\qquad \times J_2\left(\frac{v \sin \varphi_1}{\sin \alpha}\right) \exp\left(\frac{\mathrm{i}u \cos \varphi_2}{\sin^2 \alpha}\right) \mathrm{d}\varphi_1 \end{aligned}\right\} \tag{5.43}$$

and

$$\left.\begin{aligned}
I_0^{(h)} &= \int_0^{\alpha} \sqrt{\cos\varphi_1}\,\sin\varphi_1 \exp\left[ik_0\Psi(\varphi_1,\varphi_2,-d)\right](\tau_p+\tau_s\cos\varphi_2) \\
&\qquad \times J_0\left(\frac{v\sin\varphi_1}{\sin\alpha}\right)\exp\left(\frac{iu\cos\varphi_2}{\sin^2\alpha}\right)\mathrm{d}\varphi_1 \\
I_1^{(h)} &= \int_0^{\alpha} \sqrt{\cos\varphi_1}\,\sin\varphi_1 \exp\left[ik_0\Psi(\varphi_1,\varphi_2,-d)\right]\tau_s\sin\varphi_2 \\
&\qquad \times J_1\left(\frac{v\sin\varphi_1}{\sin\alpha}\right)\exp\left(\frac{iu\cos\varphi_2}{\sin^2\alpha}\right)\mathrm{d}\varphi_1 \\
I_2^{(h)} &= \int_0^{\alpha} \sqrt{\cos\varphi_1}\,\sin\varphi_1 \exp\left[ik_0\Psi(\varphi_1,\varphi_2,-d)\right](\tau_p-\tau_s\cos\varphi_2) \\
&\qquad \times J_2\left(\frac{v\sin\varphi_1}{\sin\alpha}\right)\exp\left(\frac{iu\cos\varphi_2}{\sin^2\alpha}\right)\mathrm{d}\varphi_1
\end{aligned}\right\} \tag{5.44}$$

where J_n is the Bessel function of the first kind, order n. Equations (5.40), (5.41), (5.43) and (5.44) conclude our solution of the problem. It should be noted that for $n_1 = n_2 = 1$ (therefore $\varphi_2 = \varphi_1$ and $d = 0$), Eqs. (5.43) and (5.44) reduce to $I_n^{(e)} = I_n^{(h)} = I_n$ which are then identical to the corresponding equations of Richards and Wolf[75]. Therefore, our formulation of the vector diffraction theory with spherical aberration introduced by an interface between two materials of mismatched refractive indices is consistent with previous results derived for the non-aberrated case.

We can see from the above equations that the time-independent electric and magnetic fields cannot be described by only one set of integral functions, but rather by two such sets. This is because the electric and magnetic vectors are orthogonal with respect to each other and the polarisation dependent refraction at the interface acts differently on different electric and magnetic vector directions. It is emphasized, however, that all the laws which describe the transition of these two vectors through the system are essentially the same, and the difference between the two sets of integrals does not originate from the nature of the wave (whether it is an electric or a magnetic wave), but rather depends on whether the wave incident on the interface possesses p – or s – polarisation.

5.5 Physical interpretation

Although the formal solution of our problem has been concluded by Eqs. (5.40) and (5.41) we recall here Eq. (5.13) for an analysis of our results. This latter equation can be

re-written in the following form:

$$\mathbf{E}_2(x,y,z) = K' \int\int_{\Omega_1} \mathbf{W}_0(s'_{1x}, s'_{1y}) \exp[\mathrm{i}2\pi(s'_{1x}x + s'_{1y}y)]\mathrm{d}s'_{1x}\mathrm{d}s'_{1y}, \tag{5.45}$$

where

$$s'_{1x} = \frac{n_1 s_{1x}}{\lambda} \qquad \text{and} \qquad s'_{1y} = \frac{n_1 s_{1y}}{\lambda}, \tag{5.46}$$

and

$$\begin{aligned} \mathbf{W}_0 &= \frac{\mathbf{c}(s_{1x}, s_{1y})}{s_{1z}} \exp[-\mathrm{i}(k_1 s_{1z} - k_2 s_{2z})d] \exp(\mathrm{i}k_2 s_{2z} z) \\ &= \frac{\mathbf{c}(s_{1x}, s_{1y})}{s_{1z}} \exp[-\mathrm{i}k_1 s_{1z} d] \exp[\mathrm{i}k_2 s_{2z}(z+d)] \end{aligned} \tag{5.47}$$

and K' is constant. If we extend the limit of integration in Eq. (5.45) to the range of $[-\infty, \infty]$, the latter equation becomes a Fourier integral. The physical meaning of the above extension is that the integration is performed over the full π solid angle and not only over the area of the aperture. In other words the integral operation is being performed not only in the range where

$$\left(\frac{k_1}{k_2}\right)^2 (s_{1x}^2 + s_{1y}^2) < 1 \tag{5.48}$$

but also outside of this region. Equation (5.48) represents the condition for homogeneous waves. Thus, when the integration limits are extended to $[-\infty, \infty]$, the integral representation also includes evanescent waves into the resulting field. These evanescent waves propagate perpendicular to the optical axis, i.e. very close and parallel to the plane of the aperture. These waves decay exponentially at larger z distances away from the aperture, and therefore do not contribute to the far-field distribution.

We can conclude that the electromagnetic field of the optical system under consideration can be represented in the second material as the Fourier transform of a function $\mathbf{W}_0$, which consists of an amplitude factor and two phase factors as shown by Eq. (5.47). These two phase factors represent the phase at the point of observation (P) with respect to the Gaussian focus (O), i.e. one of these phase factors represents the phase $\exp[\mathrm{i}k_2 s_{2z}(z+d)]$ as the wave propagates in the second material, the other one represents the phase $\exp(-\mathrm{i}k_1 s_{1z} d)$ that the wave would have had in the absence of the second material. The sign of the latter phase component is negative and so the total of the two phases gives the phase difference

with respect to the Gaussian focus.

There is an alternative way to write Eq. (5.45):

$$\mathbf{E}_2(x,y,z) = K' \int\int_{\Omega_1} \mathbf{W}_0'(s_{1x}', s_{1y}') \exp(\mathrm{i}k_2 s_{2z} z) \exp[\mathrm{i}2\pi(s_{1x}' x + s_{1y}' y)] \mathrm{d}s_{1x}' \mathrm{d}s_{1y}', \quad (5.49)$$

where

$$\mathbf{W}_0' = \frac{\mathbf{W}_0}{\exp(\mathrm{i}k_2 s_{2z} z)}.$$

Equation (5.49) is formally identical to the angular spectrum representation given by Goodman[32]. Thus we can conclude that our extension of the Debye integral for the optical system under consideration can also be regarded as the angular spectrum representation. The spectrum is given as an amplitude factor describing the electric (and also the magnetic) field as it traverses the system and a phase factor corresponding to the phase difference between the ray vectors along the z direction (which latter phase factor, apart from the distance d, is often referred to as the astigmatic constant).

It should also be mentioned that Eq. (5.36) is in a form identical to the corresponding equation of Wolf[115], i.e. the aberration function is represented in an *exact* form by $\Psi(\varphi_1, -d)$, as defined by Eq. (5.35). Since this function is dependent only on the azimuthal angle φ_1, it describes spherical aberration.

5.6 Paraxial approximation

Equation (5.13) can be approximated paraxially so as to correspond to the axial light distribution for a low aperture objective lens. For this we recall the expression for the electric field:

$$\begin{aligned}\mathbf{E}_2(x,y,z) = -\frac{\mathrm{i}k_2^2}{2\pi k_1} \int\int_{\Omega_1} \underline{\underline{T}}^{(e)} \frac{\mathbf{a}(s_{1x}, s_{1y})}{s_{1z}} \exp\left[-\mathrm{i}d\,(k_1 s_{1z} - k_2 s_{2z})\right] \\ \times \exp(\mathrm{i}k_2 s_{2z} z) \exp\left[\mathrm{i}k_1(s_{1x}x + s_{1y}y)\right] \mathrm{d}s_{1x} \mathrm{d}s_{1y}\end{aligned} \quad (5.50)$$

In the paraxial approximaton the strength vector $\underline{\underline{T}}^{(e)}\,\mathbf{a}$ can be considered as constant and therefore it is independent of the integration. Furthermore we can set $1/s_{1z} =$

constant and $x = y = 0$. The z components of the unit ray vectors can be approximated with the first and second terms of their expansion:

$$\left.\begin{aligned} s_{1z} &= \sqrt{1-(s_{1x}^2+s_{1y}^2)} \approx 1 - \frac{s_{1x}^2+s_{1y}^2}{2} = 1 - \frac{\sigma^2}{2} \\ s_{2z} &= \sqrt{1-(s_{2x}^2+s_{2y}^2)} \approx 1 - \frac{s_{2x}^2+s_{2y}^2}{2} = 1 - \left(\frac{n_1}{n_2}\right)^2 \frac{\sigma^2}{2} \end{aligned}\right\} \tag{5.51}$$

After rearranging Eq. (5.50) and performing further calculations, the paraxial approximation of the electric field $E_2^{(par)}$ is derived as:

$$E_2^{(par)}(0,0,z) = -\frac{\mathrm{i}k_2^2}{k_1} B \exp\left\{-\mathrm{i}k_1\left[d\left(1-\frac{n_2}{n_1}\right)+\frac{n_2}{n_1}z\right]\right\} \int_0^\beta \exp(-\frac{1}{2}\mathrm{i}k_1 D\sigma_1^2)\sigma \mathrm{d}\sigma \tag{5.52}$$

where β is the paraxial semi-angle of the aperture, B is a constant and

$$D = \left(1-\frac{n_1}{n_2}\right)d + \frac{n_1}{n_2}z. \tag{5.53}$$

After performing the integration with respect to σ, the intensity is given by:

$$\begin{aligned} I(0,0,z) = &|E_2^{(par)}(0,0,z)|^2 = I_c \left|\frac{\exp\left(-\frac{1}{2}\mathrm{i}k_1 D\beta^2\right)-1}{\frac{1}{4}k_1 D\beta^2}\right|^2 \\ &= 4I_c \frac{\sin^2\left(\frac{1}{4}k_1 D\beta^2\right)}{\left(\frac{1}{4}k_1 D\beta^2\right)^2}. \end{aligned} \tag{5.54}$$

where I_c is a constant. By putting $n_1 = n_2$ in Eqs. (5.53) and (5.54) we obtain the well-known axial intensity distribution for a single material[§]. By putting $D = 0$ in Eq. (5.53) we obtain the location of the paraxial focus, for the case of two materials, which corresponds to that obtained by geometrical optics.

In the following chapter we show how our integral formulae can be solved analytically, and we present results for the numerically computed time-averaged electric energy density as a function of the numerical aperture of the lens and the depth of focussing into the bulk specimen.

[§]If we use the integral formalism, $I_1^{(e)} = I_2^{(e)} = 0$ in the paraxial approximation, $I_0^{(e)}$ yields a form identical to Eq. (5.52).

It is emphasised that our solution for the extension of the Debye integral is an exact solution of the homogeneous wave equation and is therefore valid everywhere except in the immediate vicinity of the aperture.

CHAPTER 6

Analytical solution of the diffraction integrals

"Computers are useless; they can only give answers."

P. Picasso

6.1 Introduction

The aim of the present chapter is to show how the diffraction integrals, Eqs. (5.43) and (5.44), can be evaluated and also to present numerical results on the calculated electric field at different focussing depths. First, we overview those articles that have most extensively used numerical computations for diffraction phenomena occurring at high aperture focussing. We also include in this survey some articles that provided the theoretical background for such numerical computations and which have already been discussed in Chapter 4 in detail.

The basis of some of the numerical results referred below is due to Wolf[115] who solved the diffraction problem, for homogeneous medium of propagation, with the help of the representation of angular spectrum of plane waves, which result is identical to the Debye integral. Richards and Wolf[75] calculated quantities that are characteristic of the focussed electromagnetic field. They computed and presented the time averaged electric energy density distributions in the focal plane of a lens with different angular semi-apertures. They also presented information on the polarisation states in the focal plane. They gave details of line scans taken from electric energy density distributions to compare the fine structure of these. Later Ostenberg and Smith[72] solved Kirchhoff's diffraction integral analytically for axial positions. They presented line scans of the axial electric energy density distribution with different angular semi-apertures and criticised Luneburg's formulation pointing out that his approach was valid only at higher angular semi-apertures. Boivin

and Wolf[10] computed and presented the two dimensional structure of the time-averaged electric energy density in 0 and 45° meridional planes of a high aperture focussing system, where 0° indicates the direction of incident polarisation. They showed that the lateral resolution along the 90° meridional plane is superior over that of the 0° plane. Two years later Boivin et *al.*[11] showed the structure of the energy flow of a high aperture focussing system, in particular at the vicinity of the singularities of the Poynting vector. Moreover they computed numerically the location of these points and gave the values of the first few lateral positions. Wolf's theory was applied to stigmatic lenses by Hardy and Treves[42]. They took into account the curvature of the surface of the lens and the polarisation dependent effect of transmission on this surface and computed contours of time-averaged electric and magnetic energy density with a certain angular semi-aperture. They also showed that both the electric and the magnetic distributions possess elliptical shape in the focal plane. Recently, Kant[48] showed how Wolf's integral formulae can be expanded to a series of Gegenbauer polynomials so forming a rapidly converging series which makes needless the numerical integration of rapidly oscillating functions. Now we turn to discuss those works in which the field distribution in the image space was computed with aberration in the imaging system.

The first significant article is due to Hopkins and Yzuel[45] who used the Kirchhoff theory with lens aberration (measured as the distance between the reference sphere and the aberrated wavefront). They proposed a numerical method to evaluate their integral representation and computed axial intensity line scans for various amounts of weak spherical aberration. They also showed how spherical aberration distorts the distribution and shifts the axial location of the focus. Apart from the presentation of these line scans no attempt was made to calculate the focal shift continuously. The second significant article is due to Ling and Lee[58] who solved an interface diffraction problem using the angular spectrum representation for a system when aberration was introduced by a material whose refractive index was different from that of the medium of propagation. They presented time-averaged electric energy density distributions in the focal plane and also along the 0° meridional plane in the presence of weak spherical aberration. The authors also computed

and presented curves illustrating the focal shift phenomenon and they examined the effect of materials with different refractive indices on intensity maxima. However, Ling and Lee used a semi geometrical optics approach and an approximation to evaluate integrals occurred in their treatment. Recently an article by Kant[49] showed how lens aberrations can be taken into account by Wolf's integral equations. He formulated, computed and presented a solution for the cases of spherical aberration, curvature of field and distortion. The author presented electric energy density distributions when the first two aberrations were present in the best receiving plane and a diagram illustrating the focus shift as a function of angular semi-aperture.

Once the formal solution of the diffraction problem is obtained the next step is to derive analytical formulae to evaluate diffraction integrals which are known as rapidly oscillating functions, and are therefore hard to compute numerically. Methods have been proposed for numerical computation of such integrals, but none of them gave accurate results under extreme conditions (large aberrations or large departures from the Gaussian focus). A survey of these methods was given by Stamnes[95].

In this chapter, applying the method proposed by Kant to our equations, we solve the diffraction problem analytically for the case when the spherical aberration is directly associated with mismatched refractive index materials in the image space. We also expand the aberration function so making easier the numerical calculation.

The validity of the diffraction integral expansion is limited. One should only apply this method when the integrals are not strongly oscillating or serious loss of accuracy could occur. Strong oscillations are present when either the refractive index of the second material or the focussing depth is large. However, the method proposed here is an efficient and fast computational technique when the focussing occurs through, for example, a glass plate whose refractive index compared with silicon is not so different from that of air. We also propose a simple method by which the computation of the individual coefficients (a_s^n and b_s^n) can be simplified by a linearisation of the aberration function. This can be used as an alternative approach to the expansion of the aberration function.

Since in this work the diffraction integrals are strongly oscillating at larger penetra-

tion depths, we used numerical integration offered by the NAG (Numerical Algorithm Group, Oxford) subroutine package which applied an adaptive algorithm for the integration (D01DAF). The use of the NAG subroutine package however is only efficient when powerful computers are available. For this work a cluster of seven Hewlett-Packard Apollo 9000 computers was used and even with these powerful machines individual computations took more than 30 CPU minutes.

Computations for small penetration depths were performed by both methods and no differences were found between the results.

6.2 Evaluation of the diffraction integrals

Numerical computation of formulae like Eqs. (5.43) and (5.44) is often difficult, especially when large numbers are taken in the argument of the Bessel function and the exponential terms. Oscillations caused by these two functions can hardly be followed by numerical integration. The results of time consuming computations can be faulty or unreliable. The aim of the evaluation of the diffraction integrals is to obtain formulae that behave in a more regular manner than those of the original ones. Our diffraction integrals ($I_0^{(e),(h)}$, $I_1^{(e),(h)}$ and $I_2^{(e),(h)}$) are in the form:

$$K_n = \int_0^{\alpha} F(\cos\varphi_1)\exp[\mathrm{i}k_0\Psi(\cos\varphi_1)]J_n\left(\frac{v\sin\varphi_1}{\sin\alpha}\right)\exp\left(\frac{\mathrm{i}u\cos\varphi_2}{\sin^2\alpha}\right)\mathrm{d}\varphi_1$$

The problem with the above expression is that, although we have explicit dependence for $\varphi_1 = f(\varphi_2)$, φ_1 and φ_2 are mixed in the arguments of the functions. Thus the above integral expression is transformed so that the functions are expressed in terms of φ_2. After some calculations we obtain:

$$I_n = \frac{n_2}{n_1}\int_0^{\beta} F_2(\cos\varphi_2)\exp[\mathrm{i}k_0\Psi(\cos\varphi_2)]J_n\left(\frac{v_2\sin\varphi_2}{\sin\beta}\right)\exp\left(\frac{\mathrm{i}u_2\cos\varphi_2}{\sin^2\beta}\right)\mathrm{d}\varphi_2 \qquad (6.1)$$

where

$$\left.\begin{aligned} v_2 &= k_2(x^2+y^2)^{\frac{1}{2}}\sin\beta = k_2 r_p \sin\varphi_p \sin\beta \\ u_2 &= k_2 z\sin^2\beta = k_2 r_p \cos\varphi_p \sin^2\beta \end{aligned}\right\} \qquad (6.2)$$

and

$$F_2(\cos\varphi_2) = \frac{\cos\varphi_2}{\cos\varphi_1}F(\cos\varphi_1)$$

and β is the solid semi-angle corresponding to α.

The range of definition of function $F(\cos\varphi_2)\exp[ik_0\Psi(\cos\varphi_2)]$ can be changed from $[0,\beta]^*$ to the range of $[0,\pi]$ with the definition:

$$f(\varphi_2)\sin^{n+1}\varphi_2 = \begin{cases} F_2(\varphi_2)\exp[ik_0\Psi(\varphi_2)] & \text{when } 0 \le \varphi_2 \le \beta \\ 0 & \text{elsewhere} \end{cases} \tag{6.3}$$

According to Weierstrass' approximation theorem we can expand an at least piecewise continuous function, such as $f(\varphi_2)$, to a uniformly convergent series of orthonormal polynomials (see Appendix A):

$$f(\cos\varphi_2) = \sum_{s=0}^{\infty} a_s \Pi_s(\cos\varphi_2)$$

where $\Pi_s(x)$ is some polynomial. As the form of Eq. (6.1) shows, the most appropriate is if we choose the Gegenbauer polynomials for the expansion:

$$f(\cos\varphi_2) = \sum_{s=0}^{\infty} a_s^{(n)} C_s^{n+\frac{1}{2}}(\cos\varphi_2) \tag{6.4}$$

where $C_s^{n+\frac{1}{2}}(x)$ is the Gegenbauer polynomial of the first kind, order $n+\frac{1}{2}$ whose definition is given by Eq. (A.24). The coefficient $a_s^{(n)}$ can be calculated from the orthonormality condition:

$$a_s^{(n)} = N_s^{(n)} \int_0^{\beta} f(\cos\varphi_2)(1-\cos^2\varphi_2)^n C_s^{n+\frac{1}{2}}(\cos\varphi_2)\sin\varphi_2 \mathrm{d}\varphi_2 \tag{6.5}$$

where N_s^n is the normalisation factor

$$N_s^{(n)} = \frac{2^{2n-1}(2s+2n+1)\Gamma(n+\frac{1}{2})\Gamma(n+\frac{1}{2})\Gamma(s+1)}{\pi\Gamma(s+2n+1)}, \tag{6.6}$$

and $\Gamma(x)$ is the gamma function. With the help of Eqs. (6.3) and (6.4), Eq. (6.1) becomes:

$$I_n = \frac{n_2}{n_1}\sum_{s=0}^{\infty} a_s^{(n)} \int_0^{\pi} C_s^{n+\frac{1}{2}}(\cos\varphi_2)\sin^{n+1}\varphi_2 \; J_n(\omega\sin\varphi_2\sin\psi) \times \exp(i\omega\cos\varphi_2\cos\psi)\mathrm{d}\varphi_2 \tag{6.7}$$

*For consideration of a lens with an annular aperture the lower bound of the interval (and the lower limit of integration) should be replaced by β_a, which is the corresponding semi-angle of the annual radius.

where we defined spherical polar co-ordinates:

$$\left.\begin{aligned} \frac{u_2}{\sin^2\beta} &= \omega\cos\psi \\ \frac{v_2}{\sin\beta} &= \omega\sin\psi \end{aligned}\right\}. \tag{6.8}$$

Equation (6.7) can be evaluated as it was shown by Watson[104]. The resulting expression is:

$$I_n = 2\frac{n_2}{n_1}\sin^n\psi\sum_{s=0}^{\infty} a_s^{(n)}\mathrm{i}^s C_s^{n+\frac{1}{2}}(\cos\psi)\; j_{n+s}(\omega) \tag{6.9}$$

where $j_{n+s}(x)$ is the spherical Bessel function of the first kind, order $n+s$. Equation (6.9) shows that the field can be represented as a combination of Gegenbauer polynomials and spherical Bessel functions. This result is important, since spherical Bessel functions converge to zero rapidly at bigger arguments and also they can be represented as combinations of sine and cosine functions. As it was pointed out by Kant[48] this rapid convergence is the reason why the Fourier expansion is sufficient for this application, i.e. only the dominant terms make a significant contribution to I_n. The recurrence relation of the Gegenbauer polynomial is given by Eq. (A.25). The initial values are[1]:

$$\left.\begin{aligned} C_0^{\frac{1}{2}}(x) &= 1 & C_1^{\frac{1}{2}}(x) &= x \\ C_0^{\frac{3}{2}}(x) &= 1 & C_1^{\frac{3}{2}}(x) &= 3x \\ C_0^{\frac{5}{2}}(x) &= 1 & C_1^{\frac{5}{2}}(x) &= 5x \end{aligned}\right\} \tag{6.10}$$

The recurrence relation for the spherical Bessel function is:

$$j_{k-1}(x) + j_{k+1}(x) = (2k+1)\; x^{-1} j_k(x) \tag{6.11}$$

and the initial values are:

$$\left.\begin{aligned} j_0(x) &= \frac{\sin x}{x} \\ j_1(x) &= \frac{\sin x}{x^2} - \frac{\cos x}{x} \end{aligned}\right\}. \tag{6.12}$$

From the above results the diffraction integrals can be written:

$$\left.\begin{aligned} I_0^{(e)} &= 2\frac{n_2}{n_1}\sum_{s=0}^{\infty} a_s^{(0)}\;\mathrm{i}^s\; C_s^{\frac{1}{2}}(\cos\psi) j_s(\omega) \\ I_1^{(e)} &= 2\frac{n_2}{n_1}\sin\psi\sum_{s=0}^{\infty} a_s^{(1)}\;\mathrm{i}^s\; C_s^{\frac{3}{2}}(\cos\psi) j_{s+1}(\omega) \\ I_2^{(e)} &= 2\frac{n_2}{n_1}\sin^2\psi\sum_{s=0}^{\infty} a_s^{(2)}\;\mathrm{i}^s\; C_s^{\frac{5}{2}}(\cos\psi) j_{s+2}(\omega) \end{aligned}\right\}, \tag{6.13}$$

where, for $d \neq 0$:

$$\left.\begin{aligned}
a_s^{(0)} &= N_s^{(0)} \int_0^{\beta} \frac{\cos\varphi_2}{\sqrt{\cos\varphi_1}} \sin\varphi_1(\tau_s + \tau_p \cos\varphi_2) \exp[\mathrm{i}k_0\Psi(\cos\varphi_2)] C_s^{\frac{1}{2}}(\cos\varphi_2)\mathrm{d}\varphi_2 \\
a_s^{(1)} &= N_s^{(1)} \int_0^{\beta} \frac{\cos\varphi_2}{\sqrt{\cos\varphi_1}} \sin\varphi_1\tau_p \exp[\mathrm{i}k_0\Psi(\cos\varphi_2)] C_s^{\frac{3}{2}}(\cos\varphi_2) \sin^2\varphi_2\mathrm{d}\varphi_2 \\
a_s^{(2)} &= N_s^{(2)} \int_0^{\beta} \frac{\cos\varphi_2}{\sqrt{\cos\varphi_1}} \sin\varphi_1(\tau_s - \tau_p \cos\varphi_2) \exp[\mathrm{i}k_0\Psi(\cos\varphi_2)] \\
&\qquad\qquad \times C_s^{\frac{5}{2}}(\cos\varphi_2) \sin^2\varphi_2\mathrm{d}\varphi_2
\end{aligned}\right\}, \quad (6.14)$$

and for $d = 0$ (when $\varphi_1 = \varphi_2 = \varphi$ and $n_1 = n_2$):

$$\left.\begin{aligned}
a_s^{(0)} &= N_s^{(0)} \int_0^{\alpha} \sin\varphi\sqrt{\cos\varphi}\ (1 + \cos\varphi)\ C_s^{\frac{1}{2}}(\cos\varphi)\mathrm{d}\varphi \\
a_s^{(1)} &= N_s^{(1)} \int_0^{\alpha} \sin^3\varphi\sqrt{\cos\varphi}\ C_s^{\frac{3}{2}}(\cos\varphi)\mathrm{d}\varphi \\
a_s^{(2)} &= N_s^{(2)} \int_0^{\alpha} \sin^3\varphi\sqrt{\cos\varphi}\ (1 - \cos\varphi)\ C_s^{\frac{5}{2}}(\cos\varphi)\mathrm{d}\varphi
\end{aligned}\right\},$$

We also have a similar set of equations for the magnetic diffraction integrals $I_n^{(h)}$ and coefficients $b_s^{(n)}$:

$$\left.\begin{aligned}
I_0^{(h)} &= 2\frac{n_2}{n_1} \sum_{s=0}^{\infty} b_s^{(0)}\ \mathrm{i}^s\ C_s^{\frac{1}{2}}(\cos\psi) j_s(\omega) \\
I_1^{(h)} &= 2\frac{n_2}{n_1} \sin\psi \sum_{s=0}^{\infty} b_s^{(1)}\ \mathrm{i}^s\ C_s^{\frac{3}{2}}(\cos\psi) j_{s+1}(\omega) \\
I_2^{(h)} &= 2\frac{n_2}{n_1} \sin^2\psi \sum_{s=0}^{\infty} b_s^{(2)}\ \mathrm{i}^s\ C_s^{\frac{5}{2}}(\cos\psi) j_{s+2}(\omega)
\end{aligned}\right\}, \quad (6.15)$$

where, for $d \neq 0$:

$$\left.\begin{aligned}
b_s^{(0)} &= N_s^{(0)} \int_0^{\beta} \frac{\cos\varphi_2}{\sqrt{\cos\varphi_1}} \sin\varphi_1(\tau_p + \tau_s \cos\varphi_2) \exp[\mathrm{i}k_0\Psi(\cos\varphi_2)] C_s^{\frac{1}{2}}(\cos\varphi_2)\mathrm{d}\varphi_2 \\
b_s^{(1)} &= N_s^{(1)} \int_0^{\beta} \frac{\cos\varphi_2}{\sqrt{\cos\varphi_1}} \sin\varphi_1\tau_s \exp[\mathrm{i}k_0\Psi(\cos\varphi_2)] C_s^{\frac{3}{2}}(\cos\varphi_2) \sin^2\varphi_2\mathrm{d}\varphi_2 \\
b_s^{(2)} &= N_s^{(2)} \int_0^{\beta} \frac{\cos\varphi_2}{\sqrt{\cos\varphi_1}} \sin\varphi_1(\tau_p - \tau_s \cos\varphi_2) \exp[\mathrm{i}k_0\Psi(\cos\varphi_2)] \\
&\qquad\qquad \times C_s^{\frac{5}{2}}(\cos\varphi_2) \sin^2\varphi_2\mathrm{d}\varphi_2
\end{aligned}\right\}, \quad (6.16)$$

and for $d = 0$ (when $\varphi_1 = \varphi_2 = \varphi$ and $n_1 = n_2$), $a_s^{(n)} = b_s^{(n)}$. Although Eqs. (6.13) and (6.14) conclude the solution of the problem, we must pay attention to the equation set of (6.14) and the corresponding coefficients of the magnetic field. Referring to Eq.

(5.35) which shows that the aberration function is proportional to the focussing depth (d), the function when larger focussing depths are being calculated gives a rapidly oscillating contribution to Eq. (6.14). Thus we expand this function as a series of polynomials. For this we recall Eq. (5.35):

$$\Psi = d(n_2 \cos\varphi_2 - n_1 \cos\varphi_1).$$

Thus

$$\exp[\mathrm{i}k_0\Psi(\cos\varphi_2)] = \exp\left[\mathrm{i}k_2 d\left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right)\right]. \tag{6.17}$$

Now using the formula given by Watson[104] which is the degenerate form of the addition theorem:

$$\exp(\mathrm{i}z\cos\phi) = \sum_{n=0}^{\infty}(2n+1)\ \mathrm{i}^n\ C_n^{\frac{1}{2}}(\cos\phi)j_n(z)$$

and where for our case

$$\cos\phi = \cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1$$

and for the particular case when $n_1 \leq n_2$ and $0 \leq \varphi_2 \leq \beta \leq \frac{\pi}{2}$, from which one obtains†

$$\exp[\mathrm{i}dk_0\Psi(\cos\varphi_2)] = \sum_{n=0}^{\infty}(2n+1)\ \mathrm{i}^n C_n^{\frac{1}{2}}\left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right) j_n(k_2 d). \tag{6.18}$$

As the result of substitution from Eq. (6.18) into Eq. (6.14) the integrand carry complex numbers, which is difficult to integrate numerically. To overcome this problem the order of integration and summation is changed‡. Thus we obtain:

$$a_s^{(0)} = N_s^{(0)}\sum_{n=0}^{\infty}(2n+1)\ \mathrm{i}^n\ j_n(k_2 d)\int_0^{\beta}\frac{\cos\varphi_2}{(\cos\varphi_1)^{1/2}}\sin\varphi_1(\tau_s + \tau_p\cos\varphi_2)$$
$$\times C_s^{\frac{1}{2}}(\cos\varphi_2)C_n^{\frac{1}{2}}\left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right)\mathrm{d}\varphi_2 \tag{6.19a}$$

†We note that $C_n^{\frac{1}{2}}(x) = P_n(x)$, where $P_n(x)$ is the Legendre polynomial of first kind order n, but we used the Gegenbauer polynomials for the sake of consistency.

‡This can be done since the expansion is uniformly convergent and each individual term is continuous for $d \geq 0$ and $n_1 \leq n_2$. For $n_1 > n_2$ we have $j_n(k_1 d)$ and $C_n^{1/2}\left(\frac{n_2}{n_1}\cos\varphi_2 - \cos\varphi_1\right)$ in the expansion.

$$a_s^{(1)} = N_s^{(1)} \sum_{n=0}^{\infty} (2n+1)\ \mathrm{i}^n\ j_n(k_2 d) \int_0^{\beta} \frac{\cos\varphi_2}{(\cos\varphi_1)^{1/2}} \sin\varphi_1 \tau_p \times C_s^{\frac{3}{2}}(\cos\varphi_2) C_n^{\frac{1}{2}} \left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right) \sin^2\varphi_2 \mathrm{d}\varphi_2 \quad (6.19b)$$

$$a_s^{(2)} = N_s^{(2)} \sum_{n=0}^{\infty} (2n+1)\ \mathrm{i}^n\ j_n(k_2 d) \int_0^{\beta} \frac{\cos\varphi_2}{(\cos\varphi_1)^{1/2}} \sin\varphi_1 (\tau_s - \tau_p \cos\varphi_2) \times C_s^{\frac{5}{2}}(\cos\varphi_2) C_n^{\frac{1}{2}} \left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right) \sin^2\varphi_2 \mathrm{d}\varphi_2 \quad (6.19c)$$

and the same procedure yields similar expressions for $I_n^{(h)}$ and $(a_s^n)^{(h)}$:

$$b_s^{(0)} = N_s^{(0)} \sum_{n=0}^{\infty} (2n+1)\ \mathrm{i}^n\ j_n(k_2 d) \int_0^{\beta} \frac{\cos\varphi_2}{(\cos\varphi_1)^{1/2}} \sin\varphi_1 (\tau_p + \tau_s \cos\varphi_2) \times C_s^{\frac{1}{2}}(\cos\varphi_2) C_n^{\frac{1}{2}} \left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right) \mathrm{d}\varphi_2 \quad (6.20a)$$

$$b_s^{(1)} = N_s^{(1)} \sum_{n=0}^{\infty} (2n+1)\ \mathrm{i}^n\ j_n(k_2 d) \int_0^{\beta} \frac{\cos\varphi_2}{(\cos\varphi_1)^{1/2}} \sin\varphi_1 \tau_s \times C_s^{\frac{3}{2}}(\cos\varphi_2) C_n^{\frac{1}{2}} \left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right) \sin^2\varphi_2 \mathrm{d}\varphi_2 \quad (6.20b)$$

$$b_s^{(2)} = N_s^{(2)} \sum_{n=0}^{\infty} (2n+1)\ \mathrm{i}^n\ j_n(k_2 d) \int_0^{\beta} \frac{\cos\varphi_2}{(\cos\varphi_1)^{1/2}} \sin\varphi_1 (\tau_p - \tau_s \cos\varphi_2) \times C_s^{\frac{5}{2}}(\cos\varphi_2) C_n^{\frac{1}{2}} \left(\cos\varphi_2 - \frac{n_1}{n_2}\cos\varphi_1\right) \sin^2\varphi_2 \mathrm{d}\varphi_2 \quad (6.20c)$$

The integrand as represented in Eq. (6.19) are real and therefore can readily be computed numerically. Furthermore, the larger d values, which caused oscillations in Eq. (6.19), are present here in the argument of the spherical Bessel function which rapidly converges to zero as d increases.

One of the most desirable applications is the calculation of line scans, i.e. when $v_2 = 0$. For this it follows from Eq. (6.8) that

$$\omega = \frac{u_2}{\sin^2\beta},$$

that is, $\cos\psi = \pm 1$ and $\sin\psi = 0$. It is easy to see from the definition of the Gegenbauer polynomials Eq. (A.24) and also from the recurrence relation Eq. (A.25), that $C_n^{\frac{1}{2}}(1) = 1$.

For the case of $C_n^{\frac{1}{2}}(-1)$ we have:

$$C_n^{\frac{1}{2}}(-1) = \begin{cases} +1 & \text{if } s \text{ even} \\ -1 & \text{if } s \text{ odd} \end{cases}$$

Furthermore it is also evident from Eq. (6.13) that for this case $I_1^{(e)} = I_2^{(e)} = 0$. Thus only $I_0^{(e)}$ contributes to the axial scan and that is:

$$I_A^{(e)} = \begin{cases} 2\sum_{s=0}^{\infty} a_s^{(0)} \, (-\mathrm{i})^s \, j_s\left(\dfrac{u_2}{\sin^2\beta}\right) & \text{when} \quad u_2 < 0 \\ 2\sum_{s=0}^{\infty} a_s^{(0)} \, \mathrm{i}^s \, j_s\left(\dfrac{u_2}{\sin^2\beta}\right) & \text{when} \quad u_2 \geq 0 \end{cases} \tag{6.21}$$

where $a_s^{(0)}$ is given by Eq. (6.14) and $N_s^{(0)}$ also takes a particularly simple form:

$$N_s^{(0)} = s + \frac{1}{2}.$$

This concludes the evaluation of the diffraction integrals. The components of the electric and magnetic fields are given by Eqs. (5.40) and (5.41). From these expressions it is seen that because of $I_1^{(e)} = I_2^{(e)} = 0$, the electric field on the optical axis is always linearly polarised and the direction of polarisation coincides with the x direction.

6.3 Linearisation of the aberration function

The aim of the linearisation of the aberration function[§] is to make the calculation of coefficients $a_s^{(n)}$ and $b_s^{(n)}$ simpler so the integration can be performed on less powerful computers as well. Equation (6.1) can be re-written in a slightly modified form:

$$I_n = \frac{n_2}{n_1}\int_0^{\beta} f(\cos\varphi_2)\sin\varphi_2 \exp[\mathrm{i}k_0\Psi(\cos\varphi_2)]J_n\left(\frac{v_2\sin\varphi_2}{\sin\beta}\right) \times \exp\left(\frac{\mathrm{i}u_2\cos\varphi_2}{\sin^2\beta}\right)\mathrm{d}\varphi_2. \tag{6.22}$$

On substituting into the above equation:

$$\begin{aligned} x &= \cos\varphi_2 \\ \mathrm{d}x &= -\sin\varphi_2 \mathrm{d}\varphi_2 \end{aligned} \tag{6.23}$$

[§]The original idea of the linearisation was proposed by Dr P. Varga.

we obtain:

$$I_n = \frac{n_2}{n_1} \int_{\cos\beta}^{1} f(x) \exp[\mathrm{i}k_0 \Psi(x)] J_n \left(\frac{v_2 (1 - x^2)^{1/2}}{\sin\beta} \right) \exp\left(\frac{\mathrm{i}u_2 x}{\sin^2\beta} \right) \mathrm{d}x. \tag{6.24}$$

We define:

$$\frac{\Gamma(x)}{d} = \Psi(x) = d\left[n_2 x - \sqrt{n_2^2 - n_1^2 + n_2^2 x^2} \right] \tag{6.25a}$$

where

$$\Gamma(x) = \left[n_2 x - \sqrt{n_2^2 - n_1^2 + n_2^2 x^2} \right]. \tag{6.25b}$$

$\Gamma(x)$ can be approximated by a linear function in the following manner:

$$\Gamma(x) \approx \Upsilon + \Xi x \tag{6.26a}$$

or alternatively:

$$\Gamma(x) = \Upsilon + \Xi x + E(x) \tag{6.26b}$$

where Υ and Ξ are constants and $E(x)$ is an error function. With these expressions Eq. (6.24) can be re-written:

$$\begin{aligned} I_n = \frac{n_2}{n_1} \exp(\mathrm{i}k_0 d\Upsilon) \int_{\cos\beta}^{1} f(x) \exp[\mathrm{i}k_0 E(x)] J_n \left[\frac{v_2 (1 - x^2)^{1/2}}{\sin\beta} \right] \\ \times \exp\left[\mathrm{i}x \left(\Xi d + \frac{u_2}{\sin^2\beta} \right) \right] \mathrm{d}x. \end{aligned} \tag{6.27}$$

On introducing a spherical polar co-ordinate system:

$$\left.\begin{aligned} \omega \sin\psi &= \frac{v_2}{\sin\beta} \\ \omega \cos\psi &= \Xi d + \frac{u_2}{\sin^2\beta} \end{aligned}\right\} \tag{6.28}$$

from which we write

$$\begin{aligned} I_n = \frac{n_2}{n_1} \exp(\mathrm{i}k_0 d\Upsilon) \int_{\cos\beta}^{1} f(x) \exp[\mathrm{i}k_0 E(x)] J_n \left[\omega \sin\psi (1 - x^2)^{1/2} \right] \\ \times \exp\left[\mathrm{i}x\omega \cos\psi \right] \mathrm{d}x. \end{aligned} \tag{6.29}$$

From the above equation it is seen that only $\exp[ik_0 dE(x)]$ enters into the expressions for $a_s^{(n)}$ and $b_s^{(n)}$ in Eqs. (6.14) and (6.16). This function gives a less oscillatory contribution to the integrand (the difference between the magnitudes of $\Gamma(x)$ and $E(x)$ is $\sim 10^2$), thus the integral is readily calculable. It must be emphasised that the full expression for I_n is not simplified as the spherical Bessel function contains the term $\omega \cos\psi$ which gives rise to a slower convergence of the sum. The summation can be performed readily compared with the original integrals.

6.4 Numerical results

In this section we present results of numerical computations for the time averaged electric energy distributions performed on the model proposed in Chapter 5. For the numerical results shown below the electric field is determined at a particular point at a depth d below the surface (referred to as the *penetration* or *focussing depth*). This corresponds to the case of a scanning microscope, i.e. we wish to know the electric field at a particular point, i.e. at a steady position in the specimen (second material) when the specimen is scanned. This is of practical importance since, when the scattering is determined in Chapter 8, the knowledge of the electric field that interacts with a particle is essential. Figure 6.1 is a schematic diagram (ray refraction at the interface and the paraxial rays are not drawn) that explains how the electric energy distributions are calculated.

For each particular value of d, the field is determined at the point where the particle is situated as the probe is placed at different places. This happens when the Gaussian focus coincides with the particle (0 μm), when it is below it (e.g. 5 and 50 μm), when it is above it and in the specimen (not shown) or when it is above it and outside the specimen (e.g. -50 μm). When the probe is below the particle, the field is due to a converging beam, and when above the particle, it is due to a diverging beam. Figure 6.1 shows the probe directly below or above the particle, i.e. x or y = 0. Precisely similar considerations apply when the probe is in addition laterally displaced from the particle, i.e. x or y is positive or negative. The theory developed in Chapter 5 applies equally well to all of these cases. The electric energy distributions plotted in the diagrams given below were all calculated

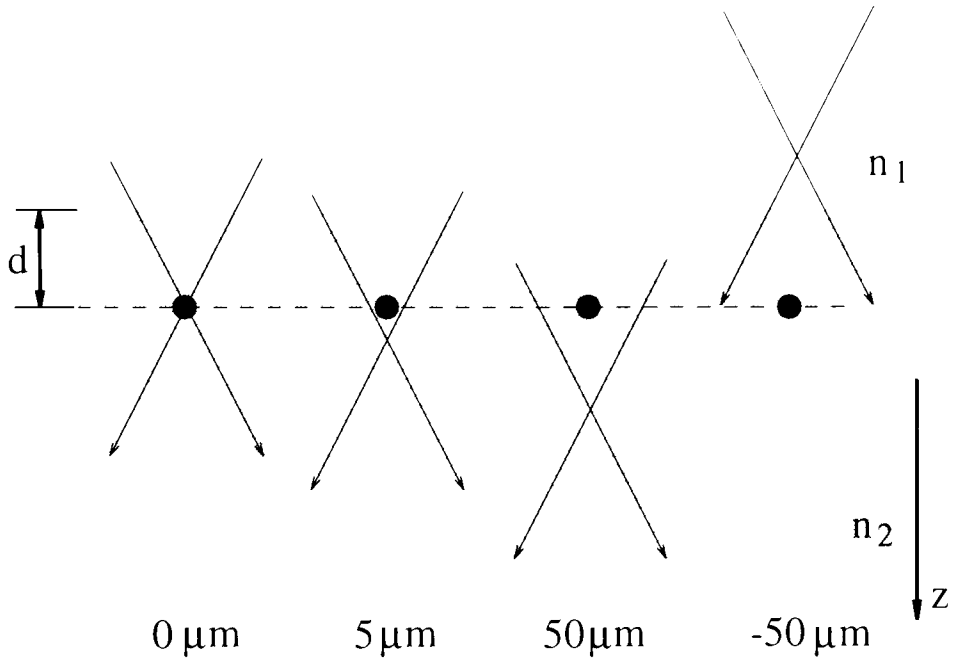


Figure 6.1: Figure showing details concerning the numerical computation.

in this way.

The basic equations of the present calculations are Eqs. (5.40) and (5.41). It is advantageous to define a quantity corresponding to the electric field that does not explicitly depend on time. For this purpose we define the *time averaged electric energy* and denote it by $\langle w_e(v,u,\vartheta_p)\rangle$. This quantity can be computed from $\mathbf{e} = (e_x, e_y, e_z)$ as follows[75] (omitting the p subscript which indicates that this co-ordinate corresponds to the point of observation):

$$\begin{aligned}\langle w_e(v,u,\vartheta)\rangle &= \frac{1}{16\pi}(\mathbf{e}\cdot\mathbf{e}^*) \\ &= \frac{A^2}{16\pi}\left\{|I_0^{(e)}|^2 + 4|I_1^{(e)}|^2\cos^2\vartheta + |I_2^{(e)}|^2 + 2\cos 2\vartheta\Re[I_0^{(e)}(I_2^{(e)})^*]\right\}.\end{aligned} \tag{6.30}$$

The time averaged electric energy density was calculated for a focussing depth of 10, 20, 50, 100, 200 and 300 μm and for the 0° and 90° meridional planes. The 0° meridional plane corresponds to the direction of incident polarisation. For each meridional plane, the electric energy distribution in the negative radial direction is a mirror image of the electric energy distribution in the positive radial direction, and so only the positive half of each image has been displayed. On the axial scale 0 corresponds to the Gaussian focus. The result of these calculations are shown in Figs. 6.2, 6.3 and 6.4.

In these figures, sub-figure (a) shows the time averaged electric energy density whilst

sub-figure (b) shows its 10 based logarithm. The horizontal axes correspond to the radial co-ordinate (r) and the dimensions are given in μm. The vertical axes correspond to the axial co-ordinate (z) and the dimensions are also given in μm. The focussing depth and the vertical axes are scaled for silicon ($n_2 = 3.5$). The figures are plotted on a grey-scale where the brightest tones correspond to the highest electric energy density values and the darkest tones correspond to the lowest electric energy density values. Each of the individual figures is normalised to the maximum of the electric energy density present in the actual figure. The intermediate tones are chosen equally between the maximum and minimum values.

The axial position of the maximum electric energy is referred to as the *diffraction focus.* The relative value of the electric energy density at the diffraction focus is a measure of the spherical aberration and is usually referred to as the *Strehl ratio.* It is known that the presence of spherical aberration causes a shift in the axial position of the diffraction focus. The amount of the shift is expressed as the distance of the diffraction focus from the Gaussian (or paraxial) focus position and is usually referred to as the *focus shift.*

It is apparent from both the linear (a) and logarithmic (b) plots and also from numerical analysis not presented here, that for each focussing depth the lateral full width at half maximum of the electric energy density at the diffraction focus (termed as lateral resolution) is superior in the 90° meridional plane than in the 0° meridional plane. It is also seen that the side lobes (secondary peaks) in the positive axial direction are better defined in the 0° meridional plane which also indicates the elliptical symmetry of the time averaged electric energy density with respect to the optical axis (z). Furthermore, for each focussing depth up to 300 μm, there always exists an electric energy density that defines a probe where the highest energy is concentrated.

From the SIRM point of view it is of primary importance to predict the maximum of the electric energy density at the diffraction focus as a function of focussing depth. For this the Strehl ratio is calculated for the depth range of 0 to 300 μm and presented in Fig. 6.5.

This shows that the electric energy density decreases by a factor of ~ 4 as the focussing

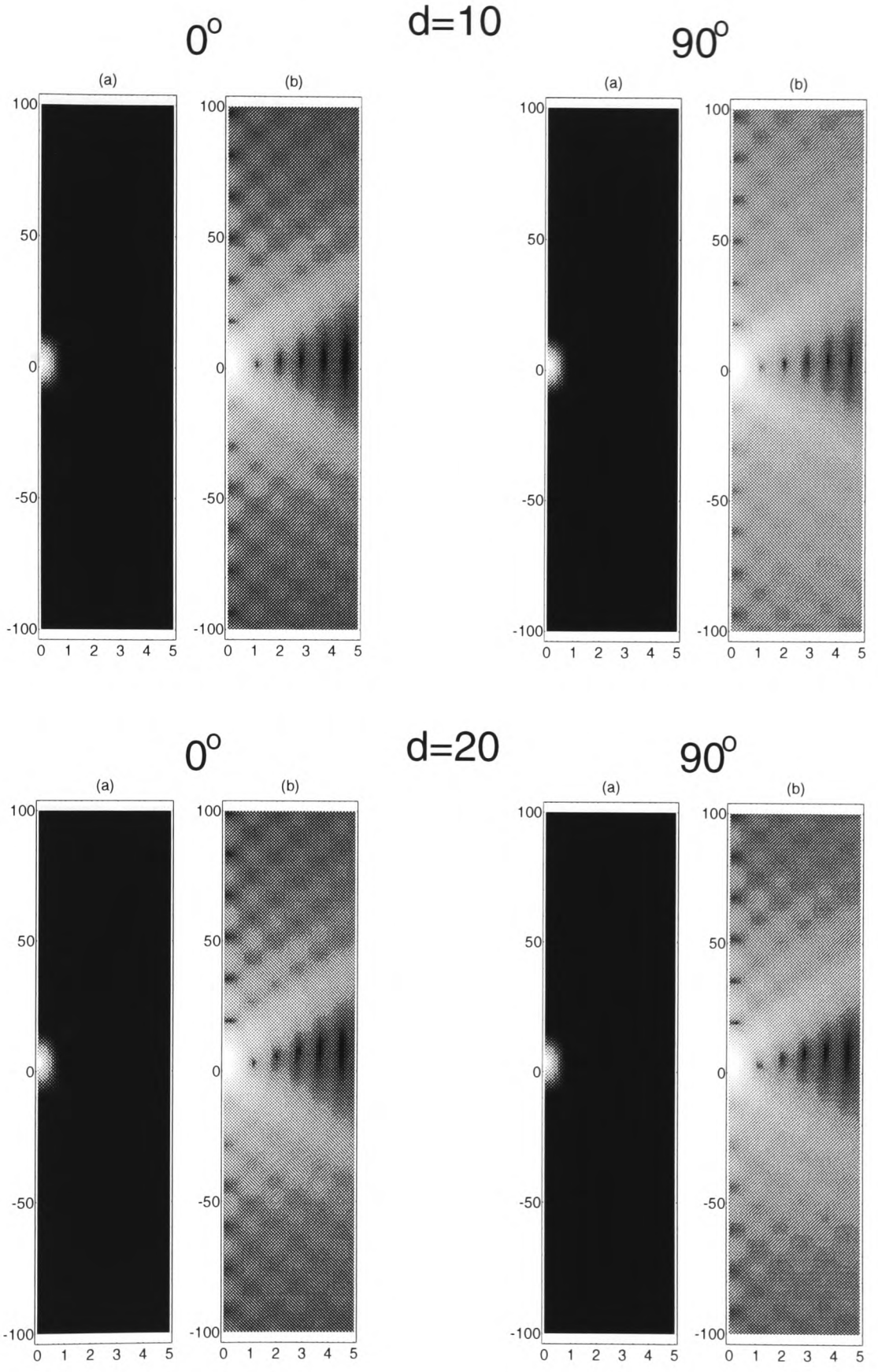


Figure 6.2: Time averaged electric energy density distribution ($\langle w_e(v,u,\vartheta)\rangle$) for focussing depths 10 and 20 μm. For details see text.

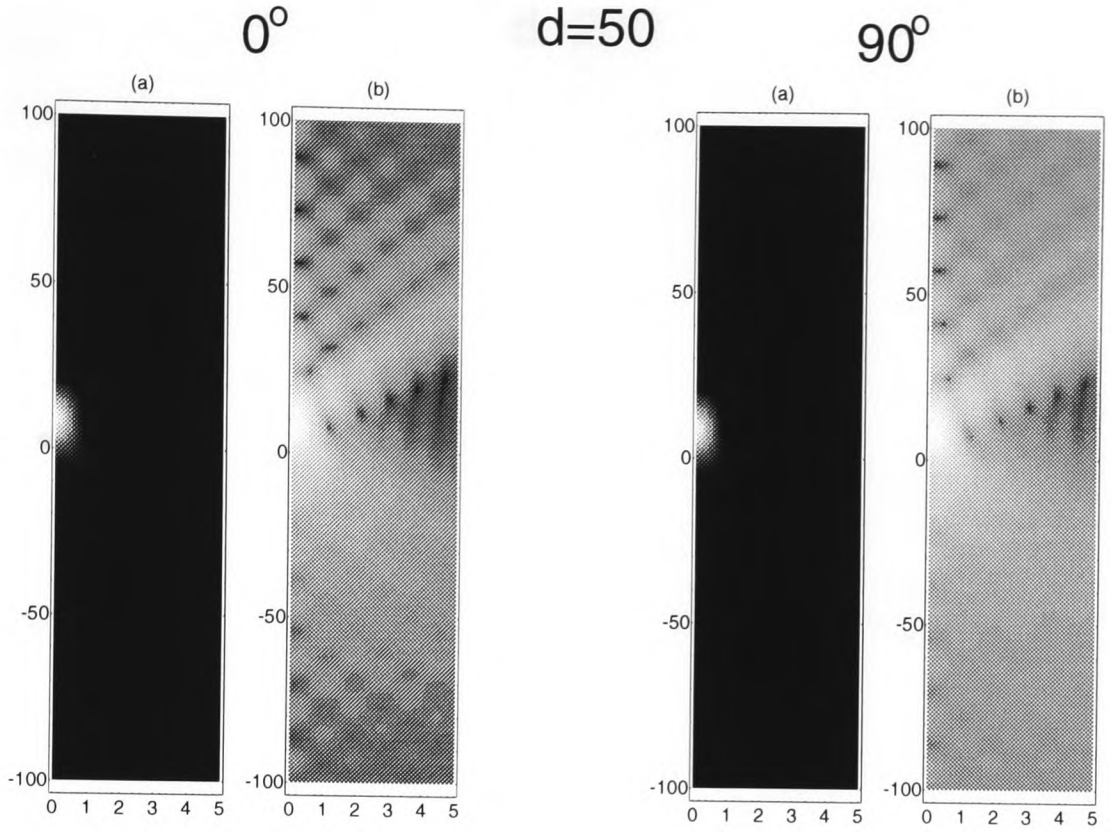


Figure 6.3: Time averaged electric energy density distribution ($\langle w_e(v,u,\vartheta)\rangle$) for focussing depths 50 and 100 μm. For details see text.

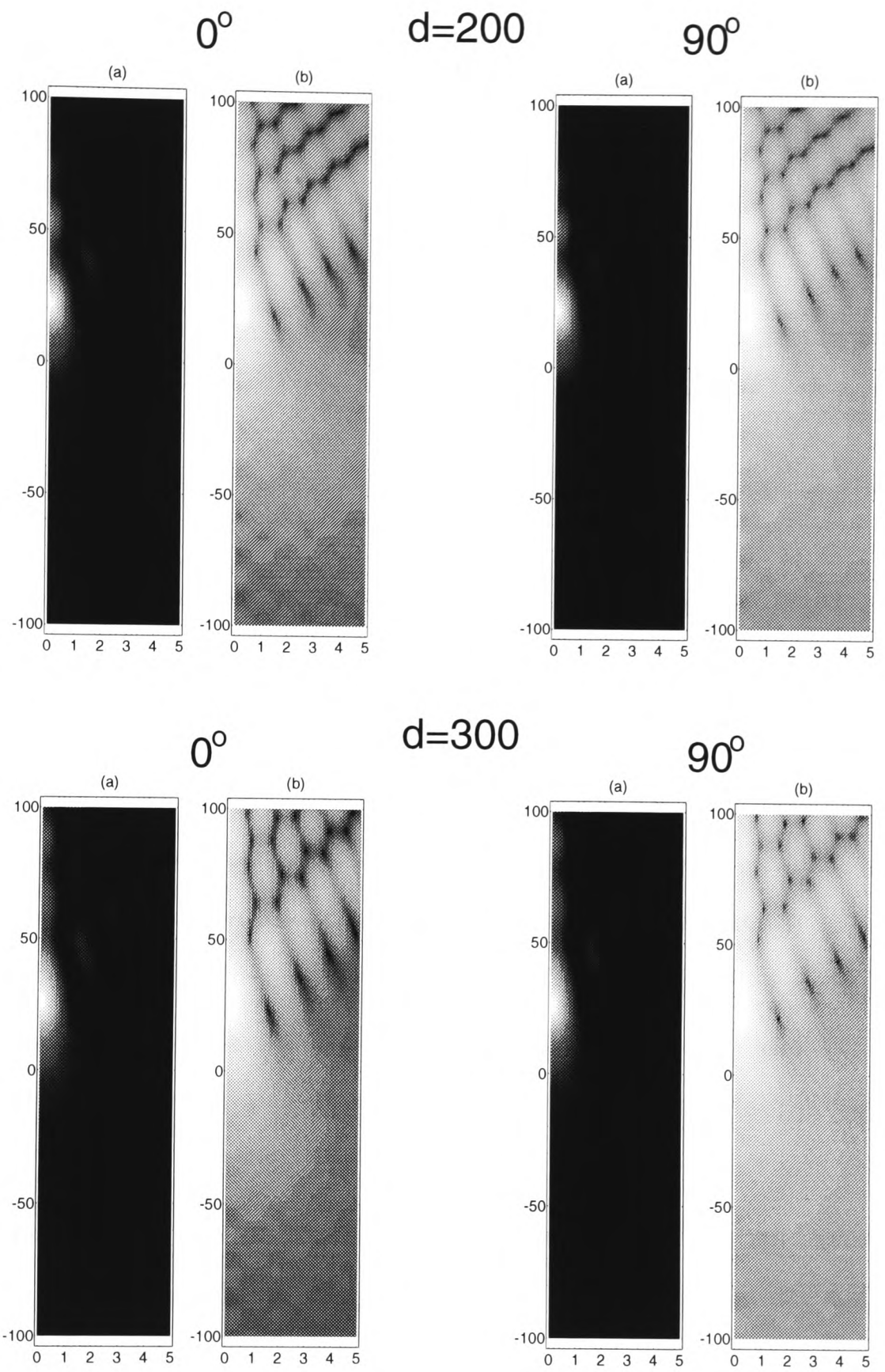


Figure 6.4: Time averaged electric energy density distribution ($\langle w_e(v,u,\vartheta)\rangle$) for focussing depths 200 and 300 μm. For details see text.

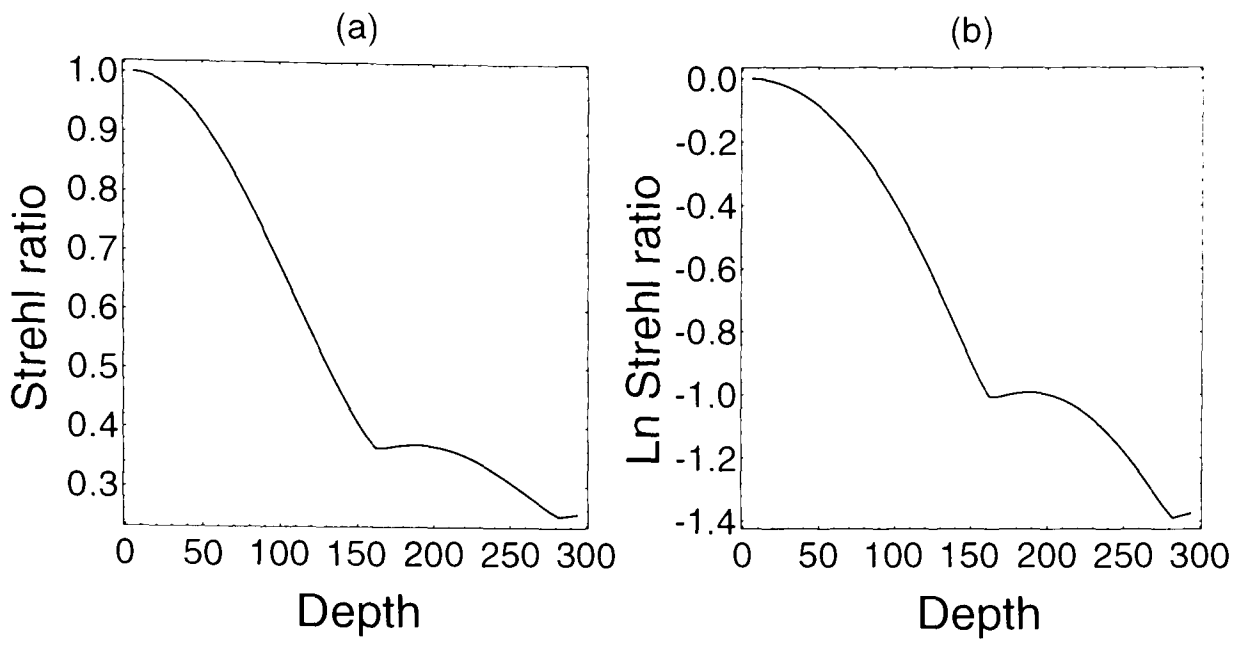


Figure 6.5: The Strehl ratio for focussing depths of 0 to 300 μm. Figure (a) is in linear scale whilst Figure (b) is in (e based) logarithmic scale.

depth increases from 0 to 300 μm. Up to a focussing depth of $\sim$ 165 μm the maximum energy at the diffraction focus is a monotonic function of the focussing depth, but at larger focussing depths this monotony is not true any more. This phenomenon was investigated further when the axial position of the diffraction focus was calculated and plotted against the focussing depth (Fig. 6.6).

This figure shows a non-linear and non-monotonic dependence of the focus shift on the focussing depth similar to the Strehl ratio. Additionally, it is seen that the focussing depths at which the irregularities occur are the same in Figs. 6.5 and 6.6. This dependence is significantly different from the linear dependence predicted by the paraxial theory (see Chapter 5).

To understand this non-monotonic behaviour, electric energy axial line scans were computed for a focussing depth range of 0 to 300 μm and the results are given in Fig. 6.7.

Figure 6.7(a) is plotted on a gray scale and shows the electric energy densities with the lowest values corresponding to the darkest tones and the highest values corresponding to the brightest tones. The scaling is linear between the darkest and brightest tones. Figure 6.7(b) shows the densities with 15 equally spaced isophotes. In Fig. 6.7(c) the energies are plotted on a three-dimensional diagram. The gray scale now corresponds to the square

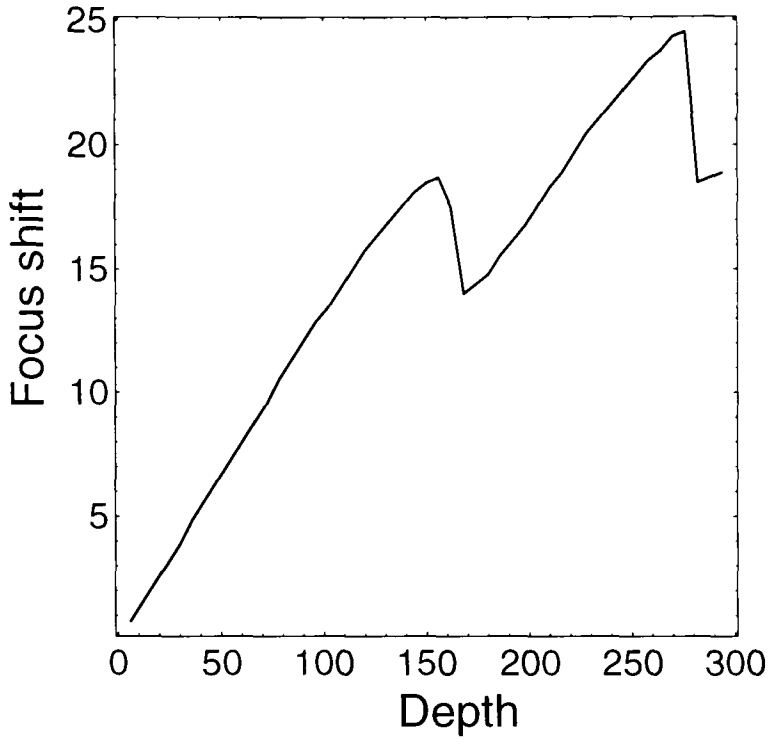


Figure 6.6: The focus shift (in μm) for focussing depths of 0 to 300 μm.

root of the electric energy values at each position, this revealing the lower values more clearly.

As the focussing depth increases from 0, the main peak moves (A in Fig.6.7(c)) in to the positive axial positions and decreases in height (i.e. darker tones). At a focussing depth of 150 μm, the secondary peak (B in Fig.6.7(c)) on the negative side takes over from the previous main peak and becomes the main peak. At a focussing depth of 280 μm, the secondary peak on the negative side (C in Fig.6.7(c)) takes over from the main peak B. As the focussing depth increases from 0, the secondary peaks on the positive side progressively increase in height but do not combine with the main peaks (at least for focussing depths < 300 μm, when as shown in Fig. 6.7 the peaks on the positive side are of low energy and become difficult to distinguish).

Fig. 6.7(c) also shows that as the focussing depth increases, the increase in height of the secondary peak B on the negative side becomes initially a shoulder to the main peak A, and then takes over from it to become the main peak. The increase in the secondary peak C and its taking over from the main peak B is just included in the display. The increase in intensity of the secondary peaks on the positive side, without combining with the main peaks, is also revealed.

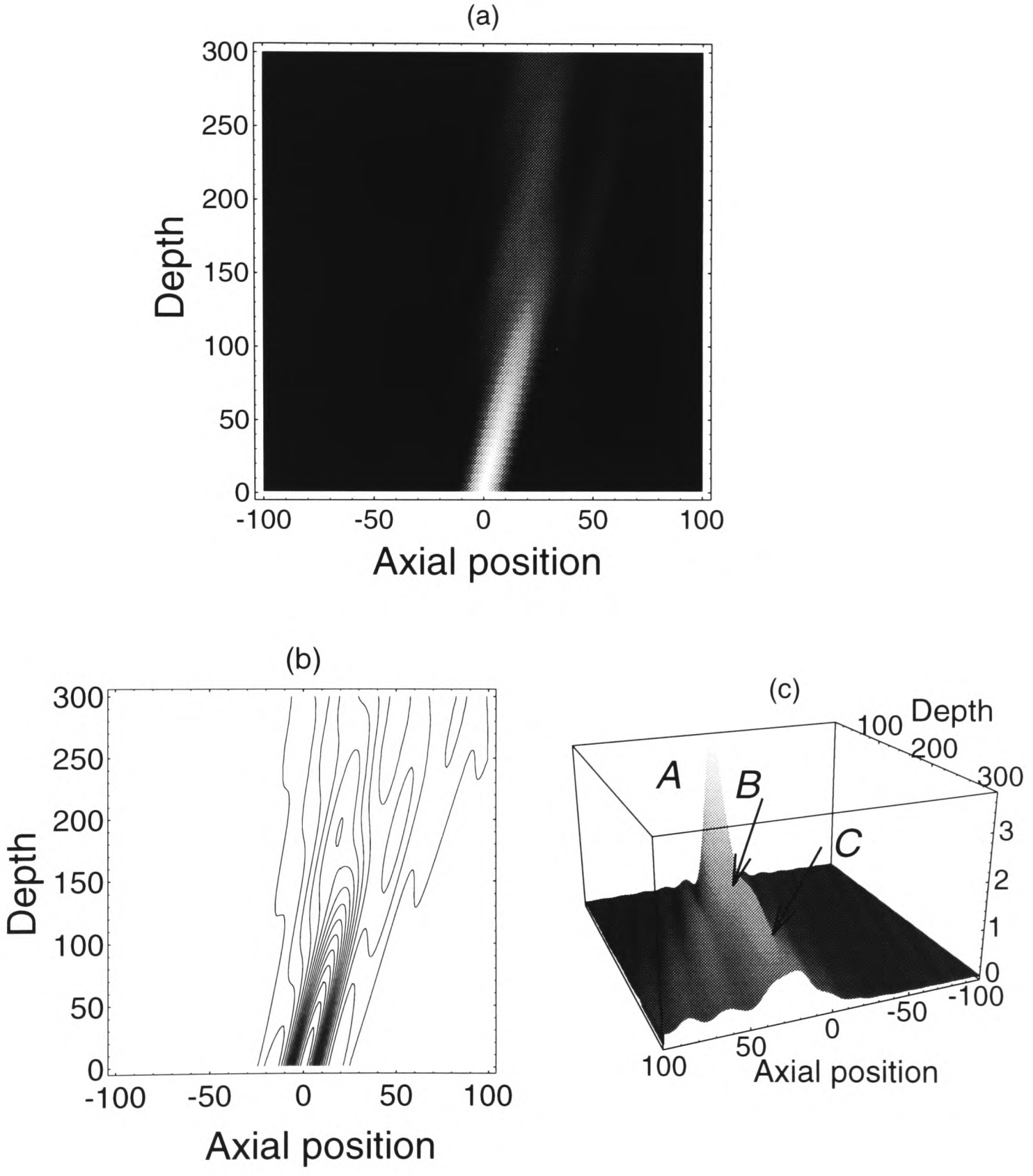


Figure 6.7: Axial line scans for a focussing depth range of 0 to 300 μm. Figure (a) shows the gray-scale representation, Figure (b) shows the corresponding contour lines, and Figure (c) is a three dimensional representation of Figure (a).

6.5 Conclusions

In this chapter we showed how analytical solutions can be obtained for the diffraction integrals of Chapter 5. We also proposed a linearisation method which makes easier and faster the computation of the polynomials for the case when larger spherical aberrations are present. The behaviour of the axial scan as presented in Fig. 6.7 and discussed subsequently fully explains the non-monotonic dependence of both the Strehl ratio and the focus shift on the focussing depth. These findings are, to the best knowledge of the author, new and have never been published in the literature. Instead, most of the authors correctly emphasize that the main peak of the electric energy density or intensity rapidly decreases as the spherical aberration increases, but no attention has been paid to the secondary peaks rising and falling and so overtaking the role of the original main peak. This behaviour is particularly apparent in Figs. 6.5 and 6.6 where the individual parts of the curves for the focussing depth (i.e. 0 to 150 μm, 150 μm to 280 μm and $>$ 280 μm) behave in a similar fashion. This means that as soon as the second rising peak overtakes the original main peak, the second peak can be characterised in exactly the same manner as the original main peak before it was overtaken.

CHAPTER 7

Interpretation of the diffraction integrals

"Some things are easy, they are only hard to do."

A. Einstein

7.1 Introduction

The purpose of this Chapter is to show the correspondence between the 'physical' and the 'optical' measures of aberration, where 'physical' refers to the focussing or penetration depth defined in Chapter 5 and 'optical' refers to the aberration coefficients generally used in optics. To some extent our present treatment is unique as it is the opposite of the usual procedure. In most cases the aim is to find an approximation for the aberration function in terms of the aberration coefficients, while here the main interest is to expand the *analytical* aberration function Ψ in terms of Zernike polynomials.

Surprisingly little has been published on the application of Zernike polynomials in optics. Apart from the pioneering work of Zernike[118], Nijboer[68,69] and Nienhuis[67], Wolf[114] reported the latest results on the diffraction theory of aberrations including the work of the above authors, and Born and Wolf[14] gave an extensive overview on the Nijboer-Zernike theory and examined the structure of diffraction images affected by primary aberrations. Born and Wolf also showed some important properties of Zernike's circular polynomials, e.g. how they can be related to the Jacobi G polynomials. Some years later Bezdidko[8] gave a review of the use of Zernike polynomials in optics. A graphical representation of some of the Zernike polynomials was given by Wang and Silva[103].

In this Chapter after giving a short review on the general properties of Zernike polynomials, the aberration function is examined in detail and it is shown how it can be expanded as a series of Zernike polynomials. It is also shown how the focussing depth

affects the number of polynomials that must be taken into account to achieve the appropriate accuracy and a graphical representation of all the Zernike polynomials needed is given. As a result of the present work it is possible to directly relate the strength of spherical aberration associated with different focussing depths to the better known Zernike coefficients.

7.2 Circular polynomials of Zernike

As discussed in Appendix A, an at least piecewise continuous function $G(x)$ can be expanded on $[a, b]$ as a series of orthogonal functions $\phi_k(x)$:

$$G(x) = \sum_{k=0}^{\infty} a_k \phi_k(x)$$

where the coefficients a_n are given by

$$a_n = N_n \int_a^b G(x) w(x) \phi_n(x) \mathrm{d}x,$$

where $w(x)$ is the weighting function and N_n is the normalisation factor. The set of functions $\phi_n(x)$ is said to be complete if the mean square error vanishes:

$$\lim_{j \to \infty} \int_a^b \left[G(x) - \sum_{n=0}^{j} a_n \varphi_n(x) \right]^2 w(x) \mathrm{d}x = 0.$$

For numerical purposes the above equation cannot be used, so we define instead*:

$$\int_a^b \left[G(x) - \sum_{k=0}^{m} a_k \phi_k(x) \right]^2 w(x) \mathrm{d}x \leq ERR. \tag{7.1}$$

During numerical evaluation the number m is to be determined in such a way that the inequality Eq. 7.1 becomes true. The definition of Zernike's circular polynomial is given[14] by:

$$\left.\begin{aligned} (U_{even})_n^m &= R_n^{(m)}(\varrho) \cos m\vartheta \\ (U_{odd})_n^m &= R_n^{(m)}(\varrho) \sin m\vartheta \end{aligned}\right\} \quad \text{if m} \neq 0$$
$$U_n^0 = R_n^{(0)}(\varrho) \qquad \text{if m} = 0 \Bigg\}. \tag{7.2}$$

*As seen in Fig. 7.2, for our case the mean square error converges monotonically as m increases.

As it was pointed out by Born and Wolf[14], the aberrations which we are concerned with are symmetrical about the meridional plane $\vartheta = 0$, thus only even polynomials enter into the series expansion of the aberration function. It is easy to show that Zernike's circular polynomials form a complete set of orthogonal functions:

$$\int_0^1 R_n^{(m)}(\varrho) R_{n'}^{(m)}(\varrho) \varrho \mathrm{d}\varrho = \frac{1}{2(n+1)} \delta_{n'n}$$

where δ is the Kronecker delta function. The normalisation is chosen so that $R_n^{(m)}(1) = 1$ and thus the weighting function $w(x)$ is unity. The definition of $R_n^{(m)}(\varrho)$ is given[118] by:

$$R_n^{(m)}(\varrho) = \sum_{s=0}^{\frac{n-m}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n+m}{2} - s\right)! \left(\frac{n-m}{2} - s\right)!} \varrho^{n-2s},$$

and the generating function:

$$\frac{\left[1 + z - \sqrt{1 - 2z(1 - 2\varrho^2) + z^2}\right]^m}{(2z\varrho)^m \sqrt{1 - 2z(1 - 2\varrho^2) + z^2}} = \sum_{s=0}^{\infty} z^s R_{m+2s}^{(m)}(\varrho).$$

For our application the special case of $m = 0$ is of importance. From the generating function it is seen that (see Appendix A)

$$R_{2n}^{(0)}(\varrho) = P_n(2\varrho^2 - 1),$$

where $P_n(\rho)$ are the Legendre functions, order n, first kind. It was shown by Nijboer[68,69] how the Zernike polynomials can be connected to the classical expansion of the aberration function, i.e. when the aberration function is expanded as a power series of ϱ, and also how a classification of aberrations can be given. Let Φ indicate the wave aberration, then

$$\Phi = \sum_m \sum_n a_{nm} \varrho^n \cos^m \varphi \tag{7.3a}$$

which expression is usually referred to as the classical expansion of the aberration. Nijboer replaced the power terms $\cos^m \varphi$ by their harmonics†:

$$\Phi = \sum_m \sum_n a_{nm} \varrho^n \cos m\varphi. \tag{7.3b}$$

†These polynomials are orthogonal since $\cos m\vartheta$ is orthogonal to $\cos m'\vartheta$

According to this, with reference to Eq. (7.2), Nijboer suggested that m gave the type and n the order of the aberration, e.g. $m = 0$, 1 and 2 correspond to spherical aberration, coma and astigmatism respectively. Perhaps the most useful property of Zernike's circular polynomials is that they balance the aberrations making it possible to obtain the maximum intensity. For example, an expression for third order spherical aberration contains the appropriate term for defocus, i.e. what the amount of defocus should be to obtain maximum intensity with the given aberration.

7.3 Expansion of the aberration function

We consider now the expansion of Eq. (5.35). For this we can write:

$$\Psi(\varphi_1) = \Psi'(\varphi_1)d = d\left[n_2 \cos\varphi_2 - n_1 \cos\varphi_1\right].$$

The explicit dependence on φ can easily be eliminated by using Snell's law:

$$\Psi(\varrho) = \Psi'(\varrho)d = d\left[\sqrt{n_2^2 - n_1^2\varrho^2} - n_1\sqrt{1-\varrho^2}\right]. \tag{7.4}$$

As the form of the above expression shows the aberration function is independent of ϑ, thus Eq. (7.4) can be expanded with Zernike's circular polynomials as follows:

$$\Psi(\varrho) = \sum_{n=0}^{\infty} A_{n0} R_n^{(0)}(\varrho) \tag{7.5}$$

where A_{n0} are the coefficients:

$$\begin{aligned} A_{n0} &= 2(n+1)\int_0^1 \Psi(\varrho) R_n^{(0)}(\varrho)\varrho \mathrm{d}\varrho \\ &= 2(n+1)d\int_0^1 \Psi'(\varrho) R_n^{(0)}(\varrho)\varrho \mathrm{d}\varrho, \end{aligned} \tag{7.6}$$

and the mean square error is given:

$$\int_0^1 \left[\Psi'(\varrho) - \sum_{n=0}^{m} A'_{n0} R_n^{(0)}(\varrho)\right]^2 \mathrm{d}\varrho \leq \frac{ERR}{d^2}. \tag{7.7}$$

7.4 Numerical results

Numerical computations have been performed to approximate the aberration function of Eq. (5.35). Equation (7.7) was computed for individual d numbers so that the limit of

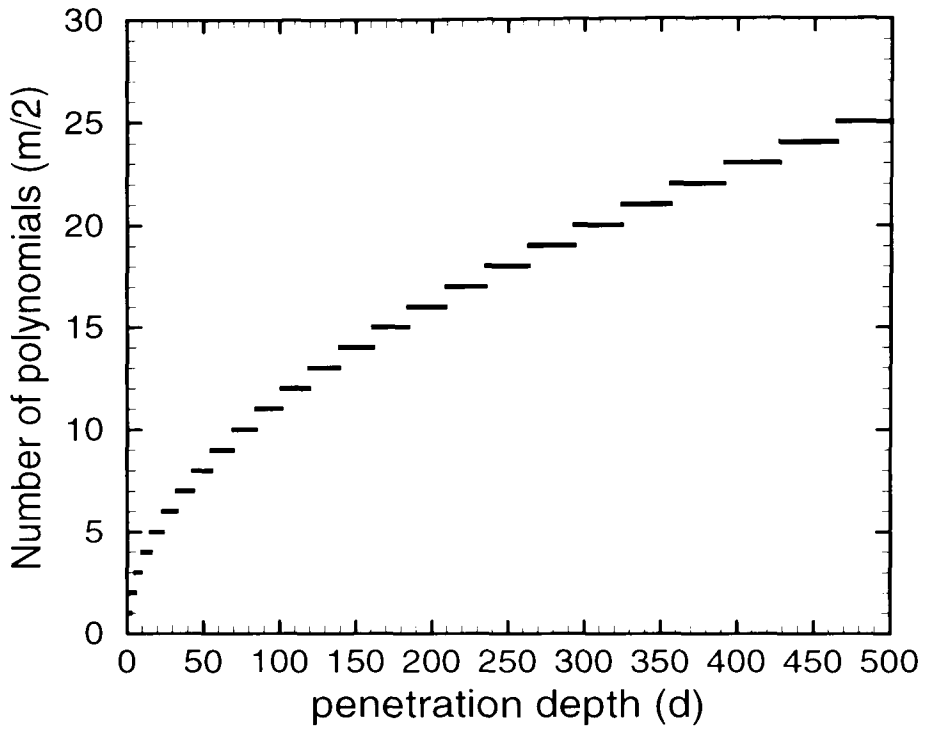


Figure 7.1: Number of Zernike polynomials needed for the expansion of the aberration function.

summation m was increased to the point when the left hand side of the expression gave a number not larger than ERR/d^2. Calculations were performed for two particular cases. First, the light was focussed from air ($n_1 = 1$) into silicon ($n_2 = 3.5$) and second, from air into glass ($n_2 = 1.5$). The result of this numerical computation is shown in Fig. 7.1.

There is essentially no difference between the number of polynomials needed for the case of silicon and glass, and 25 polynomials (corresponding to the 49th order spherical aberration) needed to be computed for a penetration depth of 500μm to obtain a mean square error of 10^{-2}.

The mean square error corresponding to different m numbers was also computed and it decreased rapidly as more polynomials were taken into account. The result of this calculation for both silicon and glass is shown in Fig. 7.2.

The coefficients A'_{n0} are numerically computed for the above two cases and plotted in Fig. 7.3. There are some differences between the coefficients at small n but as these increase the difference between these two types of coefficient vanishes.

As a result of above calculations, it is found that a maximum of 25 polynomials are needed to approximate our aberration function Eq. (5.35). We have computed and plotted in Fig. 7.4 the form of 24 Zernike polynomials $U_n^{(0)}$. The Figure shows the analytical

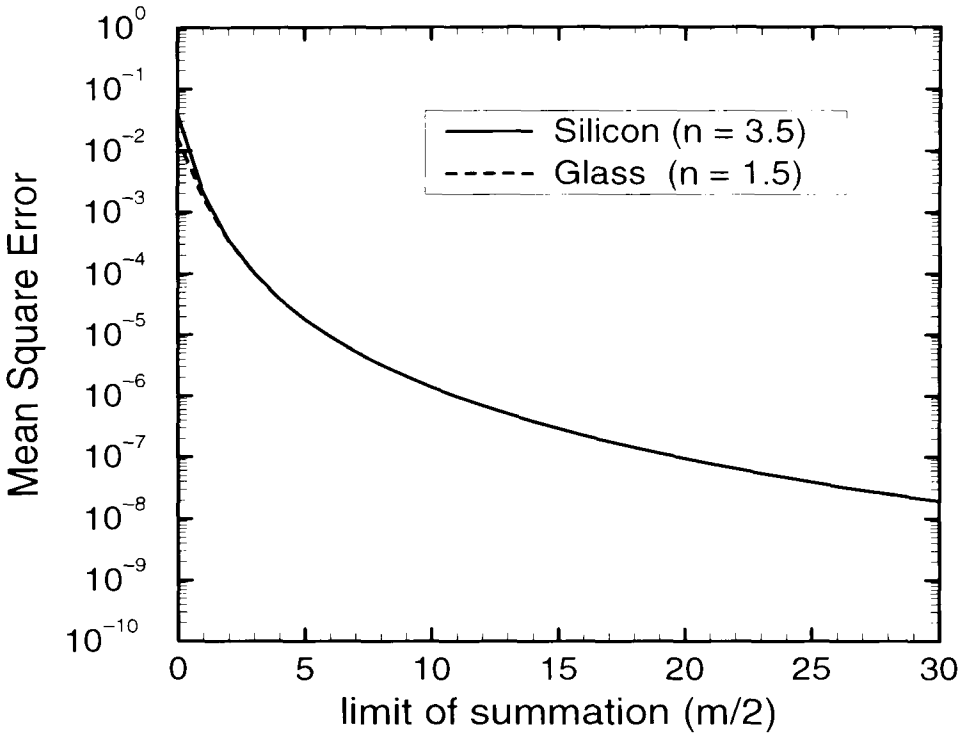


Figure 7.2: The mean square error as more polynomials are summed up.

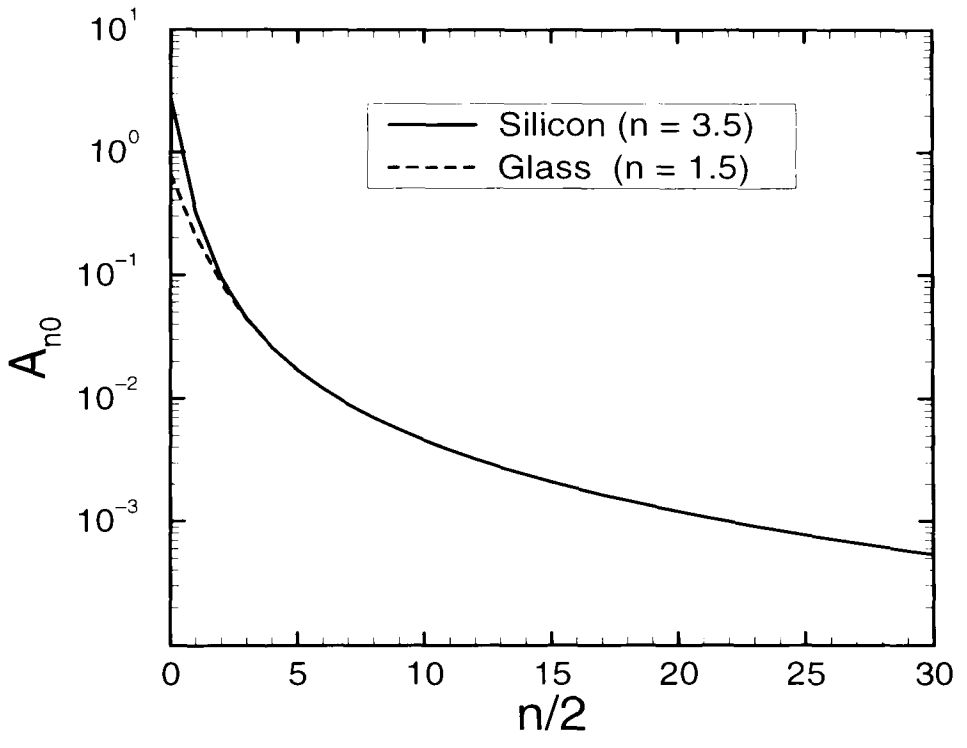


Figure 7.3: The Zernike coefficients.

aberration function $\Psi(\varrho)$, the sum of polynomials for $n = 2 \ldots 48$ ($\sum U_n^{(0)}$) and all of the corresponding individual polynomials ($U_n^{(0)}$).

The graphs are normalised individually so as to obtain the maximum grey-scale range. The corresponding aberrations are, $n = 2$ defocus, $n = 4$ primary spherical aberration, and $n = 6 \ldots 48$ higher order spherical aberrations.

The conclusions that can be drawn as a result of the above considerations are as follows. From the form of the Zernike polynomials the statement given in Chapter 5 on the type of the aberration which is described by function Ψ is clearly justified, i.e. no any other types of aberration could be represented by Ψ. Furthermore, as a consequence of the linear dependence of the aberration function on the penetration depth d, the aberration coefficients are also linear functions of d. It also follows that the expansion of the analytical aberration function does not involve difficulties, essentially only the kernel of the expression for Ψ needs to be computed.

For practical purposes the most generally used aberration coefficients are the Seidel coefficients, which originate from the following geometrical optics consideration. Suppose that a sphere is situated in the image field of an optical system with an origin at the Gaussian image point and the centre point of the sphere coincides with the centre of the aperture. This sphere shall be called the Gaussian reference sphere. Suppose that the actual wavefront of the optical system in the image field is not spherical but distorted. We consider the image of an arbitrary point situated in the object field. A ray pencil from the Gaussian image point falls to the Gaussian reference sphere and the distance from the point of intersection to the Gaussian image point is the radius of the reference sphere. This ray pencil also intersects with the aberrated wavefront. The distance between the intersections on the reference sphere and the aberrated wavefront is called the *wave aberration.* Within the accuracy of the Gaussian optics, this optical distance can be expressed as a power series. It can be shown that due to the symmetry of the optical system, only odd terms enter into this expansion in such a way that the lowest order term will be on the fourth power. The aberrations represented by the lowest order term are usually called *primary* or *Seidel aberrations.* Calculation of the higher order terms is very

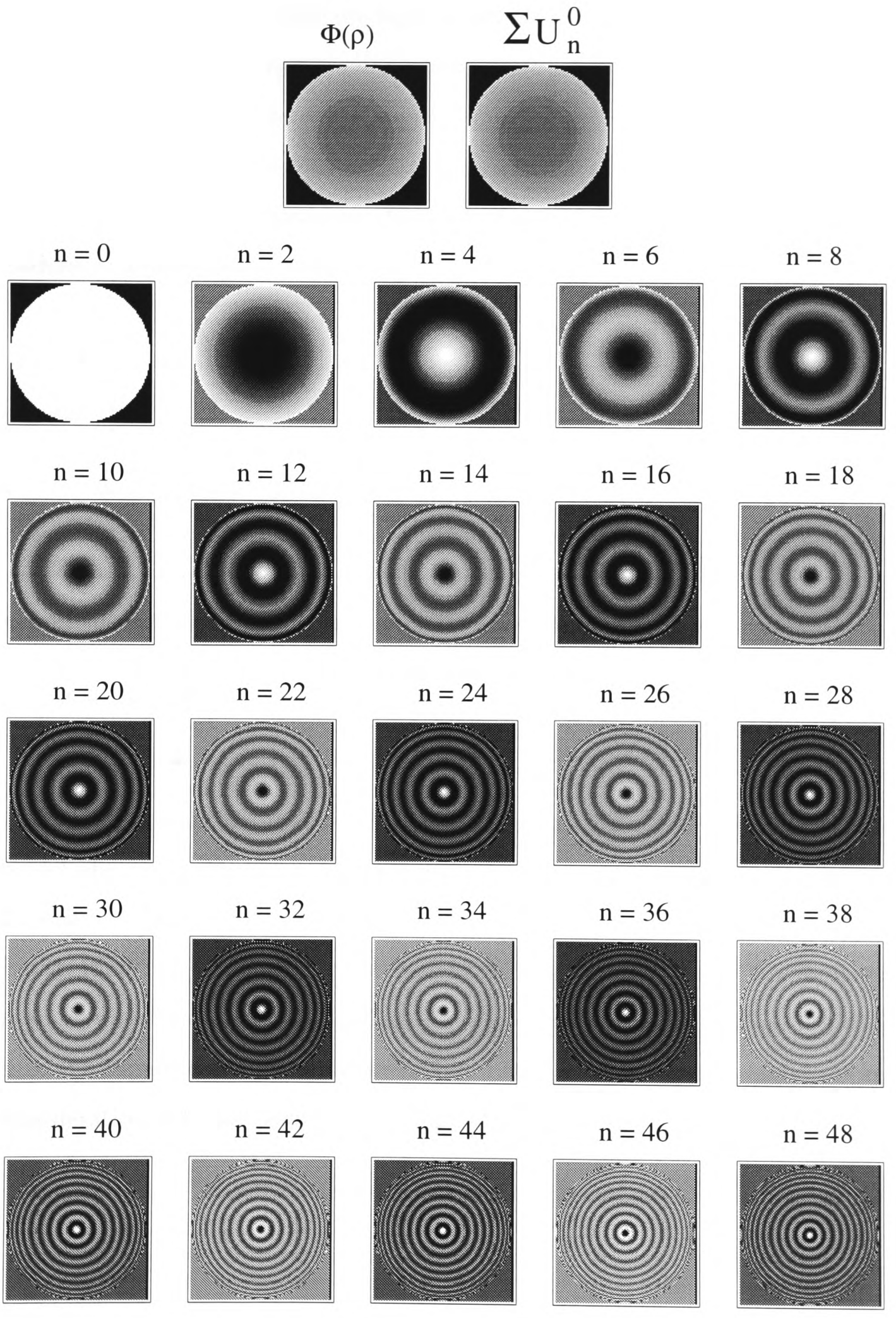


Figure 7.4: The Zernike polynomials.

laborious so they are usually ignored.

It follows from the above that the appropriate term for the spherical aberration within the primary aberration is proportional to the fourth power of the radial co-ordinate. The coefficient of the fourth order term is referred to as the *spherical aberration coefficient* and is generally used as a measure of aberration in diffraction optics.

Now considering the relationship between the Seidel and Zernike coefficients, with aberrations expressed in the form of Eq. 7.3, on putting $m = 0$:

$$\Phi = a_{00} + \sum_{n=2}^{\infty} a_{n0}\varrho^n = a_{00} + a_{20}\varrho^2 + a_{40}\varrho^4 + \dots \tag{7.8a}$$

or from Eq. (7.5)

$$\Phi = A_{00} + \sum_{n=2}^{\infty} A_{n0}R_n^{(0)}(\varrho) = A_{00} + A_{20}R_2^{(0)}(\varrho) + A_{40}R_4^{(0)}(\varrho) + \dots \tag{7.8b}$$

The coefficients from Eq. (7.8a) are given immediately. The correspondence between the Seidel and the Zernike coefficients is given for spherical aberration by:

$$\begin{aligned}\Phi &= A_{00} + \sum_{n=2}^{4} A_{n0}R_n^{(0)}(\varrho) = A_{00} + A_{20}\left(2\varrho^2 - 1\right) + A_{40}\left(6\varrho^4 - 6\varrho^2 + 1\right) \\ &= (A_{00} - A_{20} + A_{40}) + 2\left(A_{20} - 3A_{40}\right)\varrho^2 + 6A_{40}\varrho^4\end{aligned} \tag{7.9}$$

from which expression on equating the fourth power coefficients with those present in the Seidel aberration the required correspondence between the aberration coefficients is established.

Finally, it should be emphasised that the Seidel aberration is only appropriate to describe small wavefront aberrations which is a consequence of the ignorance of higher order terms in the expansion of the wave aberration. At the same time, at larger aberrations, the consideration of higher order polynomials is necessary as follows from above numerical results.

CHAPTER 8

Light scattering by particles - Theory and application to the reflection SIRM

"I do not define time, space, and motion,
as being well known to all."
I. Newton

8.1 Introduction

In previous chapters we have made a detailed analysis of the distribution of the energy present in the region of the probe when light is focussed by a high aperture lens into a silicon specimen. In this chapter we make an assessment of the scattering of light by individual particles present in such a specimen. The aim is to obtain preliminary information concerning the contrast that is likely to occur when individual oxide particles in silicon are imaged using the reflection SIRM.

The fact that the colour of rainbows is due to light scattering by particles was recognised as early as the XIVth. century by Leonardo da Vinci. The pioneering work of Maxwell in 1865 provided the basis for later rigorous scattering theories chiefly developed by Rayleigh[76], Mie[63] and Debye[24]. Comprehensive descriptions and the development of such theories were subsequently published by Bohren and Huffman[9], Born and Wolf[14], Kerker[50] and van de Hulst[101].

Most of these theories considered a plane wave incident on a particle and calculated the scattered intensity as a function of particle size and scattering angle, the results often being plotted in the form of polar diagram intensity distributions. Particles with specific shapes, e.g. spheres, plates or rods, or particles with arbitrary shapes, were considered.

However, although scattering data have been obtained for a range of materials and light wavelength, no results seem yet to have been reported for oxide particles in silicon using infra-red light. Furthermore, the data that was obtained is not generally in a form

which can be used to predict the contrast expected from imaging individual particles using a practical optical system. This is especially the case for the reflection SIRM where a high aperture lens is used to image oxide particles embedded in silicon. This is because in addition to determining the scattering occurring at the particle due to a plane wave, it is necessary to consider the illumination consisting of an aberrated focussed probe within the silicon, the back scattered light traversing the silicon/air interface before re-entering the lens, and the specific angle of acceptance of the scattered light by the lens.

None of the theoretical treatments hitherto made for the scattered light intensities were in a suitable form for assessing these additional effects for a reflection SIRM. Consequently, in the present Chapter a consistent theoretical formulation for the scattered light intensities has been developed, mainly based on the theory developed by Mie[63] and modified by Born and Wolf[14] and Kerker[50]. This formulation gives readily calculable results and can be used as a starting point to assess these additional effects.

The formulation is applied to a plane wave incident on individual spherical oxide particles in silicon. Numerical computations are performed to obtain scattered intensities for light of wavelength $1.3\mu m$, particles in the radius range 25 to 250 nm and scattering angles in the range 0 to 180°, and polar diagram distributions are plotted. These results are new and are relevant to a variety of infra-red scattering methods presently being used to image oxide particles in silicon.

Our analysis continues by considering only the light scattered in the range 180° to 165.9°, which corresponds to the light received by the lens in our reflection SIRM when investigating silicon specimens. The scattered intensities thereby obtained give an initial indication of the contrast expected from individual oxide particles in silicon when using the reflection SIRM.

In Chapter 9 a more rigorous treatment is made of the imaging using this theoretical formulation and the additional effects of the aberrated focussed probe and the silicon/air interface are taken into account.

8.2 Formulation

The method of the solution is as follows. Consider a perfect sphere which is homogeneous and isotropic for a given wavelength of light and which is embedded in a different refractive index material that is also homogeneous and isotropic. We assume that the outer material is non-absorbing and both the outer material and the sphere are non-magnetic. To obtain the result of the classical scattering theory this sphere is illuminated by a linearly polarised monochromatic plane wave. The solution shall be derived from Maxwell's equations in such a way that first these equations are transformed to a spherical polar co-ordinate system after which the general solution of these equations is represented as a combination of a linearly independent infinite series. This can be done since the time independent Maxwell's equations are separable in curvilinear co-ordinates. The coefficients of the series expansion are determined from the boundary conditions, whereby we prescribe the continuity of the tangential components of the electric and the magnetic fields.

With reference to Fig. 8.1 quantities outside the sphere will be denoted by superscript (I) and inside the sphere by superscript (II). Further, superscripts (i), (w) and (s) indicate quantities corresponding to the incident field, the field within the sphere and the scattered field respectively.

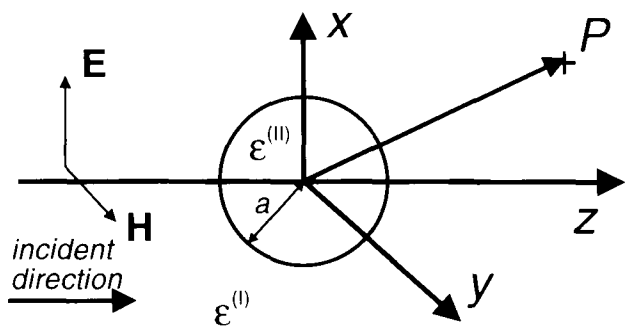


Figure 8.1: Light scattering by a sphere showing the notations used.

We assume in the following the usual time-dependence ($\exp(-i\omega t)$) of the coherent incident radiation. The time independent Maxwell's equations can be written:

$$\left.\begin{aligned} \nabla \times \mathbf{H} &= -k_1 \mathbf{E} \\ \nabla \times \mathbf{E} &= k_2 \mathbf{H} \end{aligned}\right\} \tag{8.1}$$

where

$$\left.\begin{aligned} k_1 &= \frac{\mathrm{i}\omega}{c}\left(\varepsilon + \mathrm{i}\frac{4\pi\sigma}{\omega}\right) \\ k_2 &= \frac{\mathrm{i}\omega}{c} \\ k^2 &= -k_1 k_2 \end{aligned}\right\} \tag{8.2}$$

The centre of the co-ordinate system is at the centre of the sphere and the direction of propagation of the incident light coincides with the positive z direction. The amplitude of the incident electric vector is normalised to unity and its polarisation coincides with the x direction. The components of the incident electric $\mathbf{E}^{(i)}$ and magnetic $\mathbf{H}^{(i)}$ fields are given by:

$$\left.\begin{aligned} E_x^{(i)} &= \mathrm{e}^{\mathrm{i}k^{(I)}z} \\ H_y^{(i)} &= \frac{\mathrm{i}k^{(I)}}{k_2^{(I)}}\mathrm{e}^{\mathrm{i}k^{(I)}z} \\ E_y^{(i)} &= E_z^{(i)} = H_x^{(i)} = H_z^{(i)} = 0. \end{aligned}\right\} \tag{8.3}$$

The only boundary condition is the continuity of the tangential components of the electric and magnetic field on the surface of the sphere:

$$\left.\begin{aligned} E_{tang}^{(I)} &= E_{tang}^{(II)} \\ H_{tang}^{(I)} &= H_{tang}^{(II)} \end{aligned}\right\} \qquad \text{when } r = a \tag{8.4}$$

The boundary conditions are satisfied if

$$\left.\begin{aligned} \mathbf{E} &= \mathbf{E}^{(i)} + \mathbf{E}^{(s)} \qquad \text{outside the sphere} \\ &= \mathbf{E}^{(w)} \qquad\qquad\quad \text{inside the sphere} \end{aligned}\right\} \tag{8.5}$$

We now introduce the spherical polar co-ordinate system:

$$\left.\begin{aligned} x &= r\ \sin\theta\cos\phi \\ y &= r\ \sin\theta\sin\phi \\ z &= r\ \cos\theta \end{aligned}\right\}$$

The co-ordinate transformation that takes any $\mathbf{A}^{(c)}$ vector from the Cartesian system to the spherical polar system ($\mathbf{A}^{(s)}$) is given by:

$$\mathbf{A}^{(s)} = \begin{pmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ \cos\theta\cos\phi & \cos\theta\sin\phi & -\sin\theta \\ -\sin\phi & \cos\phi & 0 \end{pmatrix} \cdot \mathbf{A}^{(c)} \tag{8.6}$$

With the above formulae the *curl* vector operation on $\mathbf{A}^{(s)}$ can readily be obtained (the (s) superscript is omitted):

$$\left.\begin{aligned} (\nabla\times\mathbf{A})_r &= \frac{1}{r^2\sin\theta}\left\{\frac{\partial(rA_\phi\sin\theta)}{\partial\theta} - \frac{\partial(rA_\theta)}{\partial\phi}\right\} \\ (\nabla\times\mathbf{A})_\theta &= \frac{1}{r\sin\theta}\left\{\frac{\partial A_r}{\partial\phi} - \frac{\partial(rA_\phi\sin\theta)}{\partial r}\right\} \\ (\nabla\times\mathbf{A})_\phi &= \frac{1}{r}\left\{\frac{\partial(rA_\theta)}{\partial r} - \frac{\partial A_r}{\partial\theta}\right\} \end{aligned}\right\} \tag{8.7}$$

From the above expression the field equations Eq. (8.1) can be obtained in spherical polar co-ordinates[66]:

$$-k_1 E_r = \frac{1}{r^2\sin\theta}\left\{\frac{\partial(rH_\phi\sin\theta)}{\partial\theta} - \frac{\partial(rH_\theta)}{\partial\phi}\right\} \tag{8.8a}$$

$$-k_1 E_\theta = \frac{1}{r\sin\theta}\left\{\frac{\partial H_r}{\partial\phi} - \frac{\partial(rH_\phi\sin\theta)}{\partial r}\right\} \tag{8.8b}$$

$$-k_1 E_\phi = \frac{1}{r}\left\{\frac{\partial(rH_\theta)}{\partial r} - \frac{\partial H_r}{\partial\theta}\right\} \tag{8.8c}$$

and

$$k_2 H_r = \frac{1}{r^2\sin\theta}\left\{\frac{\partial(rE_\phi\sin\theta)}{\partial\theta} - \frac{\partial(rE_\theta)}{\partial\phi}\right\} \tag{8.9a}$$

$$k_2 H_\theta = \frac{1}{r\sin\theta}\left\{\frac{\partial E_r}{\partial\phi} - \frac{\partial(rE_\phi\sin\theta)}{\partial r}\right\} \tag{8.9b}$$

$$k_2 H_\phi = \frac{1}{r}\left\{\frac{\partial(rE_\theta)}{\partial r} - \frac{\partial E_r}{\partial\theta}\right\} \tag{8.9c}$$

and the boundary conditions of Eq. (8.4) take the form:

$$\left.\begin{aligned} E_\theta^{(I)} &= E_\theta^{(II)}, & E_\phi^{(I)} &= E_\phi^{(II)} \\ H_\theta^{(I)} &= H_\theta^{(II)}, & H_\phi^{(I)} &= H_\phi^{(II)} \end{aligned}\right\} \quad \text{for } r = a. \tag{8.10}$$

Equations (8.8) and (8.9) with the boundary conditions Eq. (8.10) are the basic equations to be solved.

We can represent the solution of the above equations as a superposition of two linearly independent subfields. This condition is consistent with our original assumptions of Eq. (8.3). The two linearly independent subfields are taken to be $({}^{e}\mathbf{E},{}^{e}\mathbf{H})$ and $({}^{m}\mathbf{E},{}^{m}\mathbf{H})$ in such a way that

$$\left.\begin{aligned} {}^{e}E_r = E_r, \qquad & {}^{e}H_r = 0 \\ {}^{m}E_r = 0, \qquad & {}^{m}H_r = H_r \end{aligned}\right\}$$

By setting ${}^{e}H_r = H_r = 0$, substituting ${}^{e}E_r$ into Eqs. (8.9b) and (8.9c), and after some further calculations we obtain:

$$\left.\begin{aligned} \left(\frac{\partial^2}{\partial r^2} + k^2\right)(r\ {}^{e}H_\theta) &= -\frac{k_1}{\sin\theta}\frac{\partial {}^{e}E_r}{\partial \phi} \\ \left(\frac{\partial^2}{\partial r^2} + k^2\right)(r\ {}^{e}H_\phi) &= +k_1\frac{\partial {}^{e}E_r}{\partial \theta} \end{aligned}\right\} \tag{8.11}$$

The above equations with Eq. (8.8a) give a system of equations for ${}^{e}E_r$, ${}^{e}H_\theta$ and ${}^{e}H_\phi$. We have also to consider a subsidiary condition, $\nabla \cdot {}^{e}\mathbf{H} = 0$. With this condition Eq. (8.9a) takes the form:

$$0 = \frac{1}{k_1^2 r^2 \sin\theta}\ \frac{\partial}{\partial r}\left[\frac{\partial}{\partial \theta}(r\sin\theta\ {}^{e}H_\theta) + \frac{\partial}{\partial \phi}(r\ {}^{e}H_\phi)\right]$$

and we have a similar equation for the case when ${}^{m}E_r = 0$. When the radial component of the electric field vanishes the wave given by the solution is called the *magnetic* or *transverse electric* (TE) wave (i.e. the field possesses only transverse electric components), and the same applies to the radial component of the magnetic field for which the solution is called the *electric* or *transverse magnetic* (TM) wave.

The solution presented above is implied in the solution of Kerker[50] and is given explicitly in Born and Wolf[14]. At this point Kerker introduces the Herz vector on which the solution of the problem is obtained, whilst Born and Wolf use the Debye scalar potential. Since the solution to be presented here does not require the introduction of the Herz vector, we shall follow Born and Wolf. From the Debye (or Debye-Herz) potentials ${}^{e}\Pi$ and ${}^{m}\Pi$ it is possible to derive both the TE and the TM waves. Equation (8.9a) shows

that when ${}^eH_r = 0$, ${}^eE_\theta$ and ${}^eE_\phi$ can be represented as gradients of scalar quantities:

$$ {}^eE_\theta = \frac{1}{r}\frac{\partial U}{\partial \theta} \qquad {}^eE_\phi = \frac{1}{r\sin\theta}\frac{\partial U}{\partial \phi} \tag{8.12} $$

into which on substitution of

$$ U = \frac{\partial (r\ {}^e\Pi)}{\partial r} \tag{8.13} $$

we obtain

$$ {}^eE_\theta = \frac{1}{r}\frac{\partial^2 (r\ {}^e\Pi)}{\partial r \partial \theta} \qquad {}^eE_\phi = \frac{1}{r\sin\theta}\frac{\partial^2 (r\ {}^e\Pi)}{\partial r \partial \phi}. \tag{8.14} $$

By using the above formulae the solution for the field equations can be obtained:

$$ E_r = \frac{\partial^2 (r\ {}^e\Pi)}{\partial r^2} + k^2 r\ {}^e\mathrm{H} \tag{8.15a} $$

$$ E_\theta = \frac{1}{r}\frac{\partial^2 (r\ {}^e\Pi)}{\partial r \partial \theta} + k_2 \frac{1}{r\sin\theta}\frac{\partial (r\ {}^m\Pi)}{\partial \phi} \tag{8.15b} $$

$$ E_\phi = \frac{1}{r\sin\theta}\frac{\partial^2 (r\ {}^e\Pi)}{\partial r \partial \phi} - k_2 \frac{1}{r}\frac{\partial (r\ {}^m\Pi)}{\partial \theta} \tag{8.15c} $$

and

$$ H_r = k^2 r\ {}^m\Pi + \frac{\partial^2 (r\ {}^m\Pi)}{\partial r^2} \tag{8.16a} $$

$$ H_\theta = -k_1 \frac{1}{r\sin\theta}\frac{\partial (r\ {}^e\Pi)}{\partial \phi} + \frac{1}{r}\frac{\partial^2 (r\ {}^m\Pi)}{\partial r \partial \theta} \tag{8.16b} $$

$$ H_\phi = k_1 \frac{1}{r}\frac{\partial (r\ {}^e\Pi)}{\partial \theta} + \frac{1}{r\sin\theta}\frac{\partial^2 (r\ {}^m\Pi)}{\partial r \partial \phi} \tag{8.16c} $$

Both of the Debye potentials satisfy the wave equation:

$$ \nabla^2 \Pi + k^2 \Pi = 0 $$

and the boundary condition Eq. (8.4) requires the continuity of four quantities:

$$ k_1 r\ {}^e\Pi, \qquad k_2 r\ {}^m\Pi, \qquad \frac{\partial}{\partial r}(r\ {}^e\Pi), \qquad \frac{\partial}{\partial r}(r\ {}^m\Pi) \tag{8.17} $$

At this point we come to an important stage of the derivation. It must be emphasized that with Eqs. (8.15), (8.16) and (8.17) we have independent conditions for ${}^e\Pi$ and ${}^m\Pi$. Thus the solutions of the problem can be considered as two mutually independent solutions of the wave equation.

8.3 Solution of the wave equation

As it has been mentioned and is well-known the time independent wave equation is separable in the spherical polar co-ordinate system. Thus the solution of the *scalar* wave equation can be represented as an infinite series of linearly independent functions. The form of these functions follows from a set of ordinary differential equations. The coefficients of the series can be determined from the boundary conditions. The solution of the wave equation is in the following form:

$$\Pi = R(r)\Theta(\theta)\Phi(\phi) \tag{8.18}$$

where R, Θ and Φ must satisfy the ordinary differential equations:

$$\frac{\mathrm{d}^2(rR)}{\mathrm{d}r^2} + \left(k^2 - \frac{l(l+1)}{r^2}\right) rR = 0 \tag{8.19a}$$

$$\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}\Theta}{\mathrm{d}\theta}\right) + \left(l(l+1) - \frac{m^2}{\sin^2\theta}\right)\Theta = 0 \tag{8.19b}$$

$$\frac{\mathrm{d}^2\Phi}{\mathrm{d}\phi^2} + m^2\Phi = 0 \tag{8.19c}$$

where $l > |m|$ and are both integers*.

The solutions of the radial equation (8.19a) are the Riccati-Bessel functions†:

$$\psi_n(kr) = \sqrt{\frac{\pi kr}{2}}\, J_{n+\frac{1}{2}}(kr)$$

$$\chi_n(kr) = -\sqrt{\frac{\pi kr}{2}}\, N_{n+\frac{1}{2}}(kr)$$

Since the Neumann functions possess a singularity at $kr = 0$, as shown in Appendix A, they cannot represent the field *inside* the sphere. These functions are however ideal for the representation of the field *outside* the sphere when the vanishing field at infinity is imposed. It follows from Eq. (8.19) that the general solution is given by:

$$rR = c_l\, \psi_l(kr) + d_l\, \chi_l(kr). \tag{8.20a}$$

*The proof of the solution of Eqs. (8.19a) and (8.19b) is given in Appendix B.
†Some important properties of these and other special functions are given in Appendix A.

where c_l and d_l are constants and they can be determined from the boundary conditions. The solutions of Eq. (8.19b) are the associated Legendre functions

$$\Theta = P_l^{(m)}(\cos\theta) \tag{8.20b}$$

of degree l and order m where $l = m, m+1, \ldots$ [†]. Finally, the linearly independent solutions of Eq. (8.19c) are given by:

$$\Phi = a_m \cos(m\phi) + b_m \sin(m\phi). \tag{8.20c}$$

From Eq. (8.20) follows the general solution of the wave equation:

$$r\Pi = r\sum_{l=0}^{\infty}\sum_{m=-l}^{l}\{c_l\psi_l(kr) + d_l\chi_l(kr)\}\left\{P_l^{(m)}(\cos\theta)\right\}\{a_m\cos(m\phi) + b_m\sin(m\phi)\}. \tag{8.21}$$

It can be shown (see e.g. Born and Wolf[14]) that the incident plane wave can be expanded in terms of the scalar Debye potentials. This can be done by transforming the incident wave to the spherical polar co-ordinate system (Eq. (8.6)) and equating one of the corresponding terms in Eq. (8.15). The expression for the *incident* wave thus takes the form:

$$r^e\Pi^{(i)} = \frac{1}{(k^{(I)})^2}\sum_{l=1}^{\infty}\mathrm{i}^{(l-1)}\frac{2l+1}{l(l+1)}\psi_l(k^{(I)}r)P_l^{(1)}(\cos\theta)\cos\phi \tag{8.22a}$$

and

$$r^m\Pi^{(i)} = \frac{\mathrm{i}}{(k^{(I)})^2}\sum_{l=1}^{\infty}\mathrm{i}^{(l-1)}\frac{k^{(I)}}{k_2^{(II)}}\frac{2l+1}{l(l+1)}\psi_l(k^{(I)}r)P_l^{(1)}(\cos\theta)\sin\phi \tag{8.22b}$$

It also follows that the *scattered* field can be expanded in terms of scalar Debye potentials. Hence the scattered field:

$$r^e\Pi^{(s)} = \frac{1}{(k^{(I)})^2}\sum_{l=1}^{\infty}\mathrm{i}^{(l-1)}\frac{2l+1}{l(l+1)}a_l\,\zeta_l^{(1)}(k^{(I)}r)P_l^{(1)}(\cos\theta)\cos\phi \tag{8.23a}$$

and

$$r^m\Pi^{(s)} = \frac{\mathrm{i}}{k^{(I)}k_2^{(II)}}\sum_{l=1}^{\infty}\mathrm{i}^{(l-1)}\frac{2l+1}{l(l+1)}b_l\,\zeta_l^{(1)}(k^{(I)}r)P_l^{(1)}(\cos\theta)\sin\phi \tag{8.23b}$$

On substituting these results into Eq. (8.15) we obtain the components of the electric field:

$$E_r^{(s)} = \frac{1}{(k^{(I)})^2}\frac{\cos\phi}{r^2}\sum_{l=1}^{\infty}\mathrm{i}^{l+1}(2l+1)a_l\,\zeta_l^{(1)}(k^{(I)}r)P_l^{(1)}(\cos\theta) \tag{8.24a}$$

$$E_\theta^{(s)} = -\frac{1}{k^{(I)}}\frac{\cos\phi}{r}\sum_{l=1}^{\infty} \mathrm{i}^{l+1}\frac{2l+1}{l(l+1)}\left\{a_l\left[\zeta_l^{(1)}(k^{(I)}r)\right]'\left[P_l^{(1)}(\cos\theta)\right]'\sin\theta - \mathrm{i}b_l\,\zeta_l^{(1)}(k^{(I)}r)P_l^{(1)}(\cos\theta)\frac{1}{\sin\theta}\right\} \tag{8.24b}$$

$$E_\phi^{(s)} = -\frac{1}{k^{(I)}}\frac{\sin\phi}{r}\sum_{l=1}^{\infty} \mathrm{i}^{l+1}\frac{2l+1}{l(l+1)}\left\{a_l\left[\zeta_l^{(1)}(k^{(I)}r)\right]'P_l^{(1)}(\cos\theta)\frac{1}{\sin\theta} - \mathrm{i}b_l\,\zeta_l^{(1)}(k^{(I)}r)\left[P_l^{(1)}(\cos\theta)\right]'\sin\theta\right\} \tag{8.24c}$$

where *prime* indicates the differentiation of the function with respect to their arguments and we have a similar set of equations for $\mathbf{H}^{(s)}$. It will now be advantageous to introduce two relative quantities. The *relative refractive index* is given by:

$$\hat{n}^2 = \left(\frac{\hat{n}^{(II)}}{n^{(I)}}\right)^2 = \left(\frac{k^{(II)}}{k^{(I)}}\right)^2 = \frac{\varepsilon^{(II)}}{\varepsilon^{(I)}} + \mathrm{i}\frac{4\pi\sigma}{\omega\varepsilon^{(I)}}, \tag{8.25}$$

and the *size parameter* is given by:

$$q = \frac{2\pi}{\lambda^{(I)}}a \tag{8.26}$$

where a is the particle radius.

With these quantities the coefficients a_l and b_l can be written[14]:

$$a_l = \frac{\hat{n}[\psi_l(q)]'\psi_l(\hat{n}q) - \psi_l(q)[\psi_l(\hat{n}q)]'}{\hat{n}[\zeta_l^{(1)}(q)]'\psi_l(\hat{n}q) - \zeta_l^{(1)}(q)[\psi_l(\hat{n}q)]'} \tag{8.27a}$$

$$b_l = \frac{\hat{n}[\psi_l(\hat{n}q)]'\psi_l(q) - \psi_l(\hat{n}q)[\psi_l(q)]'}{\hat{n}\zeta_l^{(1)}(q)[\psi_l(\hat{n}q)]' - [\zeta_l^{(1)}(q)]'\psi_l(\hat{n}q)} \tag{8.27b}$$

Equations (8.24) and (8.27) constitute the solution of the problem.

8.4 Far field solution

If, as in most cases, the scattered field is to be of interest many wavelengths away from the particle, then mathematical approximations can be applied to simplify the formalism without introducing appreciable errors. Equation (8.24a) shows that the radial component of the electric (and also the magnetic) vector decays more rapidly than the transverse components[‡]. Thus, at the far field solution the electric and magnetic vectors consist

[‡]This is also the case when an arbitrary particle is considered (see e.g. Jackson[46]).

only of transverse components. Furthermore, we can also approximate the Riccati-Bessel functions (see Appendix A) by their asymptotic series:

$$\zeta_l^{(1)}(x) \sim (-\mathrm{i})^{(l+1)}\mathrm{e}^{\mathrm{i}x} \tag{8.28a}$$

$$\left[\zeta_l^{(1)}(x)\right]' \sim (-\mathrm{i})^{l}\mathrm{e}^{\mathrm{i}x} \tag{8.28b}$$

and we introduce[§]:

$$\pi_n(\cos\theta) = \frac{P_n^{(1)}(\cos\theta)}{\sin\theta} \tag{8.28c}$$

$$\tau_n(\cos\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta}P_n^{(1)}(\cos\theta) \tag{8.28d}$$

On substituting from Eq. 8.28 into Eq. 8.24 the electric field is given by:

$$E_\theta^{(s)} \cong \cos\phi\frac{\mathrm{e}^{\mathrm{i}k^{(I)}r}}{-\mathrm{i}k^{(I)}r}\sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)}\left[a_l\ \tau_l(\cos\theta) + b_l\ \pi_l(\cos\theta)\right] \tag{8.29a}$$

$$E_\phi^{(s)} \cong -\sin\phi\frac{\mathrm{e}^{\mathrm{i}k^{(I)}r}}{-\mathrm{i}k^{(I)}r}\sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)}\left[a_l\ \pi_l(\cos\theta) + b_l\ \tau_l(\cos\theta)\right] \tag{8.29b}$$

or by introducing

$$S_1(\cos\theta) = \sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)}\left[a_l\ \tau_l(\cos\theta) + b_l\ \pi_l(\cos\theta)\right] \tag{8.30a}$$

$$S_2(\cos\theta) = \sum_{l=1}^{\infty}\frac{2l+1}{l(l+1)}\left[a_l\ \pi_l(\cos\theta) + b_l\ \tau_l(\cos\theta)\right] \tag{8.30b}$$

the scattered electric field far from the particle is given by:

$$\left.\begin{aligned} E_\theta^{(s)} &\cong \frac{\mathrm{e}^{\mathrm{i}k^{(I)}r}}{-\mathrm{i}k^{(I)}r}\cos\phi S_2(\cos\theta) \\ E_\phi^{(s)} &\cong -\frac{\mathrm{e}^{\mathrm{i}k^{(I)}r}}{-\mathrm{i}k^{(I)}r}\sin\phi S_1(\cos\theta) \end{aligned}\right\}. \tag{8.31}$$

It should be noted that the results as shown by Eqs. (8.31), (8.30) and (8.27) are identical to those obtained by Bohren and Huffman[9], although the derivation used by the latter workers was more complex. Alternatively we can write

$$\begin{pmatrix} E_\parallel^{(i)} \\ E_\perp^{(i)} \end{pmatrix} = \frac{\mathrm{e}^{\mathrm{i}k^{(I)}(r-z)}}{-\mathrm{i}k^{(I)}r}\begin{pmatrix} S_2 & 0 \\ 0 & S_1 \end{pmatrix}\begin{pmatrix} E_\parallel^{(i)} \\ E_\perp^{(i)} \end{pmatrix} \tag{8.32}$$

[§]The function τ_n should not be confused with the Fresnel coefficients.

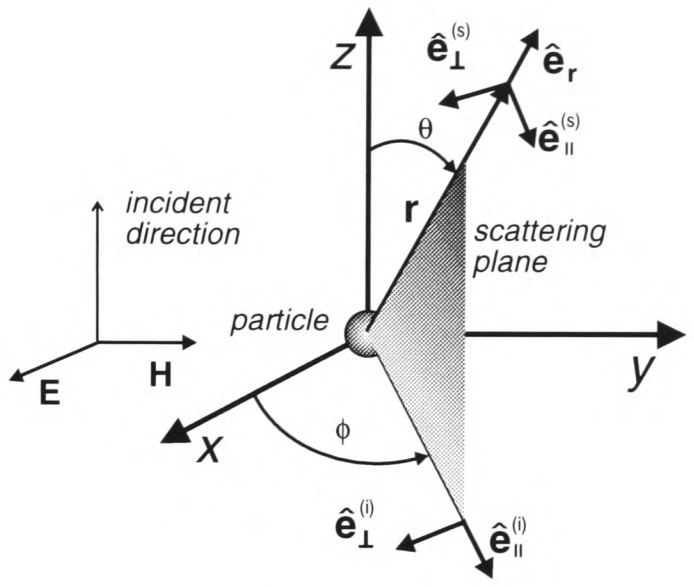


Figure 8.2: The scattering plane

where $E_{||}$ and $E_\perp$ are the parallel and perpendicular components of the corresponding electric vector with respect to the scattering plane. The scattering plane, as shown in Fig. 8.2, is the plane which contains the point of observation and the z axis (the exact definition requires the point of observation and the propagation direction of the incident illumination to be in the plane). These $\parallel$ and $\perp$ directions, therefore, are not based on a steady, but rather a continuously varying, co-ordinate system.

Equation (8.32) shows that the scattered electric field depends linearly on the incident field, and this statement is true for particles of any shape. Furthermore, Eq. 8.32 shows that the electric and magnetic fields at sufficiently large distances from the scatterer are spherical waves.

From the parameters S_1 and S_2 the Stokes parameters can easily be obtained:

$$\begin{pmatrix} I^{(s)} \\ Q^{(s)} \\ U^{(s)} \\ V^{(s)} \end{pmatrix} = \frac{1}{\left(k^{(I)}\right)^2 r^2} \begin{pmatrix} S_{11} & S_{12} & 0 & 0 \\ S_{12} & S_{11} & 0 & 0 \\ 0 & 0 & S_{33} & S_{34} \\ 0 & 0 & -S_{34} & S_{33} \end{pmatrix} \begin{pmatrix} I^{(i)} \\ Q^{(i)} \\ U^{(i)} \\ V^{(i)} \end{pmatrix} \tag{8.33}$$

where

$$\left.\begin{aligned} S_{11} &= \frac{1}{2}\left(|S_2|^2 + |S_1|^2\right), & S_{12} &= \frac{1}{2}\left(|S_2|^2 - |S_1|^2\right) \\ S_{33} &= \frac{1}{2}\left(S_1 S_2^* + S_2 S_1^*\right), & S_{34} &= \frac{\mathrm{i}}{2}\left(S_1 S_2^* - S_2 S_1^*\right) \end{aligned}\right\}. \tag{8.34}$$

When the polarisation of the incident light is parallel to the scattering plane the Stokes

parameters of the scattered light are (omitting the radial factor):

$$I^{(s)} = (S_{11} + S_{12})I^{(i)} = |S_2|^2 I^{(i)} = i_{||} I^{(i)}, \quad Q^{(s)} = I^{(s)}, \quad U^{(s)} = V^{(s)} = 0 \tag{8.35}$$

Thus, the polarisation of the scattered light is the same as that of the incident light. For the 90° polarisation, i.e. when the polarisation of the incident light is perpendicular to the scattering plane:

$$I^{(s)} = (S_{11} - S_{12})I^{(i)} = |S_1|^2 I^{(i)} = i_{\perp} I^{(i)}, \quad Q^{(s)} = -I^{(s)}, \quad U^{(s)} = V^{(s)} = 0 \tag{8.36}$$

Thus, the polarisation of the scattered light is the same as that of the incident light. When the incident light is unpolarised:

$$I^{(s)} = S_{11} I^{(i)}, \qquad Q^{(s)} = S_{12} I^{(i)}, \qquad U^{(s)} = V^{(s)} = 0 \tag{8.37}$$

and the polarisation ratio

$$P = -\frac{S_{12}}{S_{11}} = \frac{i_{\perp} - i_{||}}{i_{\perp} + i_{||}}. \tag{8.38}$$

The scattering cross-section is defined as

$$C_{sca} = \frac{2\pi}{\left(k^{(I)}\right)^2} \sum_{n=1}^{\infty} (2n+1) \left(|a_n|^2 + |b_n|^2 \right) \tag{8.39}$$

from which the scattering coefficient Q_{sca} can be calculated as $Q_{sca} = C_{sca}/\pi a^2$.

8.5 Numerical results

From the above theoretical considerations, considerable information can be obtained concerning spherical particles (scatterers) illuminated by a linearly or circularly polarised plane wave. In the case of confocal microscopy, however, the particles are illuminated by a focussed light beam. It is well known that for focussed illumination the single plane wave approach only gives reliable results when the scatterer is situated on the optical axis and exactly in the focal plane. It is therefore reasonable initially to say that the present considerations are valid when the particle is situated *exactly* in the focus and the illumination is not distorted by a substantial amount of spherical aberration¶.

¶Imposing the weak spherical aberration is not a strict condition, since the paraxial plane wave component is not distorted at any focussing depths.

It is expected from these preliminary calculations to obtain a basic understanding of the contrast mechanism of the SIRM. It will be shown in Chapter 9 that a semi-geometrical optics approach can describe the contrast mechanism of the SIRM with good accuracy when the illumination of the scatterer is different from the one imposed here.

Numerical computations of formulae Eqs. (8.27), (8.30), (8.35), (8.36), (8.37), (8.38) and Q_{sca} have been performed and the detailed results are given below.

A simple contrast mechanism is considered as follows. We wish to determine the SIRM contrast (detected signal strength) as a function of the particle size. It is assumed that the scatterer is situated exactly in focus. The probe-forming lens collects the scattered light in the backward direction ($\sim 180°$)[‖].

Given the numerical aperture of the probe-forming lens (NA_{air}) for air, the corresponding semi-angle in the second material is determined by Snell's law (with the usual notations):

$$\varphi_{eff} = \arcsin\left[\frac{n_1}{n_2}\sin\varphi_{air}\right] = \arcsin\left[\frac{n_1}{n_2}NA_{air}\right]$$

In the absence of the interface (and the second material) and by ignoring the polarisation dependent transmittance at the surface of the lens, our exact approach would be to integrate the electric vector over the solid angle φ_{air}. When the interface is present the effective numerical aperture is decreased (or increased depending on the ratio of the two refractive indices) as described by Snell's law. To calculate the light that can be collected by the probe-forming lens, the relative intensity of the scattered light within the effective numerical aperture is summed for different angles. The quantity thus computed will be referred to as the *angle dependent scattering coefficient* and denoted by $\Omega_{sca}(\alpha, q)$:

$$\Omega_{sca}(\alpha, q) = C \int_{\pi-\alpha}^{\pi} \left\{|S_1(\cos\theta, q)|^2 + |S_2(\cos\theta, q)|^2\right\} \mathrm{d}\theta \tag{8.40}$$

where α determines the solid angle in such a way that it gives the angle within which the light is detected with respect to the 180° backward direction, q is the size parameter given by Eq. (8.26) and C is a constant. For $NA_{air} = 0.85$ and an air/silicon interface, $\alpha = 14.1°$.

[‖]The effect on the back-scattered light of traversing the material/air interface before entering the lens is not considered here but is discussed in Chapter 9.

It is worth mentioning that numerical computation of formulae such as Eqs. (8.27), (8.30), (8.35), (8.36), (8.37), (8.38) is not without difficulties. It is not the aim of the present work to investigate the numerical stability of different recursion formulae, but when computer programmes are applied to a certain problem, the programmer must make sure that the computed formula is convergent. For example, the recurrence relation of the spherical Bessel function j_n of Eq. (A.12) is unstable for upward, but is stable for downward, recurrence (see e.g. Abramowitz and Stegun[1]). This means that first an estimated small value is generated and then the program computes from this small initial value the larger but more precise lower order terms. As another example, the spherical Neumann function n_n, which satisfies the same recurrence relation Eq. (A.12), is stable for upward recurrence. An excellent review of different computational methods is due to Bohren and Huffman[9] which suggests the use of a logarithmic derivative when the first derivatives of two of the Riccati-Bessel functions ($\zeta_n^{(1)}(x)$ and $\psi_n(x)$) are to be computed:

$$D_n(\rho) = \frac{\mathrm{d}}{\mathrm{d}\rho} \ln \psi_n(\rho).$$

With the help of the above equation the scattering coefficients of Eq. (8.27) can be re-written as:

$$a_n = \frac{[D_n(\hat{n}q)/\hat{n} + n/q]\psi_n(q) - \psi_{n-1}(q)}{[D_n(\hat{n}q)/\hat{n} + n/q]\zeta_n^{(1)}(q) - \zeta_{n-1}^{(1)}(q)} \tag{8.41a}$$

$$b_n = \frac{[\hat{n}D_n(\hat{n}q) + n/q]\psi_n(q) - \psi_{n-1}(q)}{[\hat{n}D_n(\hat{n}q) + n/q]\zeta_n^{(1)}(q) - \zeta_{n-1}^{(1)}(q)} \tag{8.41b}$$

where the following recurrence relations are used:

$$\psi_n'(x) = \psi_{n-1}(x) - \frac{n\psi_n(x)}{x}, \qquad \left[\zeta_n^{(1)}(x)\right]' = \zeta_{n-1}^{(1)}(x) - \frac{\zeta^{(1)}n_n(x)}{x}. \tag{8.41c}$$

Additionally the logarithmic derivative satisfies the recurrence relation:

$$D_{n-1} = \frac{n}{\rho} - \frac{1}{D_n + n/\rho}$$

which is more stable for downward recurrence. By obtaining the scattering coefficients a_n and b_n, all of the remaining quantities can be readily and efficiently computed.

To perform a straightforward check on the mathematical model, the scattered intensity $I^{(s)}$ of a SiO_2 particle of radius $a = 2.5\,\text{nm}$ was computed as a function of scattering angle for a wavelength of $\lambda_0 = 1.3\,\mu\text{m}$. The refractive index of the silicon was taken as $n^{(I)} = 3.5$ and that of the SiO_2 particle as $\hat{n}^{(II)} = 1.45 + 0\,\text{i}$ (the conductivity is zero, i.e. $\sigma^{(II)} = 0$), which is the refractive index of amorphous SiO_2 for λ_0. This computation was intended to provide an initial verification of the model since this particle size is in the Rayleigh scattering range. It is known that when Rayleigh scattering occurs a spherical particle should give an angular scattering distribution identical to that of a radiating dipole. This fact also follows from our mathematical model because when $q \ll 1$ only the first term of the series is dominant in Eq. (8.24), which then takes a form identical to that of a radiating dipole (for a proof see Born and Wolf[14]). The results of the computation are shown in Fig. 8.3, plotted on both linear and log scales. The curves correspond precisely to those expected from a radiating dipole, namely, constant intensity for perpendicular polarisation and intensity decreasing to virtually zero at 90° for parallel polarisation, with the curve perfectly symmetrical about 90°.

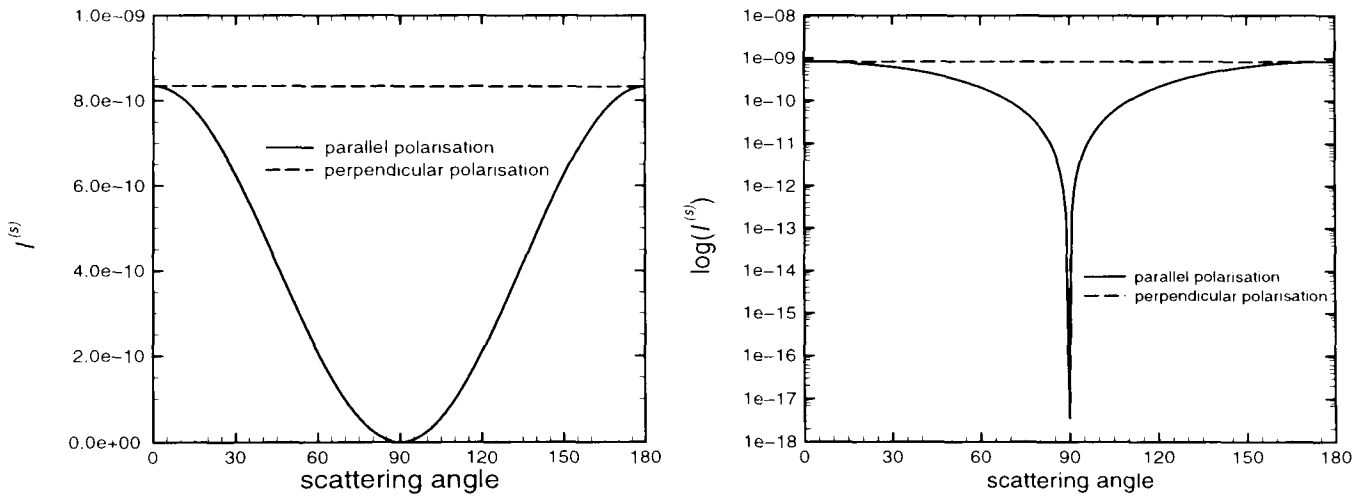


Figure 8.3: Scattering diagram of a SiO_2 particle, $a = 2.5\,\text{nm}$

A series of similar calculations were performed for particle radii $a = 25$ to $250\,\text{nm}$. The corresponding $i_{||}I^{(i)}$ and $i_{\perp}I^{(i)}$ intensities ($I^{(i)}$ is taken to be unity) are shown in Figs. 8.4 to 8.13.

The range of particle radius was chosen to cover the size of the most frequently encountered particles in practice (see Chapter 10). It is shown in the above Figures that as the

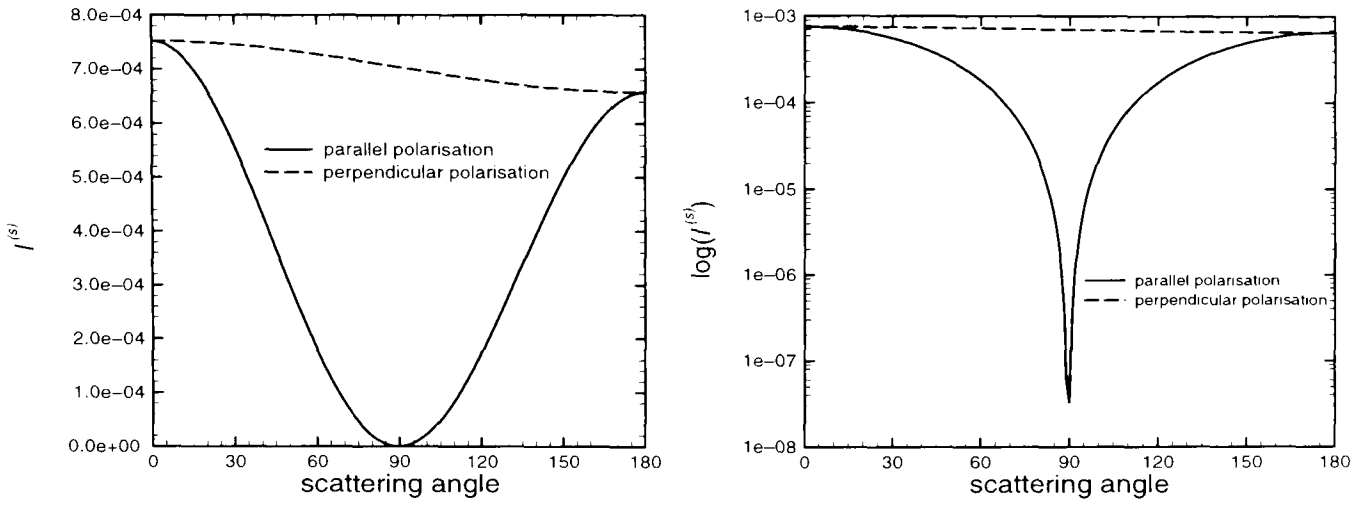


Figure 8.4: Scattering diagram of a SiO_2 particle, $a = 25\,\mathrm{nm}$

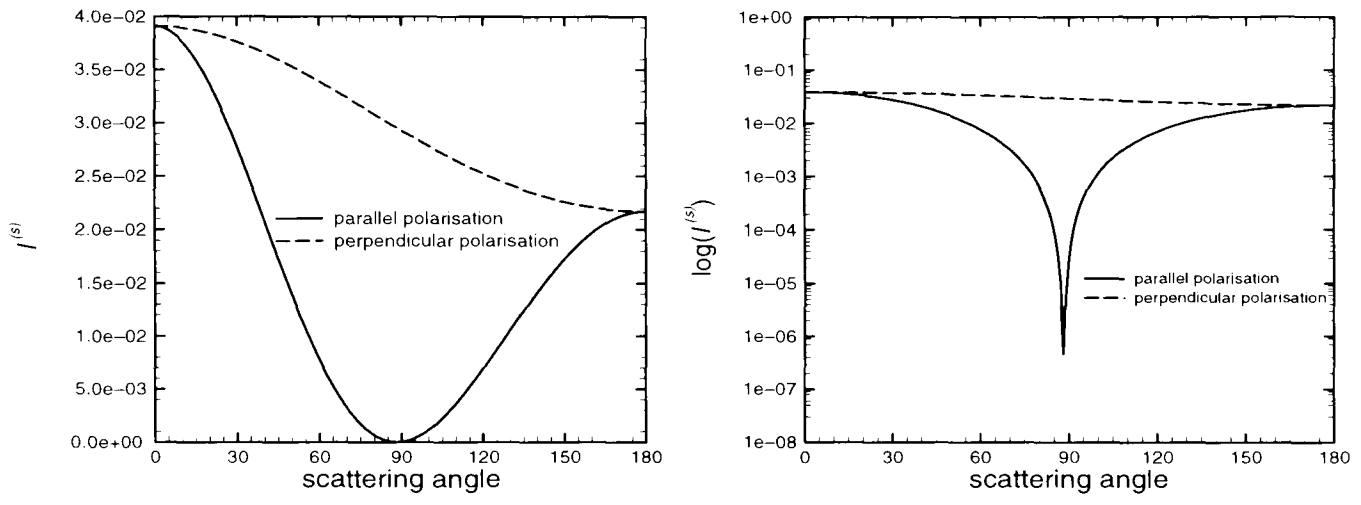


Figure 8.5: Scattering diagram of a SiO_2 particle, $a = 50\,\mathrm{nm}$

particle radius becomes larger, the scattering diagrams progressively change from symmetrical to asymmetrical, eventually exhibiting a much stronger scattering in the forward (0°) direction.

As seen in the logarithmically scaled Figures, as the particle radius becomes larger up to $a = 150\mathrm{nm}$, the minimum of the scattered $i_{||}$ intensity gradually shifts from 90° towards lower angles. At larger particle radii the intensity corresponding to the minimum becomes close or equal to the intensity corresponding to the backward (180°) direction. This behaviour of the backward scattering is further emphasized in Fig. 8.14, where the normalised scattered intensity corresponding to unpolarised illumination is shown on a linear scale as a function of the scattering angle for different particle radii.

Having described the *relative* intensities of the light scattering by SiO_2 particles, we turn our attention to *absolute* intensities. The absolute polarised scattered intensity was

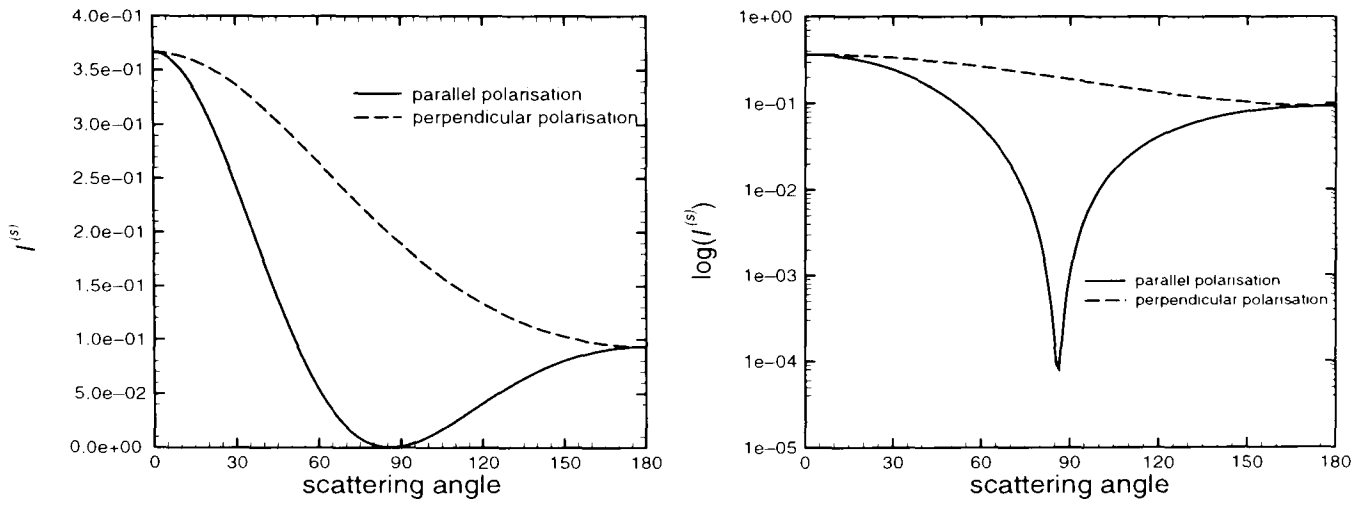


Figure 8.6: Scattering diagram of a SiO_2 particle, $a = 75\,\mathrm{nm}$

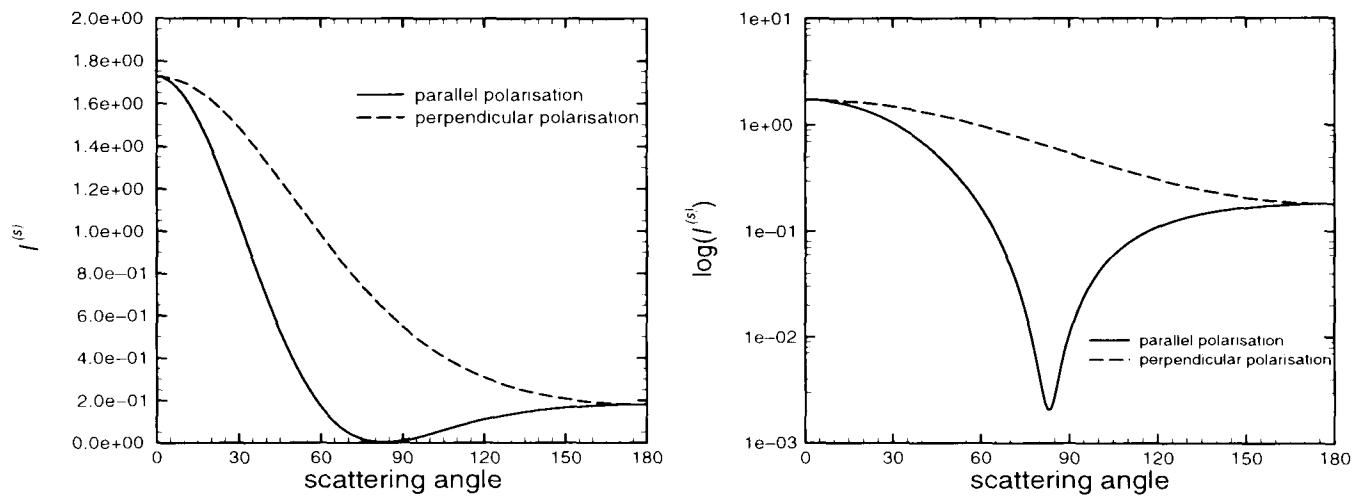


Figure 8.7: Scattering diagram of a SiO_2 particle, $a = 100\,\mathrm{nm}$

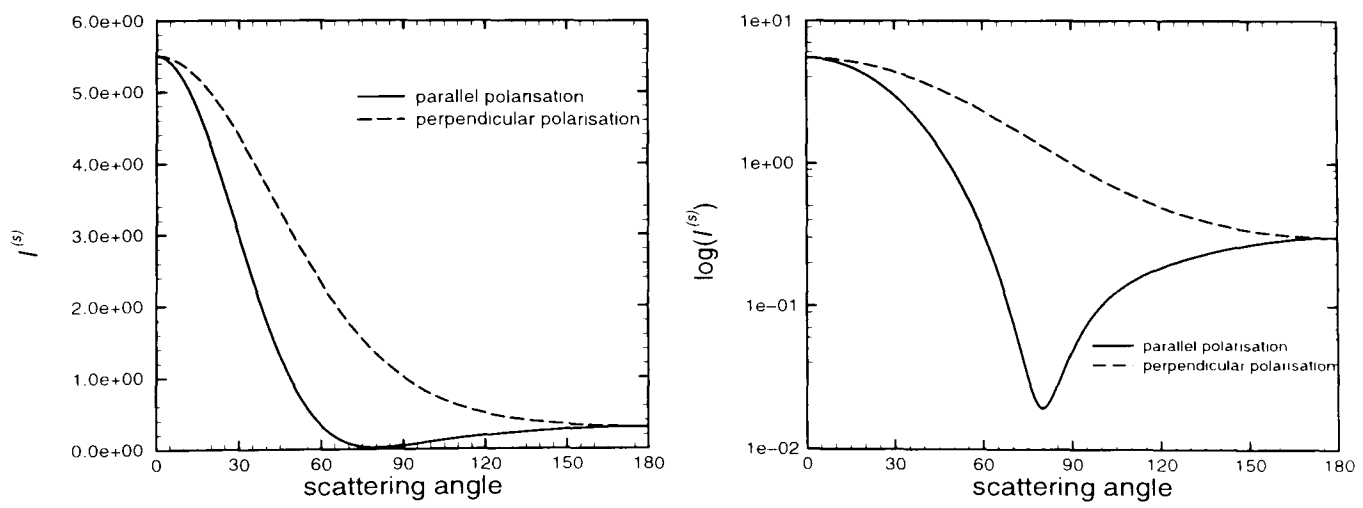


Figure 8.8: Scattering diagram of a SiO_2 particle, $a = 125\,\mathrm{nm}$

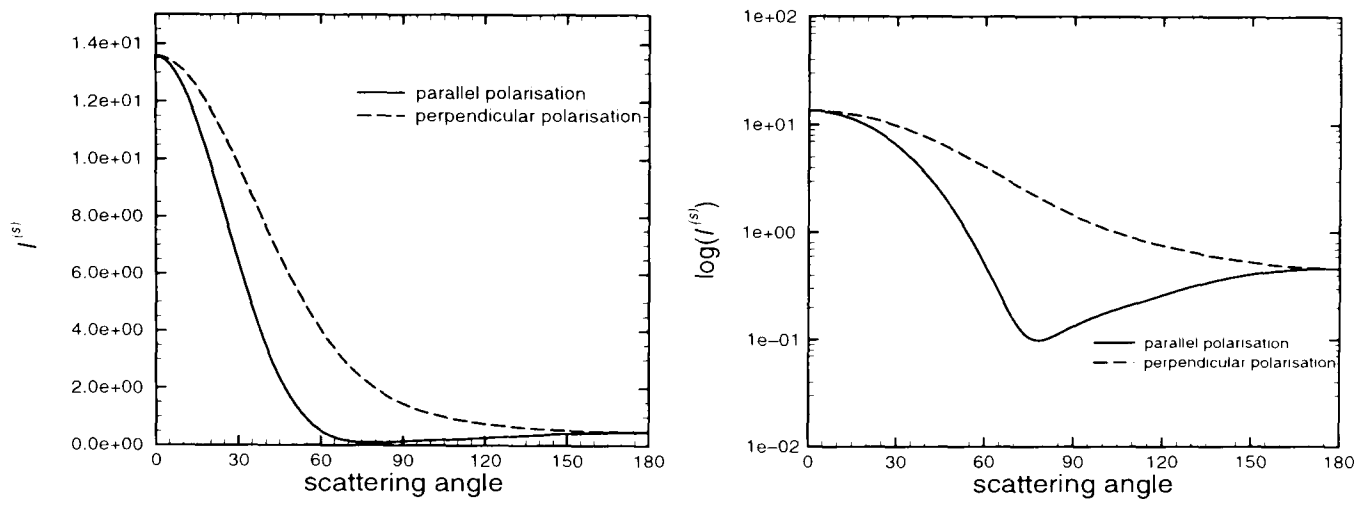


Figure 8.9: Scattering diagram of a SiO_2 particle, $a = 150\,\mathrm{nm}$

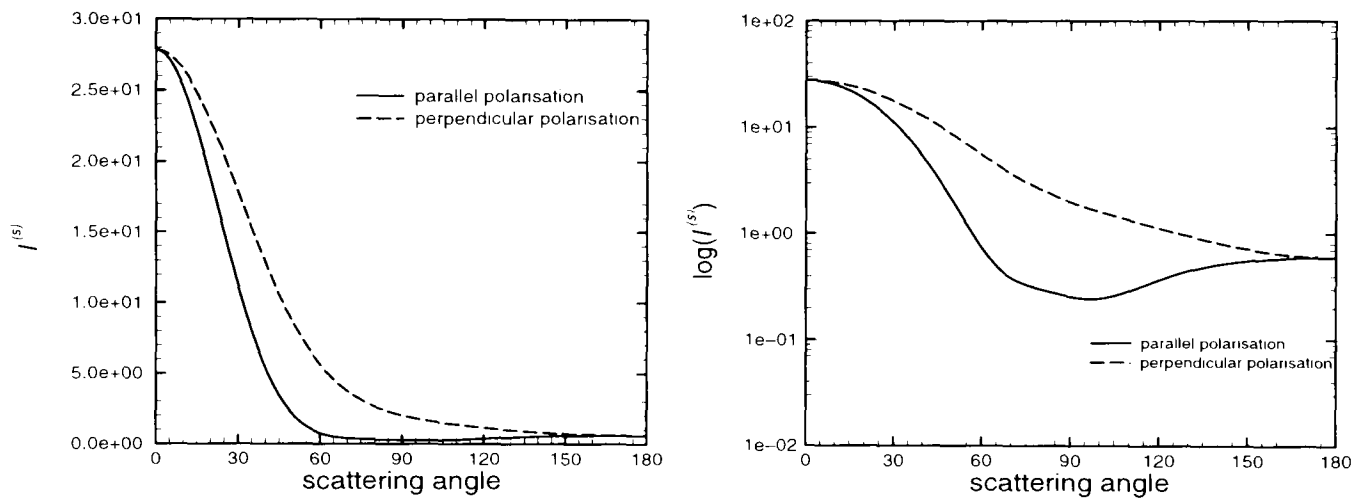


Figure 8.10: Scattering diagram of a SiO_2 particle, $a = 175\,\mathrm{nm}$

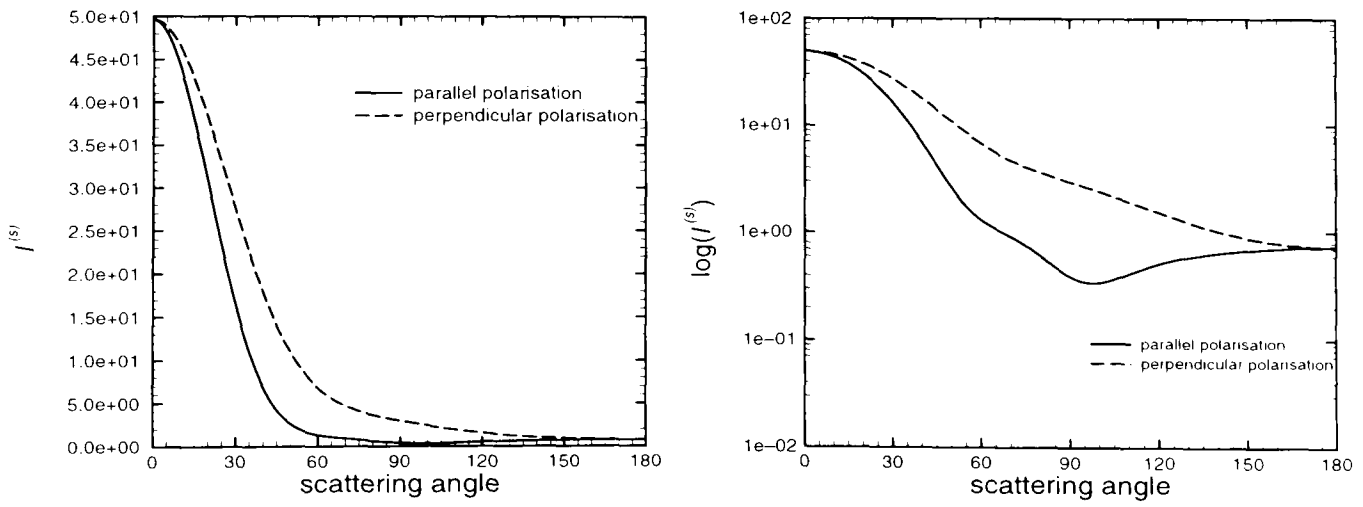


Figure 8.11: Scattering diagram of a SiO_2 particle, $a = 200\,\mathrm{nm}$

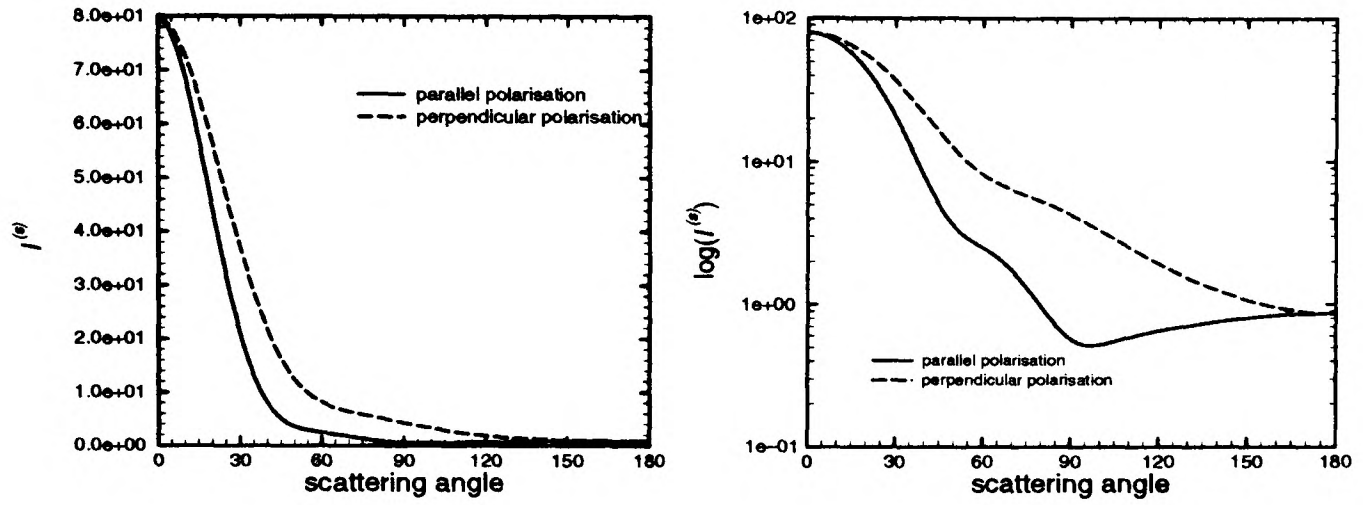


Figure 8.12: Scattering diagram of a SiO_2 particle, $a = 225\,\mathrm{nm}$

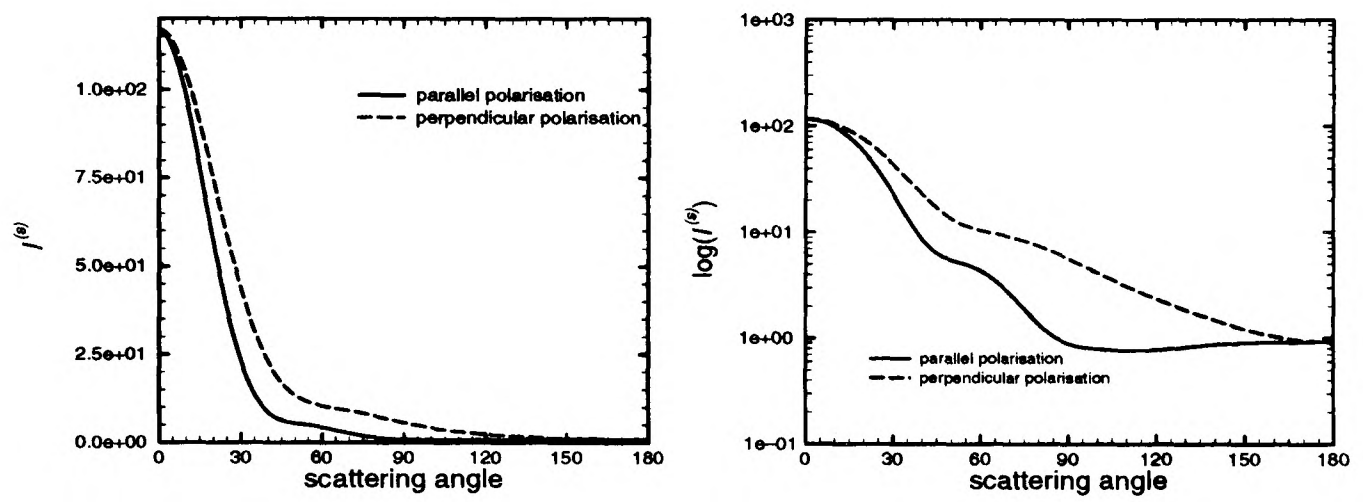


Figure 8.13: Scattering diagram of a SiO_2 particle, $a = 250\,\mathrm{nm}$

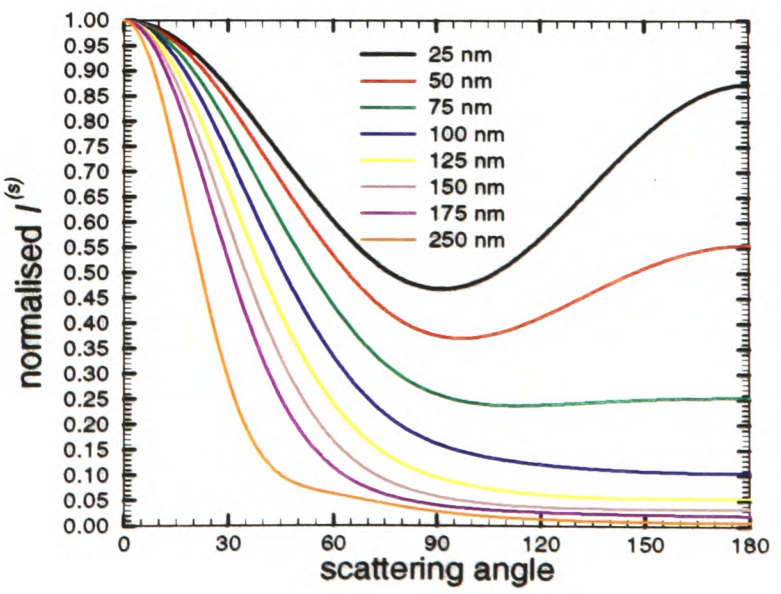


Figure 8.14: Scattering diagram of SiO_2 particles. Unpolarised illumination. Normalisation is applied.

plotted on a different scale (linear or log) for each particle radius in Figs. 8.4 to 8.13. However, for a better comparison the absolute unpolarised scattered intensity is now plotted on the same scale (log) for each particle radius in Fig. 8.15, together with the corresponding polarisation ratio P curves (Eq. 8.38).

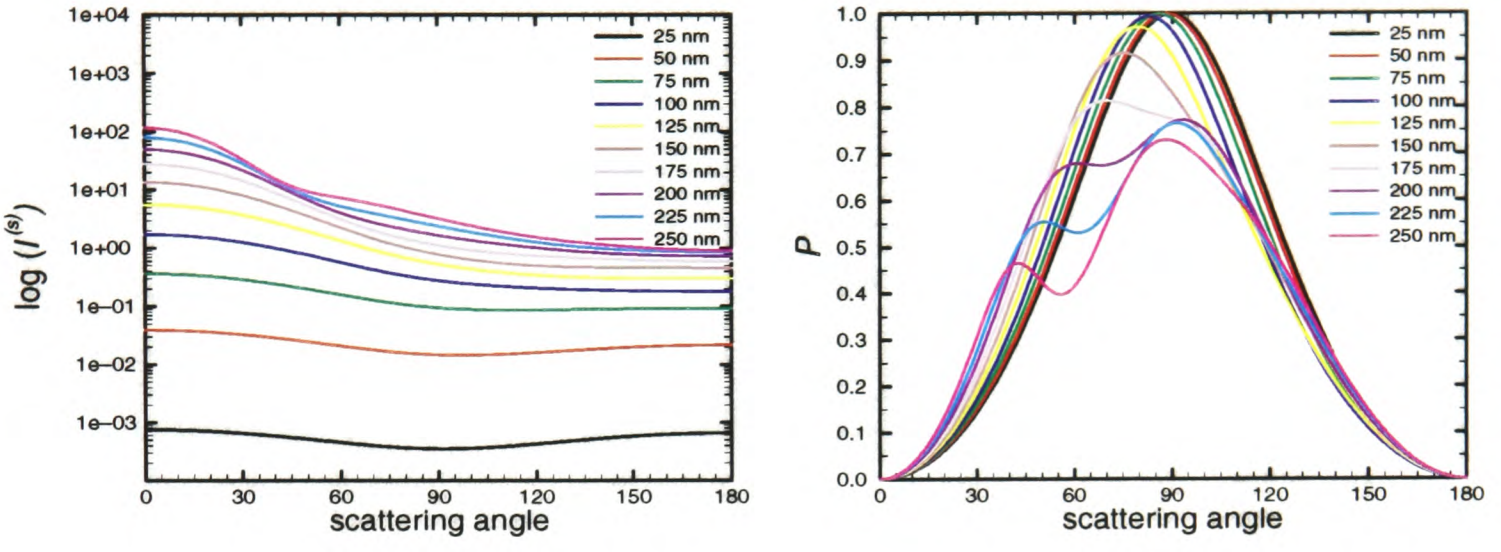


Figure 8.15: Scattering diagram of SiO_2 particles. Unpolarised illumination. No normalisation. The polarisation ratio P is also shown.

If we confine our attention now to the scattering angles of special interest for reflection mode SIRM, i.e. from 165.9° to 180°, Fig. 8.15 shows that no significant variations occur in either the scattered intensity or the polarisation ratio P as the angle changes. For the large particles, the absolute intensity of backward scattering is approximately two orders of magnitude lower than the corresponding forward scattering values, while for the small particles, the analogous absolute intensities are similar. Furthermore, for the reflection SIRM ($\sim 180°$ scattering) in order to cover the full particle size range, a dynamic intensity range of $\sim 10^3$ is required, compared with the transmission SIRM ($\sim 0°$ scattering), where a dynamic intensity range of $\sim 10^5$ is required. However, with the transmission SIRM, the contrast is also determined by the light blocking (or absorption) by the particle, but this does not affect the contrast for the reflection SIRM.

Based on the above numerical results, the angle dependent scattering coefficient Ω_{sca} for $\alpha = 14.1°$, corresponding to integration from 165.9° to 180° and determined as a function of particle radius is shown in Fig. 8.16. This curve is in effect a measure of the particle contrast expected for the reflection SIRM. The contrast progressively increases

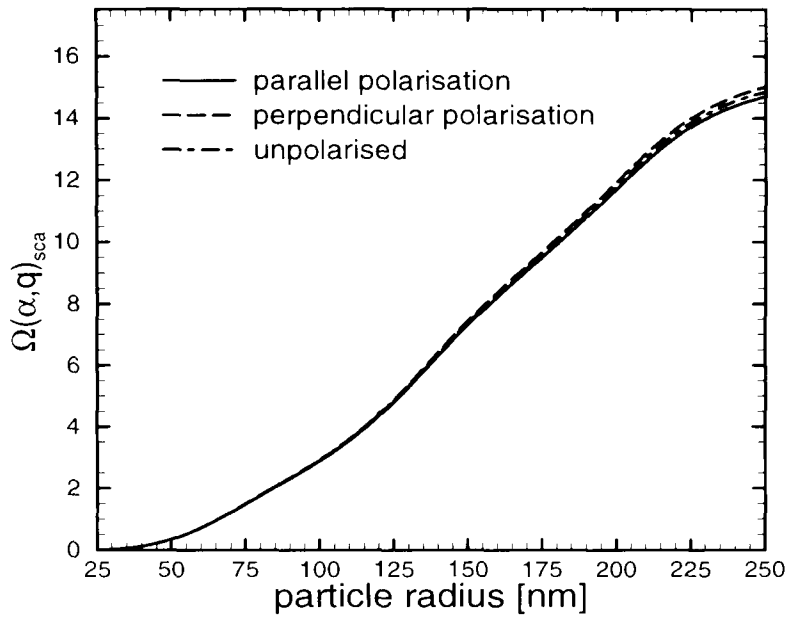


Figure 8.16: Angle dependent scattering coefficient $\Omega_{sca}(\alpha, q)$ for $\alpha = 14.1°$. The continuous curve corresponds to the parallel, the dashed to the perpendicular and the dot-dashed to the unpolarised illumination.

as the particle size increases, but is insignificantly dependent on the state of polarisation of the incident illumination.

This Figure represents the ultimate goal of this Chapter as it shows the signal strength that can be expected when SiO_2 particles are measured with the SIRM operating in the reflection mode. It shows that from a particle radius range of $\sim$ 85 nm to 250 nm the signal strength changes only 7.5 times. It can be concluded from Fig. 8.16 that the detectable signal strength is not expected to show rapid changes at larger particle radii, in fact, the correspondence between signal strength and particle radius is approximately linear in the range of 85 to 200μm and the curve also exhibits asymptotic behaviour for particle radii greater than $\sim$ 200μm. The signal strength at smaller particle radii, however, remains of particular interest because the detectable minimum particle radius can be predicted from those data. For this reason, the data shown in Fig. 8.16 are re-plotted on a log – log scale in Fig. 8.17.

This figure shows that the whole scattering range can be approximated by different functions depending on the radius of the particle. In particular, in the radius range of 25 to 40μm, the scattering follows the Rayleigh scattering (q^6 dependence). Beyond this radius, the scattering can be approximated by q^3, q^2 and $q^{1.2}$ functions for radius ranges

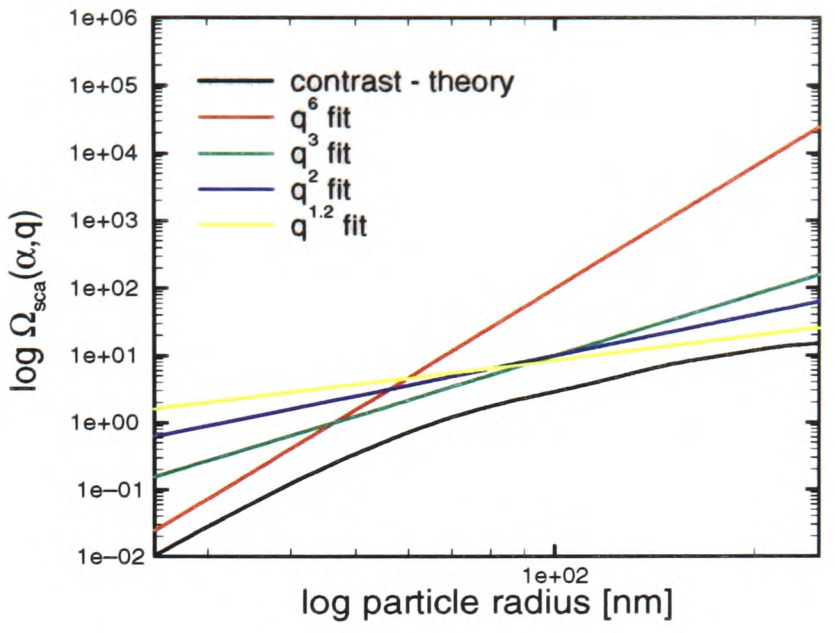


Figure 8.17: Log of the angle dependent scattering coefficient $\Omega_{sca}(\alpha, q)$ for $\alpha = 14.1°$. Coloured curves correspond to different fittings.

of 40 to 100μm, 100 to 200μm and 200 to 250μm respectively. It is emphasized that the slopes of the fitted curves are relevant, but not the vertical positions which are arbitrary. We shall return to this point in Chapter 10 where a comparison is made between these theoretical results and experimental data.

An alternative representation of the scattered intensities is obtained by plotting them onto polar diagrams. Thus, Fig. 8.18(a) shows absolute scattered intensity for unpolarised light plotted on a log scale using the data of Fig. 8.15, and Fig. 8.18(b) shows normalised scattered intensity for unpolarised light plotted on a linear scale using the data of Fig. 8.14.

Polar diagrams have similarly been plotted in Figs. 8.19 to 8.29 showing scattered intensity for polarised light using the data of Figs. 8.3 to 8.13 respectively. Each polar diagram corresponds to a particular particle radius and is plotted with the same linear scale that was used above in the corresponding Figure.

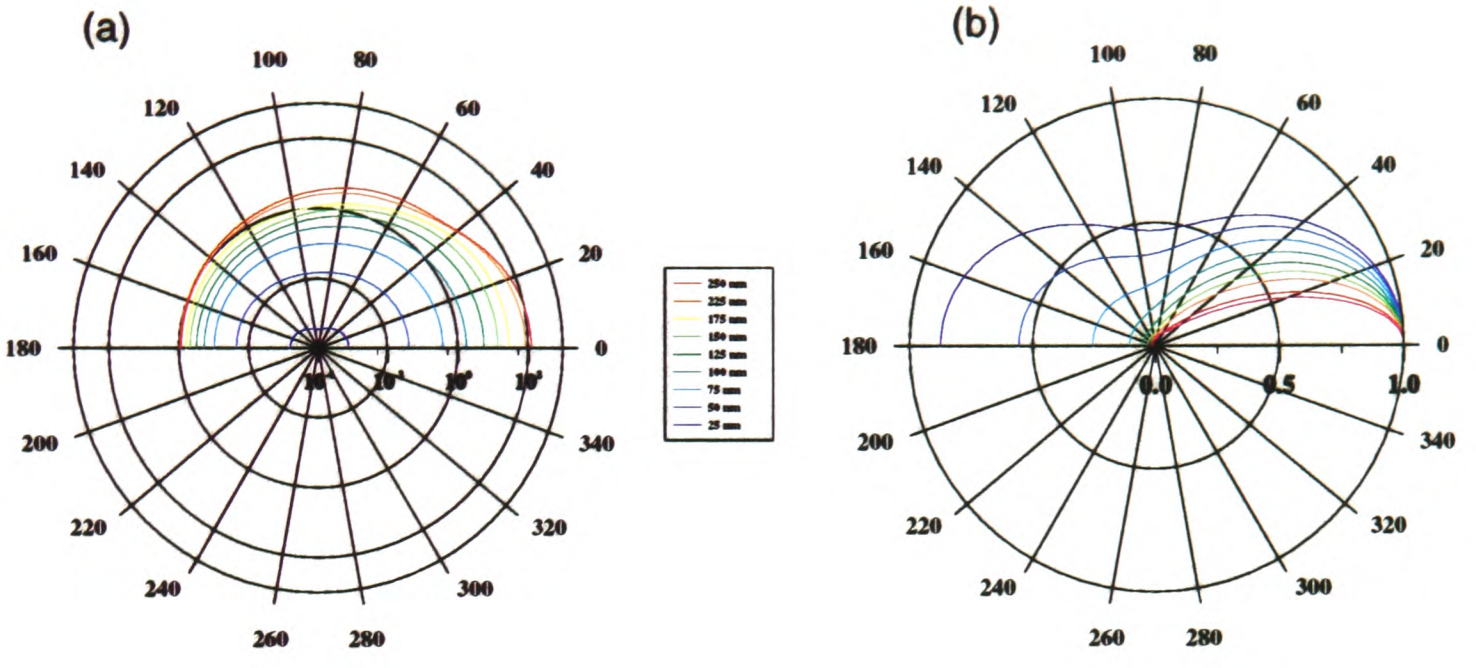


Figure 8.18: Log of scattered intensity (a), and normalised scattered intensity (b) for various particle radii. Unpolarised illumination.

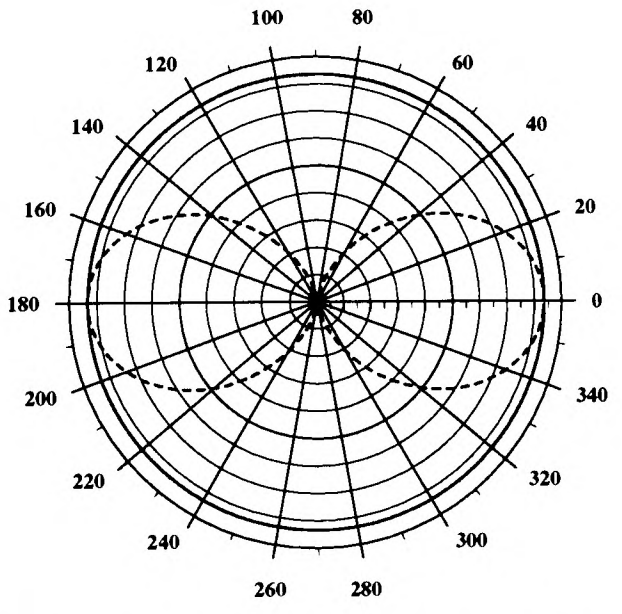


Figure 8.19: Scattering diagram of a SiO_2 particle, $a = 2.5\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

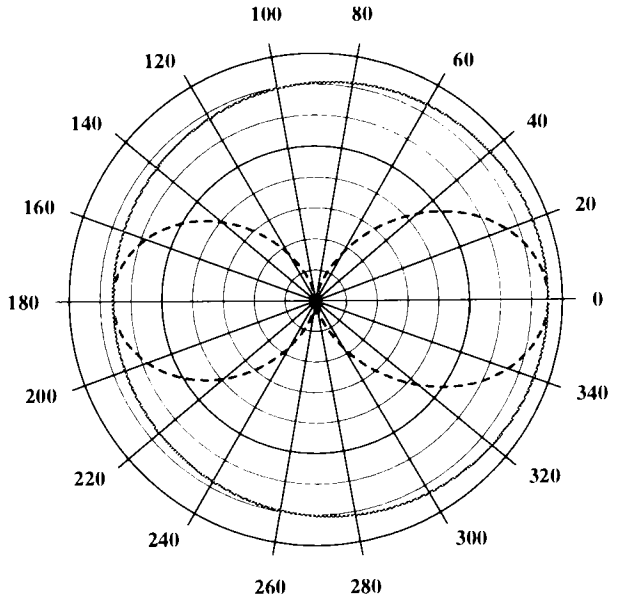


Figure 8.20: Scattering diagram of a SiO_2 particle, $a = 25\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

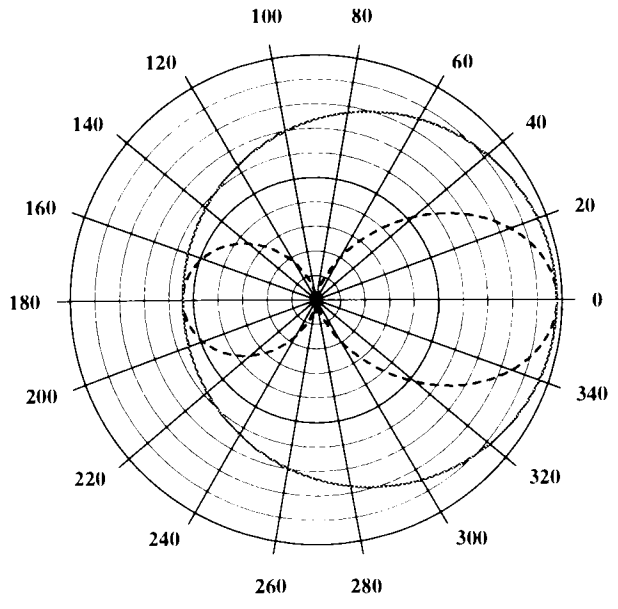


Figure 8.21: Scattering diagram of a SiO_2 particle, $a = 50\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

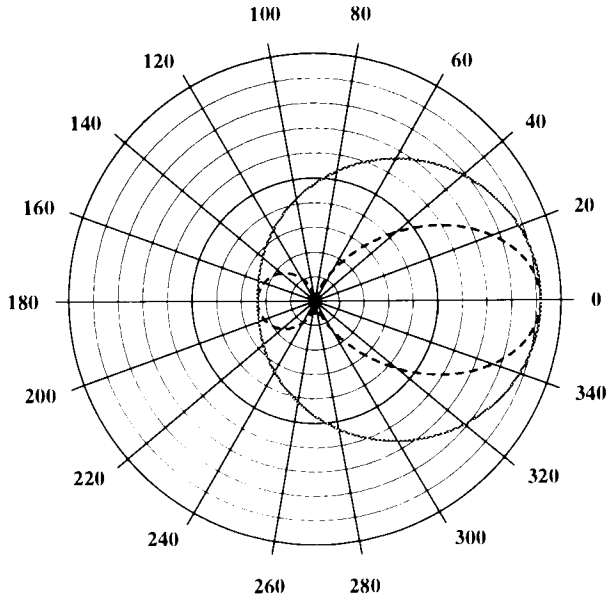


Figure 8.22: Scattering diagram of a SiO_2 particle, $a = 75\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

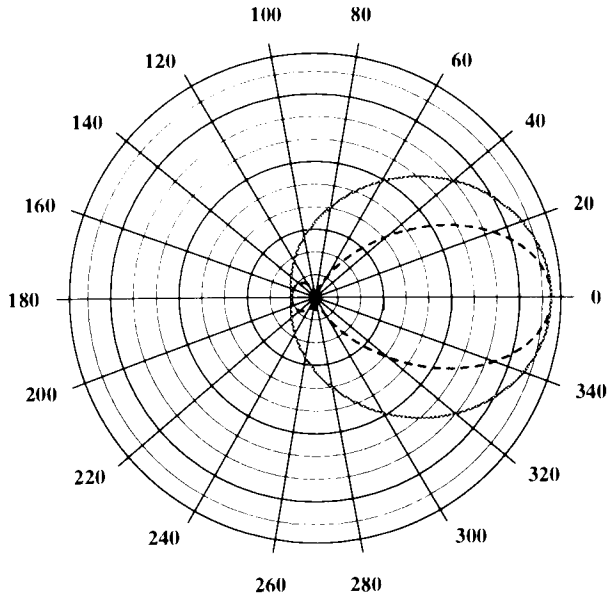


Figure 8.23: Scattering diagram of a SiO_2 particle, $a = 100\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

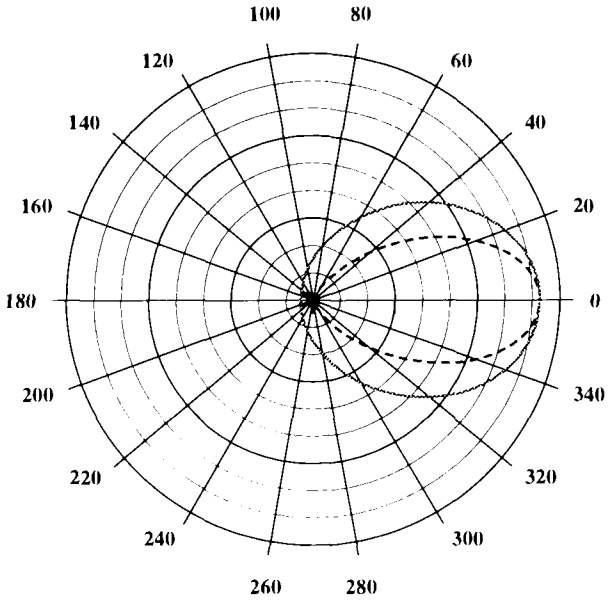


Figure 8.24: Scattering diagram of a SiO_2 particle, $a = 125\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

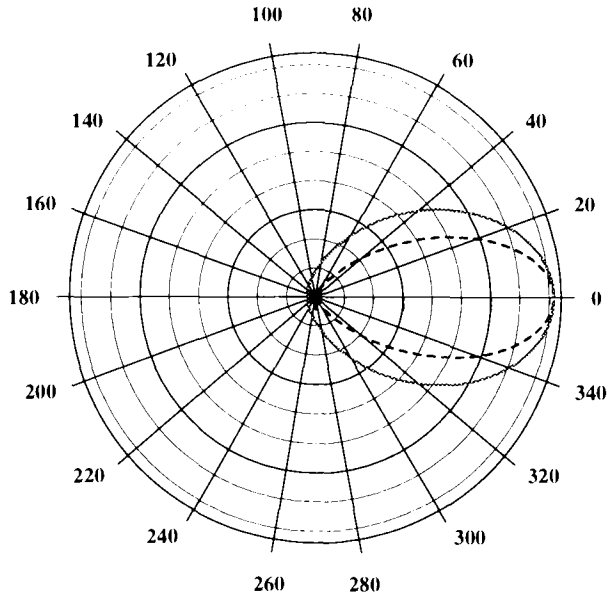


Figure 8.25: Scattering diagram of a SiO_2 particle, $a = 150\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

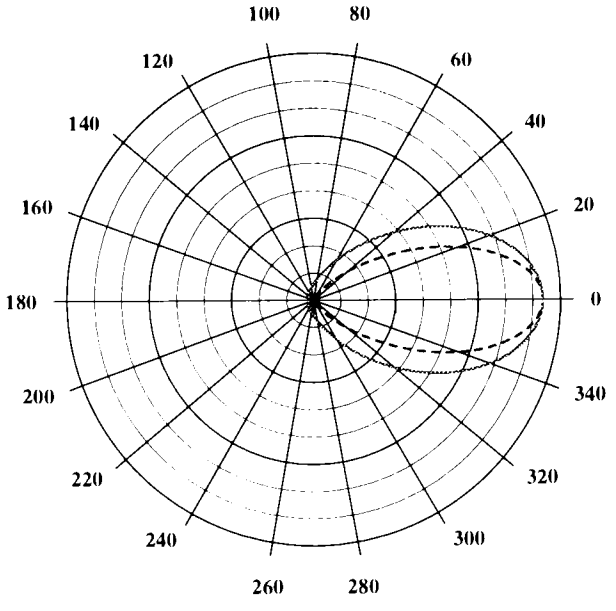


Figure 8.26: Scattering diagram of a SiO_2 particle, $a = 175\,\text{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

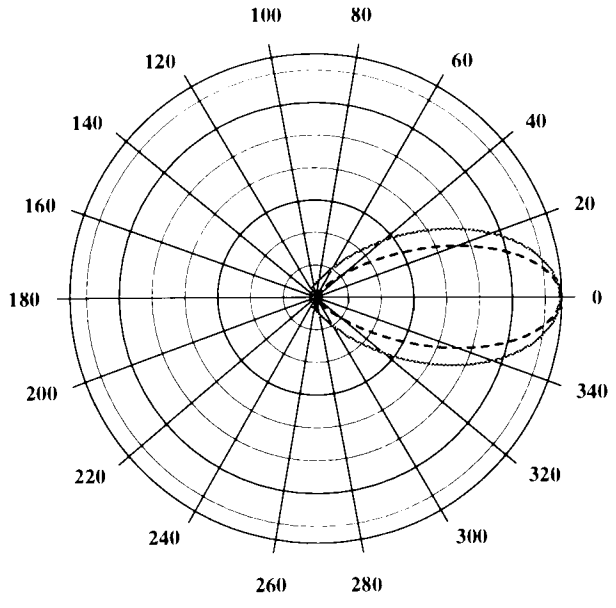


Figure 8.27: Scattering diagram of a SiO_2 particle, $a = 200\,\text{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

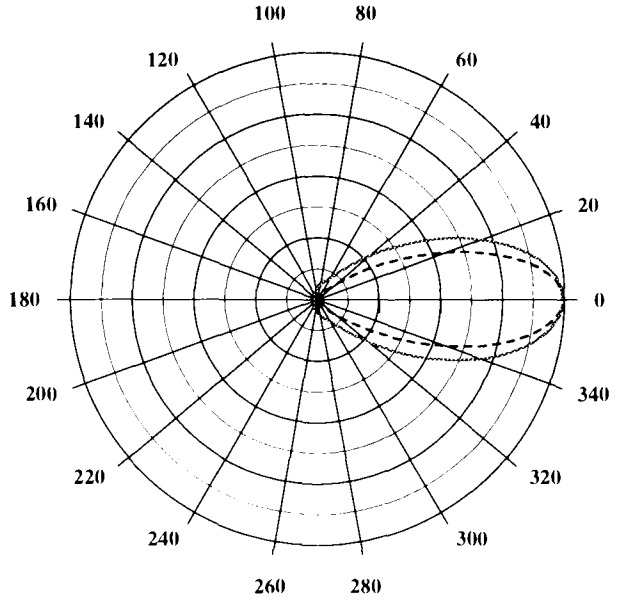


Figure 8.28: Scattering diagram of a SiO_2 particle, $a = 225\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

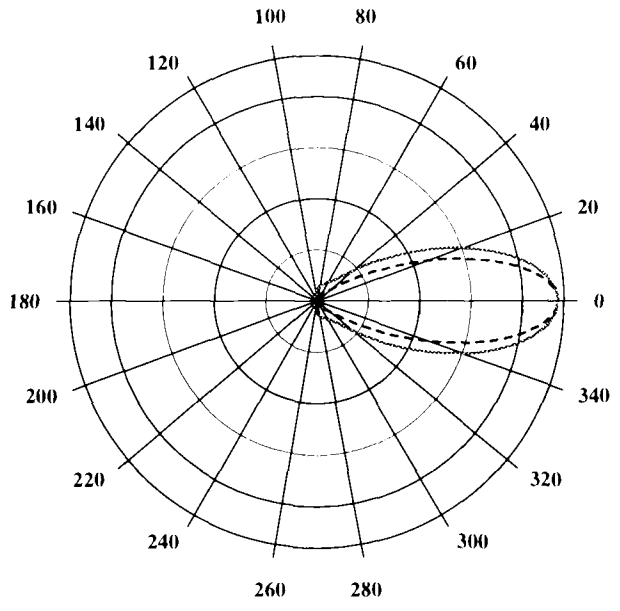


Figure 8.29: Scattering diagram of a SiO_2 particle, $a = 250\,\mathrm{nm}$. The continuous curve corresponds to perpendicular polarisation, the dashed curve corresponds to parallel polarisation.

CHAPTER 9

Solution of the diffraction integrals in the presence of a scatterer in the focal region - Backward propagation

"The more accurate the calculations become, the more the concepts tend to vanish into air."

R.S. Mulliken

9.1 Introduction

The aim of the present chapter is to give a formal solution to the problem when the scatterer (sphere) is illuminated by plane waves with different incident directions and the scattered light traverses the interface between two dielectric materials in such a way that it propagates *towards* the lens. This situation is of practical interest since it occurs when, for example, light scattered by oxide particles in silicon is collected by the probe-forming lens. In this chapter we only consider the rigorous but formal solution of the backward scattering, i.e. the case which corresponds to the reflection confocal SIRM. Our formulation is exact and to the best knowledge of the author there were no attempts made to solve this problem previously.

Before proceeding with the formal solution it is worth to discuss how the solution is obtained. Consider a spherical particle embedded in a dielectric material. The refractive index of the outer material is prescribed to be real, whilst that of the particle can possess real and imaginary parts. These conditions are identical to those imposed in Chapter 8. The particle is illuminated by light focussed close to the particle. It is not assumed that the particle is on the optical axis neither that it is in the focal plane. The mathematical description of the illumination of the particle is given by Eqs. (5.13) and (5.14) of Chapter 5.

9. Backward propagation

9.1. Introduction

As shown by Eq. (5.49), Eq. (5.13) can be regarded as a representation of the sum of plane waves propagating within a solid angle and with different direction cosines. This fact is the basis of the present solution of the problem, where we apply results obtained in Chapter 8, and the electric and the magnetic fields can be described as the sum of individual plane waves. Unfortunately, the kernel in Eq. (5.49) which describes the strength and initial phase of each plane wave cannot be directly applied in the (x, y, z) co-ordinate system to our problem, because the initial condition of Eq. (8.3) no longer applies. Instead, we use local co-ordinate systems which are set for each plane wave in such a way that the strength vector $\mathbf{c}$ determines the x direction. The origin of each local co-ordinate system coincides with the particle P. The wave vector $\mathbf{s}$ is perpendicular to $\mathbf{c}$ and these two vectors together with their cross-product define the $(\bar{x}, \bar{y}, \bar{z})$ co-ordinate system. These co-ordinate systems are mutually independent of the (x, y, z) system. Since the scattering tensor $\underline{\underline{S}}$ and $\mathbf{c}$ are both given in the spherical polar co-ordinate system (r, ϑ, φ), they are transformed to the corresponding Cartesian co-ordinate systems determined by $\mathbf{c}$ and $\mathbf{k}$. The next step is to shift this latter co-ordinate system to the origin of the (x, y, z) system and to perform a co-ordinate system rotation to transform $(\bar{x}, \bar{y}, \bar{z})$ to (x, y, z). The resulting expression describes a directed spherical wave whose direction depends on the incident plane wave illuminating the particle. In order to be able to handle the spherical wave so obtained we decompose it into plane wave components by a Fourier-transform and the resulting expression is a function which consists of the sum of plane waves whose direction depends on the direction and phase of the incident plane wave. Different incident plane wave directions are considered so that the field is expressed as a superposition of individual plane waves. For this we integrate the expression for the plane waves in such a way that all incident directions are considered (over the ψ solid angle corresponding to the angle of total internal reflection). This field is then transmitted linearly through the interface and so the same principle can be applied as in Chapter 5. This means that the outside field at the vicinity of the interface is described by Wolf's integral formulae and, by means of an appropriate co-ordinate transformation, the outside field is equal to the inside field expressed at the vicinity of the interface. The electric and magnetic fields can

then readily be determined before the lens.

The imaging by the lens is considered with the application of Luneburg's off axis diffraction integral. In this interpretation off-axis sources situated in the object space of an imaging system can be treated and thus its application is convenient for our case. Application of this integral results in the electric and magnetic field just in front of the pin-hole. The intensity measured by the pin-hole can either be considered by taking the size of the pin-hole to be infinitely small, or alternatively to integrate over the finite area of it.

It must be emphasized, that the mathematics of the above problem is far from trivial. This method can be computed numerically but the precision of the result depends mainly on the precision of the individual computation steps. It is likely that an analytical expression can be found to approximate the scattering tensor for the given solid angle since within this the value of the tensor does not change significantly (see Chapter 8).

9.2 Formulation and solution

The solution is considered with the help of the geometry shown in Fig. 9.1

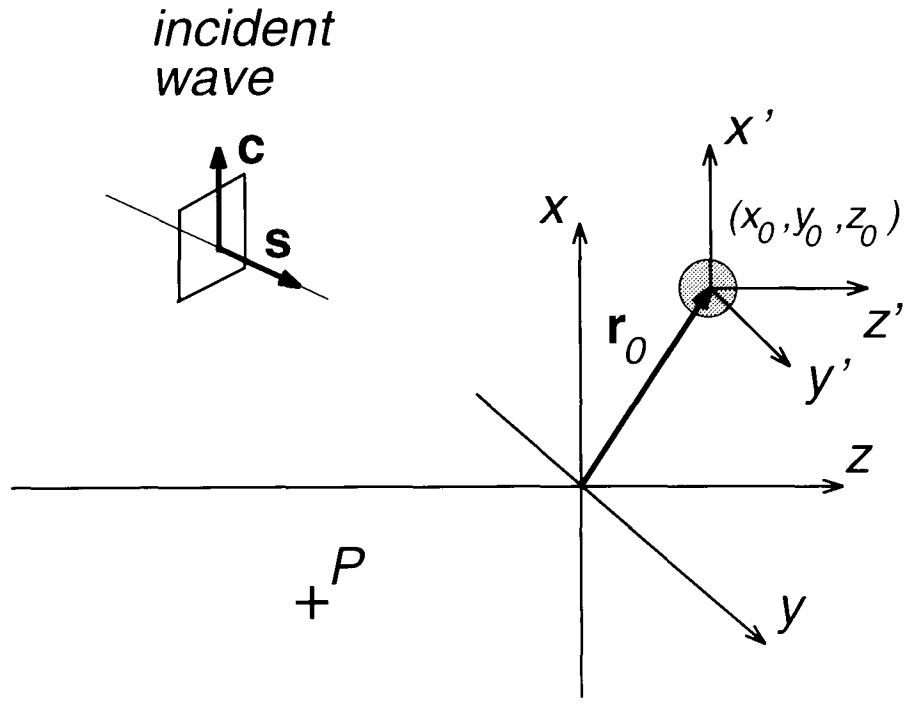


Figure 9.1: Diagram showing the co-ordinate systems used.

The incident plane wave is characterised by a propagation vector $\mathbf{s}$ and a strength vector $\mathbf{c}$. The scatterer is placed at the point (x_0, y_0, z_0) and an $\mathbf{r}_0$ vector points from the origin of a steady Cartesian co-ordinate system (x, y, z) to the centre of the scatterer. A second Cartesian co-ordinate system $(\bar{x}, \bar{y}, \bar{z})$ is defined in such a way that its origin coincides with the origin of (x, y, z) and its base vectors are:

$$(\bar{\boldsymbol{i}}, \bar{\boldsymbol{j}}, \bar{\boldsymbol{k}}) = \left(\frac{\mathbf{c}}{|\mathbf{c}|}, \frac{\mathbf{s} \times \mathbf{c}}{|\mathbf{s} \times \mathbf{c}|}, \frac{\mathbf{s}}{|\mathbf{s}|} \right). \tag{9.1}$$

A third Cartesian co-ordinate system (x', y', z') is attached to the scatterer in such a way

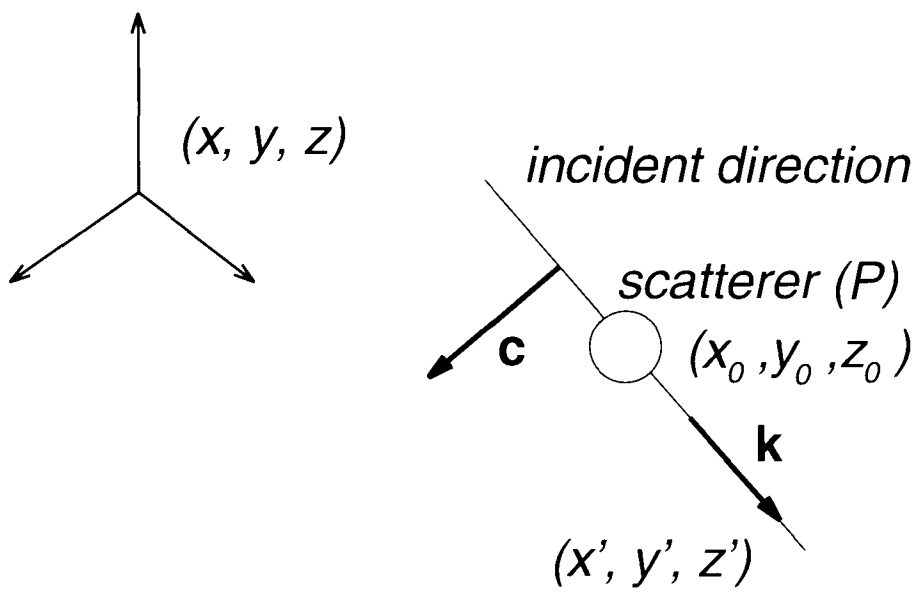


Figure 9.2: Diagram showing the local co-ordinate systems.

that its origin coincides with centre of the scatterer and its base vectors are (see Fig. 9.2):

$$(\boldsymbol{i}', \boldsymbol{j}', \boldsymbol{k}') = (\bar{\boldsymbol{i}} + \mathbf{r}_0, \bar{\boldsymbol{j}} + \mathbf{r}_0, \bar{\boldsymbol{k}} + \mathbf{r}_0). \tag{9.2}$$

The scattered electric field at point P, with polar co-ordinates in the (x', y', z') co-ordinate system of $(r'_p, \vartheta'_p, \varphi'_p)$, is given from Eq. 8.31:

$$\begin{pmatrix} E_{r'} \\ E_{\vartheta'} \\ E_{\varphi'} \end{pmatrix} = \frac{\mathrm{e}^{\mathrm{i}k^{(I)}r'_p}}{-\mathrm{i}k^{(I)}r'_p} \begin{pmatrix} 0 & 0 & 0 \\ 0 & \cos\varphi'_p S_2 & 0 \\ 0 & 0 & -\sin\varphi'_p S_1 \end{pmatrix} \cdot \begin{pmatrix} E_x^{(i)} \\ 0 \\ 0 \end{pmatrix} = \underline{\underline{\mathrm{S}}}_{(r', \vartheta', \varphi')} \cdot \mathbf{E}^{(i)} \tag{9.3}$$

and from Eqs. 5.28 and 5.30 the strength vector:

$$\mathbf{c}(\vartheta, \varphi) = A(\vartheta_1) \begin{pmatrix} c_x \\ c_y \\ c_z \end{pmatrix} = A(\vartheta_1) \begin{pmatrix} \tau_p \cos\vartheta \cos^2\varphi + \tau_s \sin^2\varphi \\ \tau_p \cos\vartheta \sin\varphi \cos\varphi - \tau_s \sin\varphi \cos\varphi \\ -\tau_p \sin\vartheta \cos\varphi \end{pmatrix} \tag{9.4}$$

where ϑ and φ indicate the direction of the incident plane wave and

$$\tau_s = \frac{2\sin\vartheta\cos\vartheta_1}{\sin(\vartheta_1+\vartheta)} \qquad \tau_p = \frac{2\sin\vartheta\cos\vartheta_1}{\sin(\vartheta_1+\vartheta)\cos(\vartheta_1-\vartheta)}, \tag{9.5}$$

are the Fresnel coefficients. We note that in the present derivation the notation of azimuthal and rotational angles are interchanged with respect to those of Chapters 5, 6 and 7 which was done to keep the original notations of Wolf in the above chapters. It follows from the definition of (x', y', z') and Eqs. (9.1) and (9.2) that the scattered electric field is given from Eq. (9.3):

$$\begin{pmatrix} E_{r'} \\ E_{\vartheta'} \\ E_{\varphi'} \end{pmatrix} = \underline{\underline{S}}_{(r',\vartheta',\varphi')} \begin{pmatrix} |\mathbf{c}(\vartheta,\varphi)|\,\exp[ik\Psi(\vartheta,d)] \\ 0 \\ 0 \end{pmatrix}. \tag{9.6}$$

The propagation vector $\mathbf{s}$ is given in (x, y, z) system by:

$$\mathbf{s} = \sin\vartheta\cos\varphi\,\boldsymbol{i} + \sin\vartheta\sin\varphi\,\boldsymbol{j} + \cos\vartheta\,\boldsymbol{k}. \tag{9.7}$$

The co-ordinate transformation that takes the $(r', \vartheta, \varphi')$ system to (x', y', z') is:

$$\underline{\underline{R}} = \begin{pmatrix} \sin\vartheta_s\cos\varphi_s & \cos\vartheta_s\cos\varphi_s & -\sin\varphi_s \\ \sin\vartheta_s\sin\varphi_s & \cos\vartheta_s\sin\varphi_s & \cos\varphi_s \\ \cos\vartheta_s & -\sin\vartheta_s & 0 \end{pmatrix} \tag{9.8}$$

with which we write

$$\begin{pmatrix} E_{x'} \\ E_{y'} \\ E_{z'} \end{pmatrix} = \underline{\underline{R}} \cdot \begin{pmatrix} E_{r'} \\ E_{\vartheta'} \\ E_{\varphi'} \end{pmatrix} \tag{9.9}$$

We now translate the (x', y', z') system into $(\bar{x}, \bar{y}, \bar{z})$ system so as point P has the coordinates

$$\left.\begin{aligned} x'_p &= \bar{x}_p + x_0 \\ y'_p &= \bar{y}_p + y_0 \\ z'_p &= \bar{z}_p + z_0 \end{aligned}\right\}. \tag{9.10}$$

The electric field is now given in $(\bar{x}, \bar{y}, \bar{z})$ system. We express the electric field in (x, y, z) system by a rotation transformation. Let us define α, β and γ as the angles which the base vectors of $(\bar{x}, \bar{y}, \bar{z})$ system make with the base vectors of (x, y, z):

$$\left.\begin{aligned} \alpha &= \arccos[\mathbf{c}\cdot\boldsymbol{k}] \\ \beta &= \arccos[(\mathbf{s}\times\mathbf{c})\cdot\boldsymbol{j}] \\ \gamma &= \arccos[\mathbf{s}\cdot\boldsymbol{i}] \end{aligned}\right\}. \tag{9.11}$$

The effect of the rotation on the electric vector can be written as:

$$\begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} = \underline{\underline{\mathrm{A}}} \cdot \begin{pmatrix} E_{\bar{x}} \\ E_{\bar{y}} \\ E_{\bar{z}} \end{pmatrix} \tag{9.12}$$

where[3]

$$\underline{\underline{\mathrm{A}}} = \begin{pmatrix} \cos\gamma\cos\beta\cos\alpha - \sin\gamma\sin\alpha & \cos\gamma\cos\beta\sin\alpha + \sin\gamma\cos\alpha & -\cos\gamma\sin\beta \\ \sin\gamma\cos\beta\cos\alpha - \cos\gamma\sin\alpha & -\sin\gamma\cos\beta\sin\alpha + \cos\gamma\cos\alpha & \sin\gamma\sin\beta \\ \sin\beta\cos\alpha & \sin\beta\sin\alpha & \cos\beta \end{pmatrix}$$

After this transformation we obtain a 'directed' spherical wave ($E^{(s)}$):

$$E^{(s)}_{(x,y,z)}(\vartheta, \varphi; x_0, y_0, z_0; x_p, y_p, z_p)$$

which depends on the direction (ϑ, φ) and phase ($\exp[\mathrm{i}k\Psi(\vartheta, d)]$) of the incident plane wave, the position of the scatterer (x_0, y_0, z_0) in the (x, y, z) system and on the position where the field is detected (x_p, y_p, z_p).

This spherical wave, however, cannot be treated readily when the light traverses the interface towards the lens. It is known that any kind of waveform can be constructed as the superposition of plane waves. This principle, in its mathematical form, is stated by the angular spectrum representation. Thus, we Fourier analyse the 'directed' spherical wave to obtain the plane wave components. The Fourier transform can be defined between functions $A(f_x, f_y)$ and $U(x, y, z)$ by[32]:

$$A(f_x, f_y, z) = \int\!\!\int_{-\infty}^{\infty} U(x, y, z) \exp[2\pi\mathrm{i}(f_x x + f_y y)]\mathrm{d}x\mathrm{d}y \tag{9.13}$$

where, as a physical interpretation, the spherical wave is expressed as a sum of plane waves by Eq. 9.13. These plane waves propagate with directional cosines λf_x and λf_y. Let us now indicate the two-dimensional Fourier transformation by $\Im$. From now on we shall omit the subscripts for point P (p), and the indication of the (x, y, z) system. The plane waves from which the 'directed' spherical wave can be constructed are given by:

$$\mathbf{E}^{(a)}(\vartheta, \varphi; x_0, y_0, z_0; \frac{f_x}{\lambda}, \frac{f_y}{\lambda}, z) = \Im\left\{\mathbf{E}^{(s)}(\vartheta, \varphi; x_0, y_0, z_0; x, y, z)\right\} \tag{9.14}$$

where the '(a)' superscript indicates that the electric vector is expressed in terms of plane waves.

Having obtained the plane wave components of the scattered field, we can express the electric field inside the *second* material and at the vicinity of the interface by:

$$\mathbf{E}^{(a)}\left(\frac{f_x}{\lambda},\frac{f_y}{\lambda},-(d-\delta);\mathbf{r}_0\right)=C\int_0^{\psi}\int_0^{2\pi}\mathbf{E}^{(a)}[\vartheta,\varphi;\mathbf{r}_0;\frac{f_x}{\lambda},\frac{f_y}{\lambda},-(d-\delta)]\mathrm{d}\vartheta\mathrm{d}\varphi \tag{9.15}$$

where C is a constant and ψ is the angle of total reflection for the two given materials. The electric field inside the *first* material but at the vicinity of the interface is expressed by Wolf's integral:

$$\begin{aligned}&\mathbf{E}^{(r)}\left(\frac{f_x}{\lambda},\frac{f_y}{\lambda},-(d+\delta);\mathbf{r}_0\right)=\\&C'\int\int_{\Omega_1}\mathbf{W}(\mathbf{s}_1)\exp\left\{\mathrm{i}k_1\left[s_{1x}\frac{f_x}{\lambda}+s_{1y}\frac{f_y}{\lambda}-s_{1z}(d+\delta)\right]\right\}\mathrm{d}s_{1x}\mathrm{d}s_{1y}\end{aligned} \tag{9.16}$$

where Ω_1 corresponds to the 90° solid semi-angle, $\mathbf{s}_1=(s_{1x},s_{1y},s_{1z})$ is the wave vector in the first material corresponding to $\mathbf{s}$ in the second material and k_1 is the wave number of the first material. From Eqs. 9.15 and 9.16 the field can readily be expressed in the same way as in Chapter 5. The electric field so obtained in air, at the interface is denoted by

$$\mathbf{E}^{(r)}(x,y,-d,x_0,y_0,z_0).$$

We consider the effect of the light traversing the lens by introducing Luneburg's[60] off-axis diffraction integral for our case:

$$\mathbf{E}^{(p)}=-\frac{\mathrm{i}k_1}{2\pi}\exp(\mathrm{i}k_0C_0)\int\int\sqrt{|K_0|}\mathbf{U}_0\exp\{\mathrm{i}k_1[a(x-\xi)+b(y-\eta)-c(d+\zeta)]\}\mathrm{d}\sigma \tag{9.17}$$

where a, b and c are directional cosines of the normal of a phase front at $(x,y,-d)$, (ξ,η,ζ) is the point where the electric field is to be determined (at the pin-hole), and C_0 is a reference wavefront. Other notations are the same as in Chapter 4.

When the signal of the pin-hole is of interest, the field must be determined at the plane of the pin-hole and integrated within the area of the pin-hole.

It is emphasised that the analytical solution or the equations that would be suitable for numerical computations require additional work.

CHAPTER 10

Experimental results and interpretation

"Nature will tell you a direct lie if she can."
C. Darwin

10.1 Introduction

The aim of this chapter is to discuss the practical aspects of the Oxford SIRM including the presentation of images taken by the SIRM operating in different imaging modes, and to compare experimental results to those of theory. First a general discussion is made on the system design from the mechanical, electronic and optical points of view. Subsequently a detailed discussion is presented on the most important imaging mode, namely, the reflection confocal mode. The imaging capabilities of the microscope are demonstrated by a number of images taken from a wide range of specimens. Focal series and cross sectional images are presented which show the effect of the spherical aberration on the images. Imaging modes, different from reflection confocal, are discussed and images are presented corresponding to each imaging mode. Contrast data measured in the reflection confocal mode are presented and compared with the theoretical values obtained in Chapter 8.

10.2 System design

There are several important points to be considered when a SIRM is designed. In particular an early decision should be made on the required magnification, resolution, sensitivity and stability of the system.

10.2.1 MECHANICAL DESIGN

The mechanical stability of the SIRM is important since the required sensitivity and resolution are high. In principle, there is a possibility to build a scanning microscope where all components are firmly fixed, but this solution prevents rapid modifications of the optical system. Instead, a system needed to be designed which unified the flexibility and the rigidity. For laboratory purposes the microscope was built on a high stability honeycomb structure optical bench that provided a good mechanical vibration isolation, mainly from the floor. Also the bench is useful for damping internal vibrations originating from the system itself, e.g. from the mechanical scanner.

The previous Oxford SIRM used an optical rail system with carriers that proved to be accurate and stable and so it was adapted for the new system. In order to fulfill the stability and flexibility conditions a set of new mechanical parts was designed to replace the old post holder/post system which was difficult to align and fix with high accuracy. New mechanical components called posts were designed to match both the rail carriers directly and also a previously introduced micro bench/platform system. These components are now available with different heights (Fig. 10.1 (A)) as well as adaptor plates with different thicknesses (Fig. 10.1 (B)). The micro bench system is flexible and plates are available at different lengths (Fig. 10.2 (A)). This system is adapted from various original designs and it benefits from all of the advantages and avoids most of the disadvantages of those. Each of these plates consists of a central slot running along the length of the plate. These slots were milled with extra care and play positioning roles. At the sides of the plates several threaded holes are available to fix various optical components firmly to the micro bench plate. Optical components may either be fixed to the micro bench system or directly to the posts. The micro bench system offers several advantages over conventional mechanical assemblies, e.g. unified beam height and easy centering, and provides a rigid but extremely flexible mechanical system. Some additional mechanical components were designed and made by the author including gimbal mirror holders (Fig. 10.2 (B)), rotation stages (Fig. 10.2 (C)), and 3D translator assemblies. All

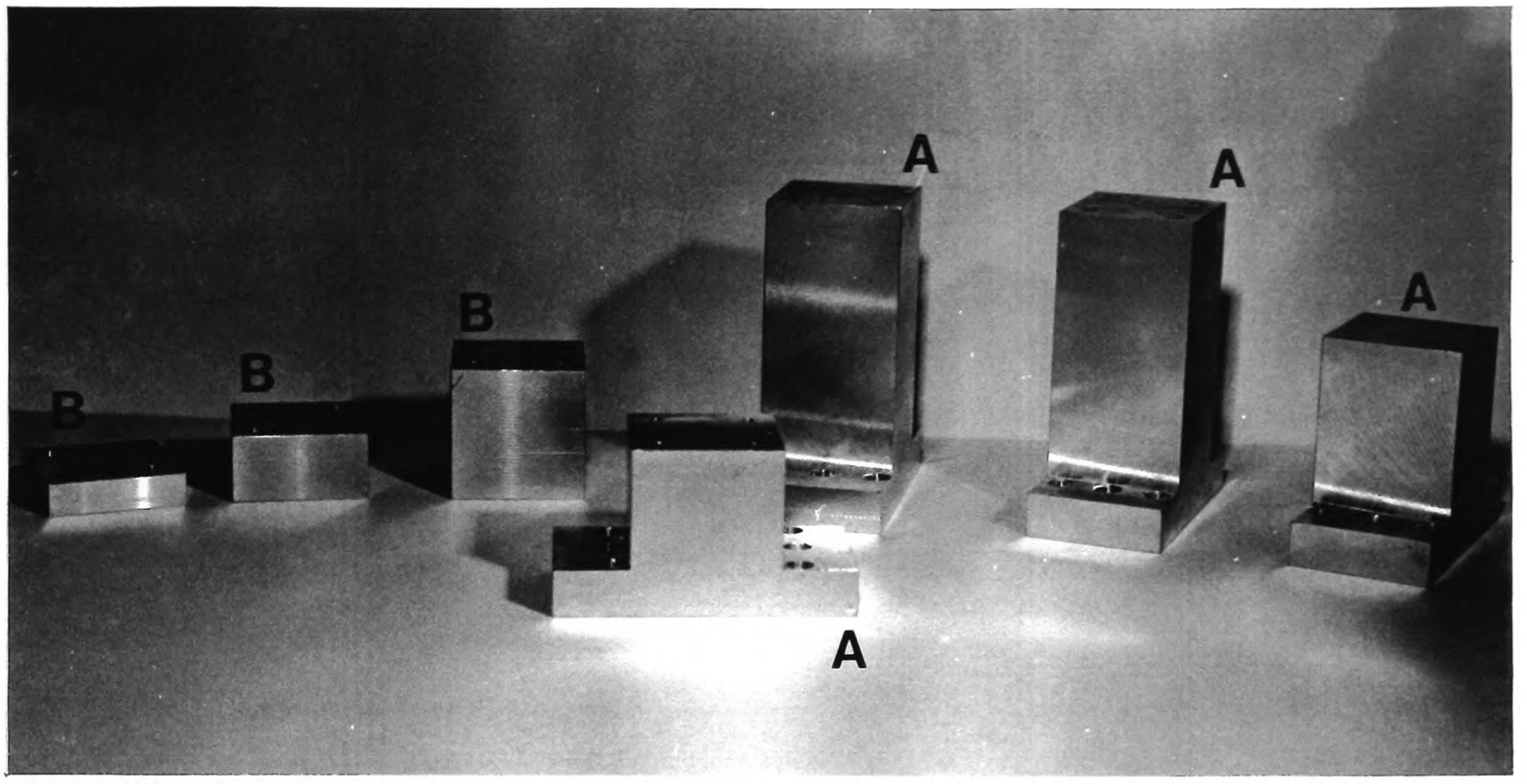


Figure 10.1: The new post system with accessories.

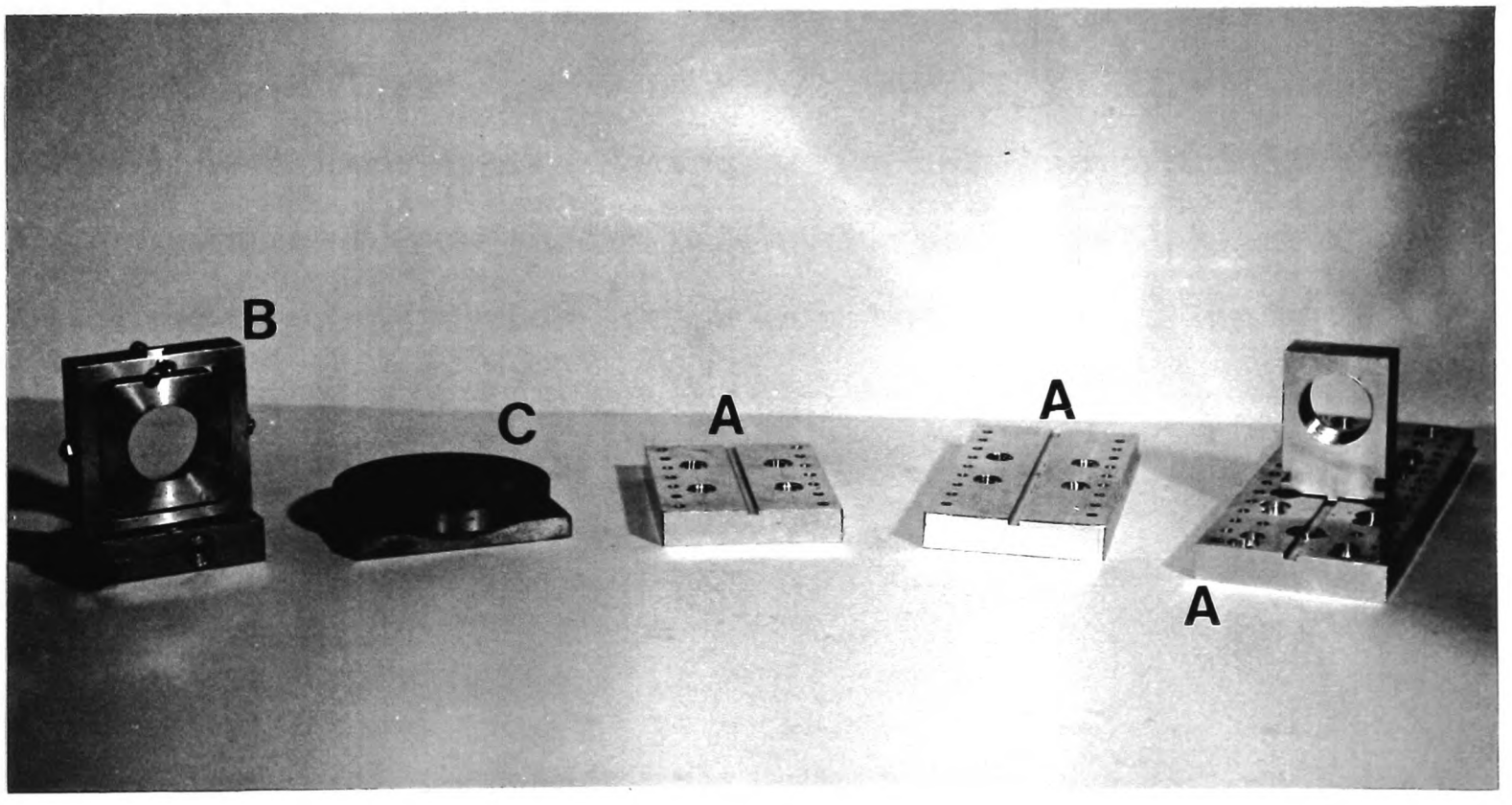


Figure 10.2: The new microbench system with some accessories.

of these components were tested in a microscope for rigidity, stability and self-oscillations.

The stable and rigid mechanics is of little value if mechanical vibration is introduced from inside the system. A vibration source can be a scanner or any moving component in the system, e.g. a mechanical chopper, or an electronic unit with its transformer on the bench. A problem can be encountered with liquid cooled lasers since as the coolant flows through the laser it causes vibration. For interferometric application the stability of the above described system could still not be high enough in which case other solutions would need to be found.

The probe-specimen relative movement is performed in the SIRM by means of a pair of orthogonal mechanical scanners providing the x and the y direction scanning. These scanners are based on interferometric precision translator stages driven by linear motors. The scanner stack is mounted on a pair of coarse, stepper motor driven translator stages providing the coarse x and y movement of the specimen over a wide range (up to 100 mm). The specimen is mounted on a removable specimen holder magnetically held to a base plate, which is attached firmly to the scanners. This removable specimen holder, which was designed and made by the author, provides the opportunity of rapid and convenient specimen change. Correct repositioning is made possible by a number of strong magnets mounted into the frame of the specimen holder. The mechanical construction of the frame and the specimen holder ensures that the holder does not get stuck and so misplaced. The measured accuracy of the repositioning is better than $\sim$ 10 μm. The scanner unit can be finely positioned in the z direction by a low voltage piezo stack-type translator in the range of 0 to 100 μm. In addition to the fine positioning, a coarse axial movement of the specimen can be performed by a stepper motor driven micrometer.

10.2.2 ELECTRONIC DESIGN

Some of the electronic components of the system were available when the design of the new SIRM started. These were the laser power supply, stepper motor drivers and the control/data acquisition computer with the software.

The system is controlled by a 486SX CPU based IBM PC computer with a 300 MByte

hard disk and 8 MByte RAM. The data acquisition is performed by a multi function interface card accommodated in the computer. The positioning of the scanners (analogue signal) is also provided by this card together with the parallel digital outputs which are used to control the stepper motor drivers. The A/D (analogue-digital) and D/A (digital-analogue) converters possess 14 bit resolution which is sufficient to position the scanners with a resolution of $\sim$ 100 nm, depending on the settings of the scanner driver. The resolution of the A/D input would in principle be sufficient to measure contrast levels down to $\sim$ 0.03% but this level is limited by noise as discussed later. The detailed description of these electronic components can be found elsewhere[57].

To improve the positioning accuracy, especially in the z direction, a piezo translator was used. Due to the special design of the microscope electronics it was more advantageous to design the driver electronics rather than to purchase one. The driver circuit, which was designed by the author, had to be constructed to fit the system without the need of modifying the existing software and hardware elements. Originally, the fine z movement was performed by a stepper motor driven micrometer and therefore the signal which controlled the movement was binary (a direction and a step bit). The interface was designed to convert these signals to an analogue signal which was suitable to drive the piezo power amplifier. The interface circuit essentially consists of a 14 bit up/down counter and a 14 bit D/A converter IC. The binary resolution was selected to match the required positioning accuracy of 30 to 50 nm. This accuracy cannot be expected in a real system due to mechanical reasons, but an accuracy of similar order is ideal for the SIRM. The interface electronics also consists of an analogue circuit that receives the signal from the position sensor built into the piezo translator. This signal is amplified by a closed loop PD (proportional and differentiated) circuit which ensures the minimum overshoot when high spatial frequency components are in the input signal.

A similar electronics design was used to provide a closed loop control for the X and Y scanners. The high voltage electronics is accommodated in a separate box and can deliver a 100 V peak-to-peak signal at 1 kHz when a pure capacitive load of 10 μF is applied (this is the typical capacitance of a low voltage piezo stack). This was achieved by a pair



Figure 10.3: The piezo driver and power amplifier for the fine z movement.

of complementary power transistors which were driven in a Darlington configuration to give a low current input signal. Although the voltage-translation characteristic of a piezo translator is not linear, with the closed loop control a linear characteristic was realised. Both the interface and the piezo amplifier are shown in Fig. 10.3.

For the DPC imaging mode a need arose for a high quality differential amplifier. This was designed and built using FET (field effect transistor) ICs to provide a high input impedance and low noise. The resulting amplifier possesses the same or less noise than a Keithley current amplifier. This electronics unit is shown in Fig. 10.4(a) together with a low noise AC amplifier (b) and the Keithley current amplifier (c). The AC amplifier was used in some of the transmission experiments whilst the Keithley amplifier was generally used to amplify the detector signal. The noise in the SIRM images partly arose from the electronics, with the most important contribution from the laser and the laser power supply ($\sim 0.1\%$). Recently, a new laser power supply was obtained and this significantly decreased the electronic noise.

The block diagram of the SIRM electronics is shown in Fig. 10.5.

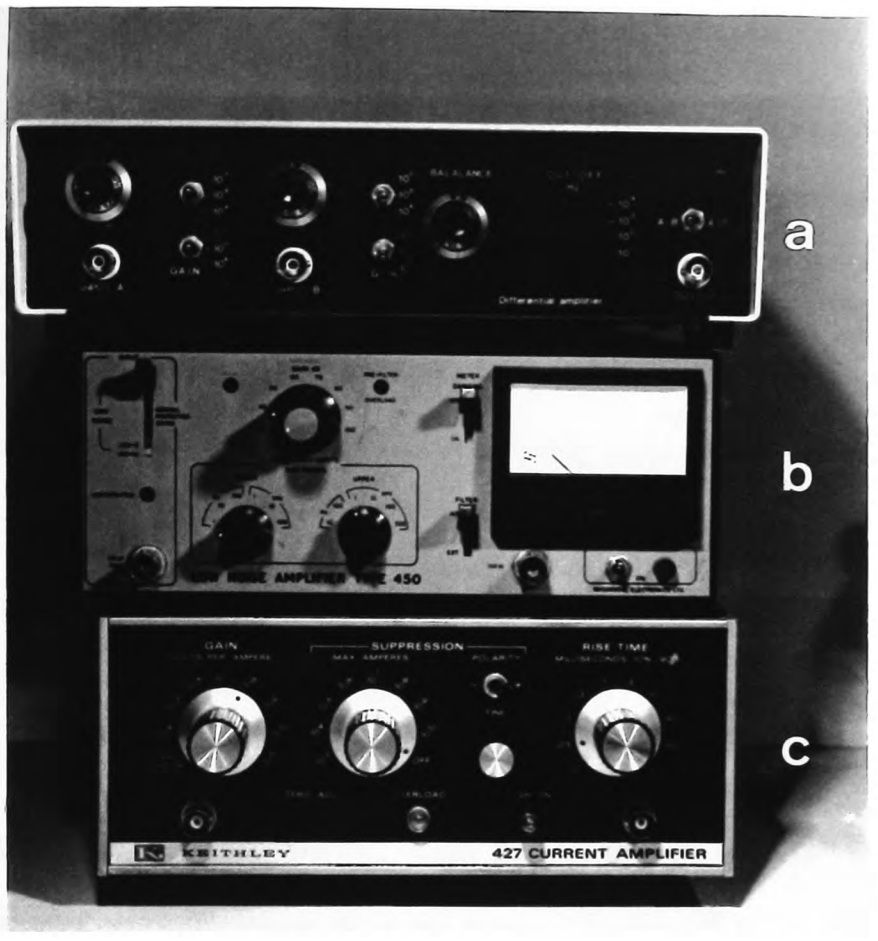


Figure 10.4: The amplifiers. Amplifier (a) is a low noise differential amplifier used for DPC imaging.

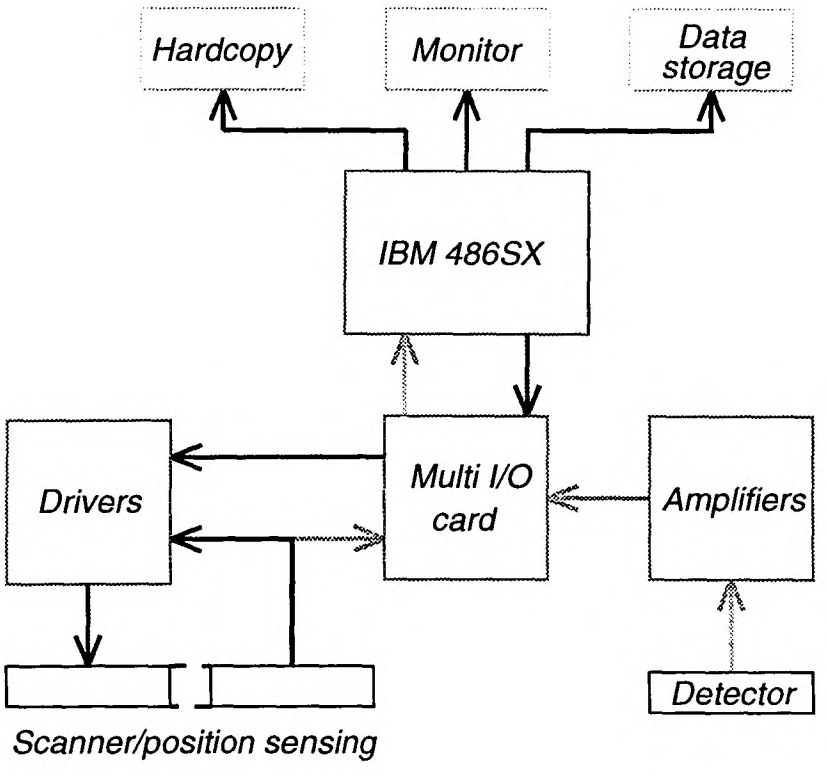


Figure 10.5: The block diagram of the SIRM electronic system.

10.2.3 OPTICAL DESIGN

The optical design usually starts with the selection of the appropriate light source. There are two major types of laser that can be used in the SIRM. One of these is a coherent gas or solid state laser that operates at the required wavelength. Use of a coherent laser is essential for an interference microscope. Gas or solid state lasers also have high output powers which may be needed for some applications, e.g. dark field SIRM. There are however many disadvantages of these lasers. From an optical point of view the coherence can cause unwanted interference effects in the system. This interference can be eliminated by using high quality anti-reflection coatings on the surfaces of the optical components. Interference may also occur inside the specimen when the light is scattered by particles or the rough back surface of the specimen. This interference cannot be eliminated easily. Speckle formation due to the coherence may produce fluctuations in light intensity. When the optics is aligned perfectly a self-coupling effect can occur, i.e. light is coupled back to the laser and this coupling causes fluctuations in the output light of the laser.

Semiconductor lasers automatically eliminate these problems because they have low coherence. Such lasers however are usually not available at high power, but recently became available with $\sim$ 100 mW at 1.3 μm wavelength. The main disadvantage of semiconductor lasers is that they usually have higher noise levels than gas or solid state lasers.

Also important in the optical design is the type of scanning. Various scanning methods have been used and these can be divided into two types, on-axis and off-axis. The on-axis type uses a stationary optical beam and a mechanically scanned specimen, while the off-axis type uses a scanned optical beam and a stationary specimen. The main advantage of on-axis scanning is that the system is free from off-axis aberrations (coma and curvature of field), provided that the probe forming lens is perfect. The field of view is solely determined by the peak-to-peak movement of the scanners and is not limited by the lens. The linearity of the scan does not depend on the optics, only on the quality of the scanners. The main disadvantage of on-axis scanning is that mechanical scanning of

the specimen requires high quality mechanical scanners. The major requirement is a fast movement over a large area, e.g. up to 1 mm. This demand has not been satisfied yet by any of the scanners available on the market. Extremely good accuracy, repeatability and linearity of the scanning movement are required

The alternative way is to use off-axis scanning and this can be done by either galvano, acousto-optical or rotating type scanners. The most common is the Nipkow disk which is utilised in most direct-view or tandem scanning microscopes. The main advantage of this type of scanning is that images can be produced in real-time which is not possible with most on-axis type microscopes when recording a single image can take from several seconds up to several minutes. The price that must be paid for the faster frame time of the off-axis type microscope is the dominance of off-axis aberrations in the image, the difficult optical design and the strongly limited field-of-view.

For the present SIRM, we continued to use on-axis scanning. The main reasons were the need for a large field-of-view and high flexibility, which could not have been provided by an off-axis system.

10.3 Reflection confocal measurements

10.3.1 INTRODUCTION

In this work we have mainly concentrated on the reflection confocal SIRM measurements. This is because this imaging mode has newly been developed and the performance of the system has never been studied before. The experimental set-up is shown in Fig. 10.6(a). For this imaging mode a high performance Nikon infra-red lens is used with a continuously adjustable spherical aberration correction collar. The free working distance of this lens is $\sim$ 1 mm and it possesses a numerical aperture of 0.85. The lens is finite tube-length corrected (210 mm) and has a magnification of 80 $\times$. A 30 mW semiconductor laser is mounted on a block of three translators (a) to allow the accurate positioning along x, y and z directions. The exit aperture of the laser is imaged to an intermediate image plane by a lens which is also mounted on a block of three translators (b). The intermediate image is

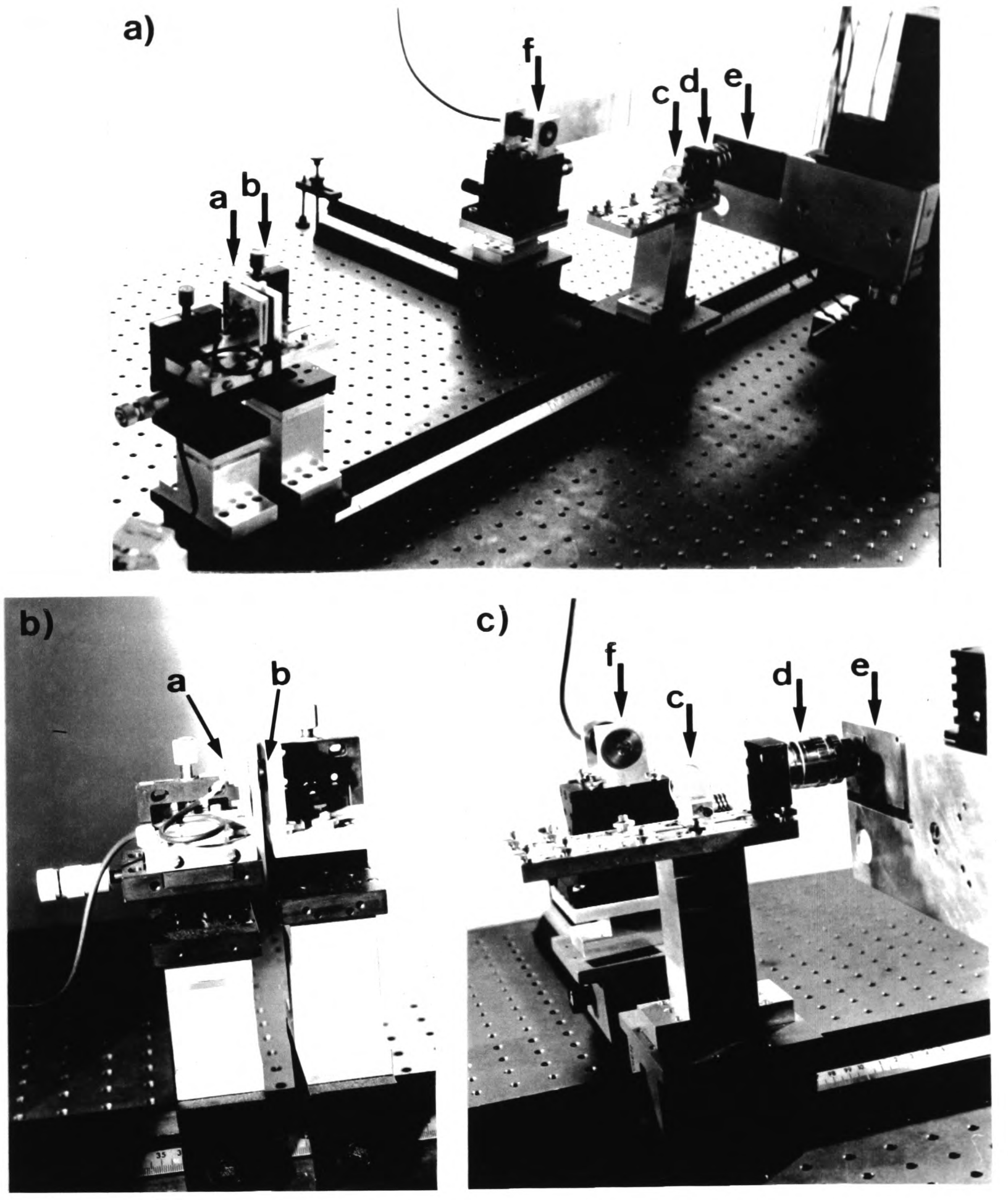


Figure 10.6: The reflection confocal SIRM. Figure (a) shows the whole system, whilst Figures (b) and (c) show details of the optics.

formed at 210 mm from the shoulder of the objective lens (d) which is firmly mounted on a micro bench. The light emerging from the imaging lens (b) traverses a non-polarising beam-splitter (c) and is incident on the input aperture of the objective lens (d). The numerical aperture of the imaging lens (b) is selected to match the numerical aperture of the laser (a) or serious light loss would occur. The distance at which this lens must be placed from the intermediate image plane is calculated from geometrical optics. The intermediate image plane of the imaging lens (b) coincides with the object plane of the objective lens (d) and is imaged into the specimen (e). Light reflected/scattered from the specimen is collected by the objective lens (d) and is directed via the beam splitter to the pin-hole/detector complex (f). Details of the laser/imaging lens assembly are shown in Fig. 10.6(b) and of the detector assembly in Fig. 10.6(c). The pin-hole/detector complex (f) is mounted on a robust three-axes translator that provides a submicron accuracy for the alignment.

The alignment procedure used, owing to the difficulty of the invisible laser radiation, is initially to align coarsely most of the optical components on the optical axis using a visible light (usually HeNe) laser. The semiconductor laser is then replaced and its correct x, y and z position determined by putting an opaque and diffuse screen at the required intermediate image position and observing the screen on an infra-red camera. Once the intermediate image is formed the opaque screen is removed and the correction collar of the objective lens is set to 0 mm. A mirror is placed and scanned axially through the focal plane, the reflected light is observed by the infra-red camera and the laser and the lens are aligned until the image is close to the pin-hole. After the laser is aligned the next step is to align the pin-hole, and for this large pin-hole sizes are initially used. Maximisation of the peak intensity and minimisation of the side lobes as the mirror scans through the focal plane enables the alignment to be made. The optimum pin-hole size is determined by measuring the FWHM and the peak intensity of axial intensity scans and by comparing these results with the theoretical predictions of Wilson[113].

10.3.2 THE SPECIMENS

To demonstrate the imaging capabilities of the SIRM a series of specimens was chosen. These specimens were prepared by Dr Robert Falster and his colleagues of MEMC Electronic Materials SpA. in Novara, Italy. The specimens were n-type (001) Czochralski silicon wafers 650 μm thick that had been well polished on the front side and less well polished on the back side. The wafers had been heat treated to produce oxide precipitate particles from the interstitial oxygen present in the silicon lattice. The wafers had received a three-step heat-treatment of 900°C for 15 minutes, 800°C for 4 hours and 1000°C for 4 hours (Table 10.1). The first anneal produced a surface denuded zone and dissolved some of the oxygen precursors in the bulk, the second nucleated the oxide particles and the third caused the particles to grow. Some wafers were further annealed at 1000°C, which caused the particles to grow further. On going from specimen *BM-0* to specimen *BM-6*, the particle size progressively increased while the particle number density remained approximately the same.

In addition to the specimens listed in Table 10.1, other specimens that had been given different heat treatments were examined. These included similar specimens with different oxide particle sizes and/or number densities, and specimens containing oxide particles that had been subsequently heat treated so that copper diffused in and precipitated at the oxide particles.

For the specimens examined in reflection, the SIRM was focussed through the front surface into the specimen. For the specimens examined in transmission, the specimens were additionally well polished (at Oxford) on the back side. The SIRM could then be focussed through either the front or back surface. The additional polishing was performed so as not to lose resolution from light leaving the specimen through a non-smooth surface.

Specimen	1st anneal	2nd anneal	3rd anneal	Additional anneals
BM-0	900°C/15 min.	800°C/4 hr.	1000°C/4 hr.	none
BM-1	900°C/15 min.	800°C/4 hr.	1000°C/4 hr.	1000°C/2 hr.
BM-2	900°C/15 min.	800°C/4 hr.	1000°C/4 hr.	1000°C/4 hr.
BM-3	900°C/15 min.	800°C/4 hr.	1000°C/4 hr.	1000°C/16 hr.
BM-4	900°C/15 min.	800°C/4 hr.	1000°C/4 hr.	1000°C/32 hr.
BM-5	900°C/15 min.	800°C/4 hr.	1000°C/4 hr.	1000°C/85 hr.
BM-6	900°C/15 min.	800°C/4 hr.	1000°C/4 hr.	1000°C/160 hr.
BB-5	1100°C/15 min.	800°C/4 hr.	1000°C/4 hr.	1250°C/35 sec. RTA

Table 10.1: Heat-treatments details of specimens examined.

10.3.3 EXPERIMENTAL RESULTS

These measurements were aimed at two goals. First, to investigate the contrast mechanism occurring when particles with different size are imaged in the reflection confocal mode. Second, to determine the optimum conditions at which the microscope gives the best image quality. The ultimate aim is to image small particles with very high number density.

Images were recorded, unless otherwise stated, under the following conditions. The focussing depth was $\sim$ 300 μm below the front surface. The correction collar of the microscope objective lens was set to this value on the calibrated scale and fine adjustments were made on observing a number of cross-section (x-z) images of particles (see Fig. 10.13). The images presented in the following are raw images, i.e. no image processing was applied. The signal of the detector was amplified by the Keithley current amplifier at 10^7, 10^8 and 10^9 A/V settings with no integration applied. No offset was added to or subtracted from the images. For reflection images, the contrast was taken as $(s_p - s_b)$, where s_p is the particle signal and s_b is the background signal, and generally $s_p \gg s_b$. For transmission images, the contrast was taken as $(s_b - s_p)/s_b$, and generally $s_b \gg s_p$. For reflection images, the contrast is given in mV, for transmission images, in per cent.

The specimens *BM-0* to *BM-6* were initially examined in plan-view, i.e. x-y images, using the SIRM in the reflection confocal mode. The particle number density was $\sim$ 10^8 cm^{-3} and this gave image spots that were well separated and suitable for line-trace

contrast measurements.

For specimen *BM-0*, no particles were detected. For *BM-1*, the particle signals were comparable with the noise level and although the particles could be detected, they were difficult to record.

For *BM-2*, images are shown in Fig. 10.7 (scale bar 20 μm). The six images comprise a focal series with the spacing between successive images of $\Delta z = 1.3$ μm. The two bright spots correspond to two particles. As the focus changes, the spots become progressively stronger in intensity and then weaker, but the spot diameter does not change significantly. The FWHM for the stronger spots is ~ 0.7 μm, which corresponds to the lateral resolution. The lateral resolution is the same for all individual particles imaged with the reflection confocal SIRM for specimens *BM-2* to *BM-6*.

The axial FWHM can be obtained from plan-view images by taking a line trace across a particular bright spot and measuring the intensity maximum, and repeating this for a number of focussing depths. By plotting these intensities against axial depth, the axial FWHM can be directly determined. This FWHM value is usually referred to as the depth or axial resolution and can be used to calculate particle number densities from single plan-view images. This is done by counting the individual particles revealed in the image and dividing this by the area of the plan-view image multiplied by the depth resolution. The two particles of Fig. 10.7 gave depth resolutions of ~ 5.2 and 2.5 μm respectively.

Images taken from specimen *BM-3* are shown in Fig. 10.8. The scale bar corresponds to 40 μm. Successive images comprise a focal series with an axial spacing of $\Delta z = 1.3$ μm. On comparing these particles to those in Fig. 10.7, individual particles exhibit similar behaviour in intensity as the axial focus position is changed but show higher contrast which is indicated by a lower background level. Although the background level in subsequent images appears to change, it is not the case since the absolute background level for these images does not vary. The virtual change is caused by a software contrast normalisation procedure, i.e. when lower intensities are measured on particles, the programme normalises the contrast so that the highest intensity level is stretched to the maximum brightness level. The axial FWHM measured by the method described above gave depth resolutions

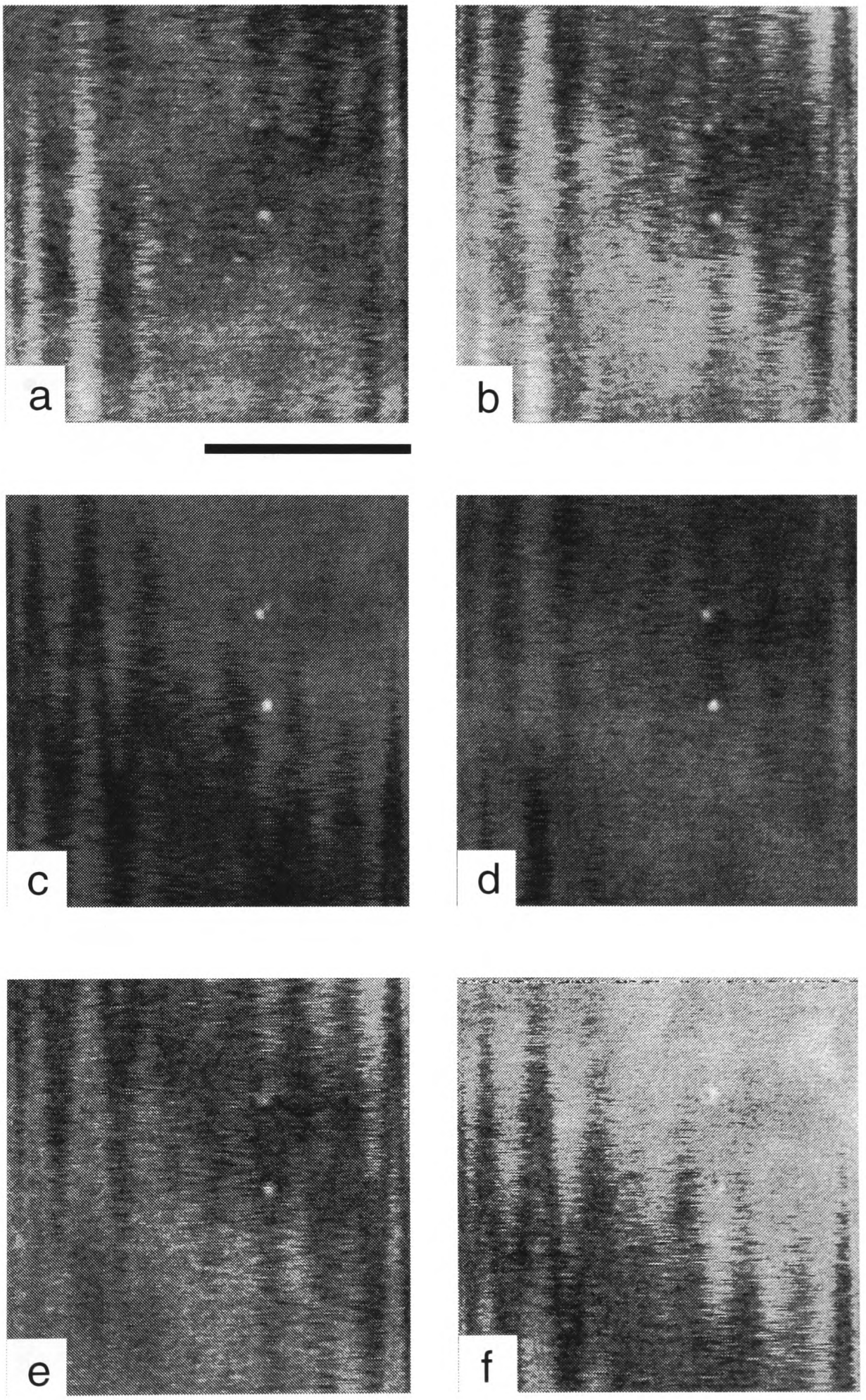
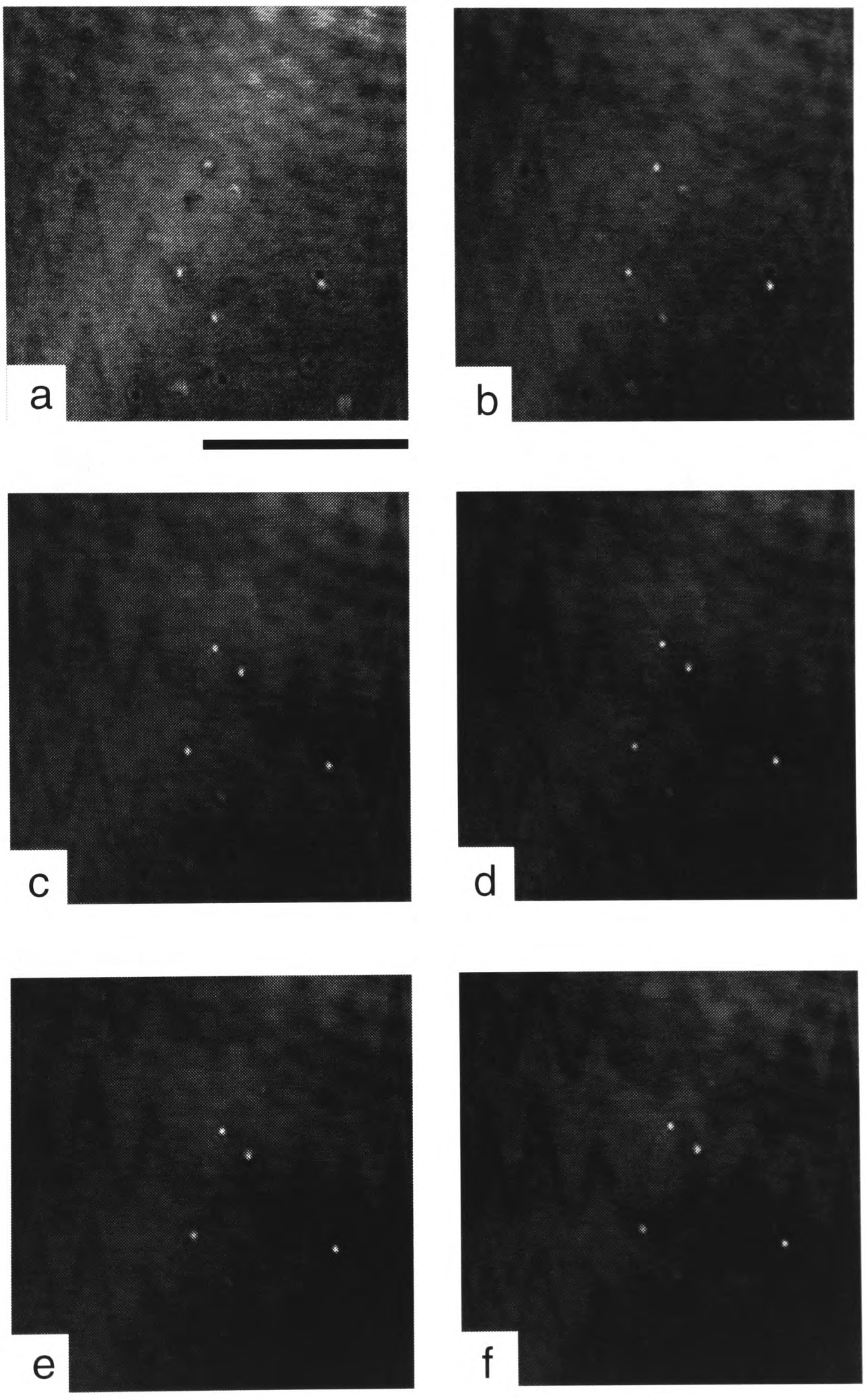


Figure 10.7: Focal series from specimen *BM-2*. The scale bar corresponds to 20 μm. The axial spacing between images is $\sim$ 1.3 μm.


a
b
c
d
e
f

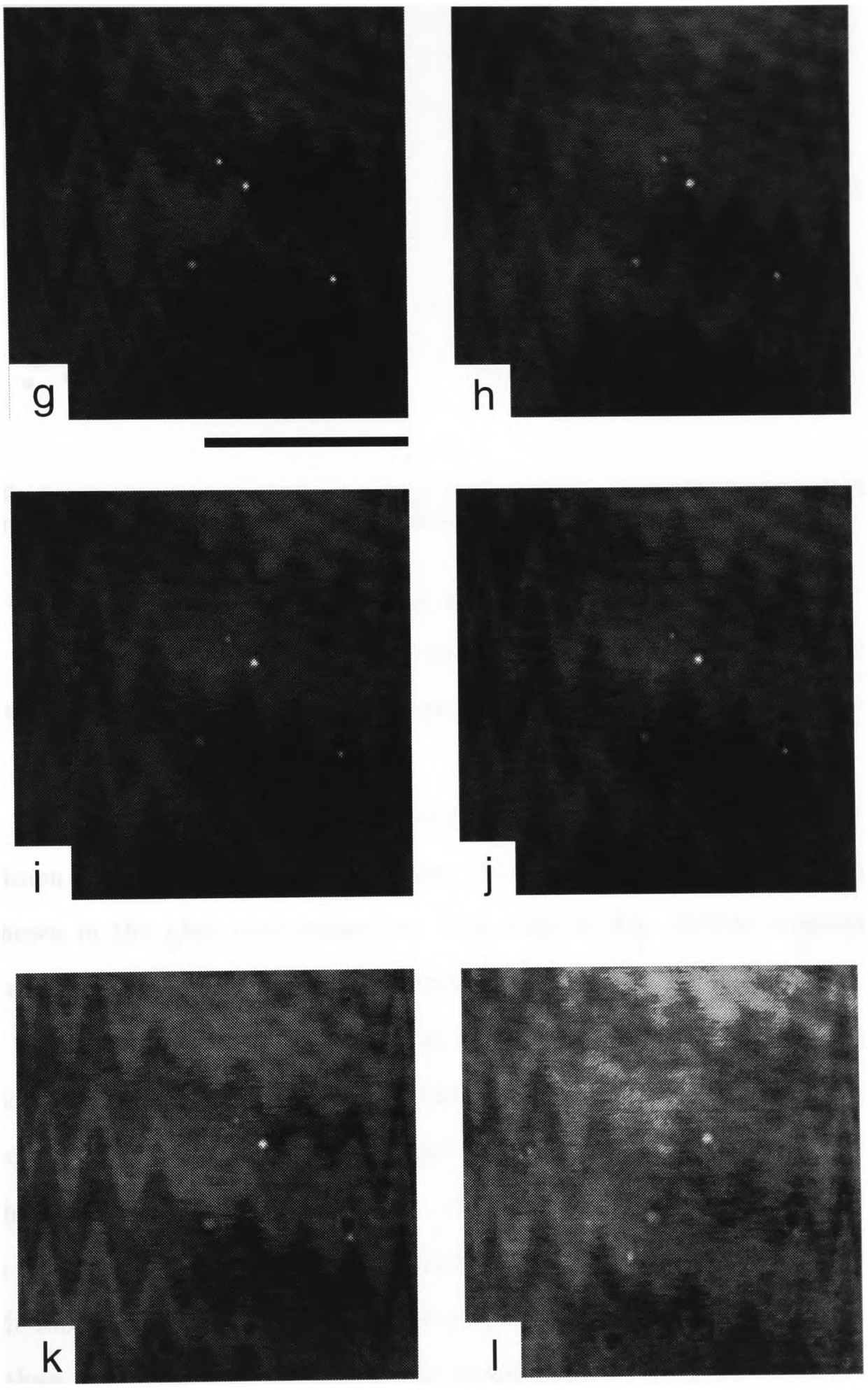


Figure 10.8: Focal series from specimen *BM-3*. The scale bar corresponds to 40 μm. The lateral spacing between images is $\sim$ 1.3 μm.

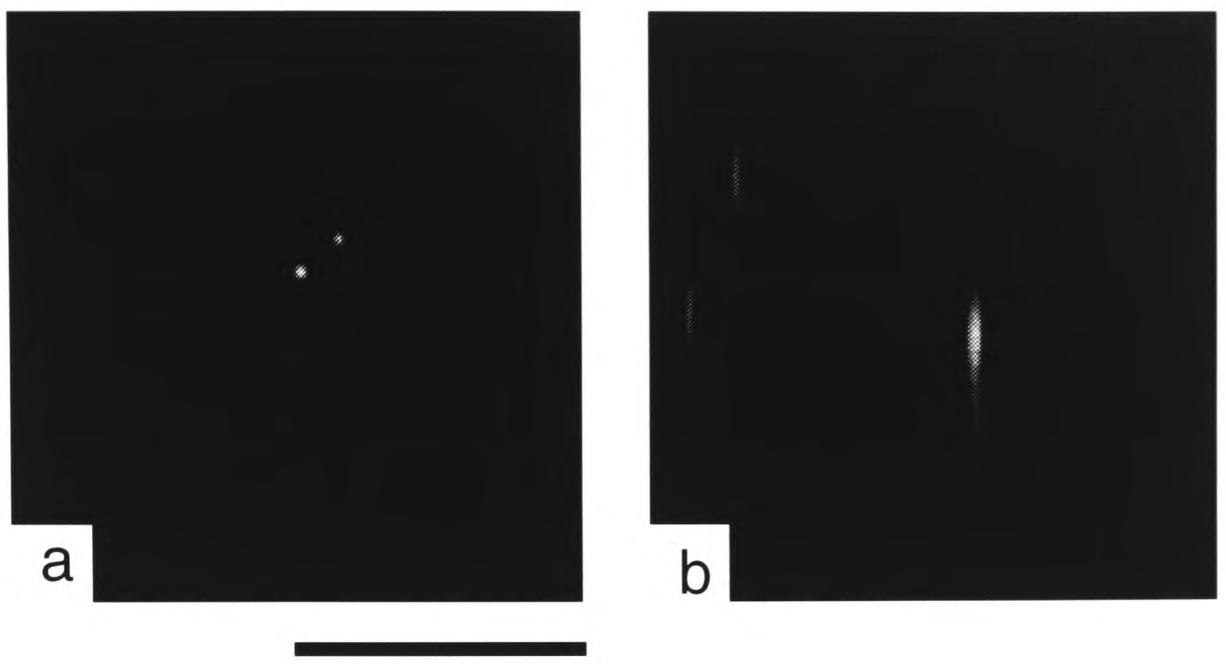


Figure 10.9: A pair of plan-view and cross-section image from specimen *BM-5*. Total depth imaged is 50 μm and the scale in lateral direction corresponds to 40 μm.

of $\sim$ 5.2 - 6.5 μm, but some particles gave a value of $\sim$ 9.1 μm.

Images taken from specimen *BM-4* (not presented) are similar to those shown in Fig. 10.8, but the particles exhibit higher contrast. The axial FWHM values are $\sim$ 5.1 - 6.6 μm.

In Fig. 10.9 a plan-view (x-y) image (a) and a cross sectional (x-z) image (b) are shown from specimen *BM-5*, where the cross section (b) is taken at the middle of the brighter particle shown in the plan-view image (a). The scale at Fig. 10.9(a) corresponds to 40 μm both along horizontal and vertical directions, whilst in Fig. 10.9(b) the scale only applies to the horizontal direction. The total vertical size of Fig. 10.9(b) is 50 μm. In the cross-sectional image the elongated bright 'spots' correspond to individual oxide particles. A line trace taken from the main bright 'spot' of image (b) in the vertical direction, and running through the middle of the particle, reveals that the intensity distribution of the particle is symmetrical at the main peak and less symmetrical at the side lobes (tails), i.e. the side lobe closer to the bottom of the image (closer to the surface of the specimen) is longer than that closer to the top of the image (farther from the specimen surface). Some additional particles are also seen in the cross-sectional image which are at different x-y planes than the observation plane of Fig. 10.9(a). These particles are less bright

than the particle discussed above and this is because of they are situated at different lateral positions than that of the image plane. Cross-sectional images can also be used to determine the axial resolution of the microscope. Vertical line traces through the middle of the particle reveal the intensity distribution. By taking the FWHM of this distribution, the depth resolution is given.

In Fig. 10.10 a focal series is shown from specimen *BM-6*. The scale bar corresponds to 10 μm. Successive images comprise a focal series with an axial spacing of $\Delta z = 3.5$ μm. The images reveal a cluster of closely separated oxide particles. Such clusters arise because of the long particle growth anneal (160 hours). They were previously observed by TEM[57]. Figure 10.10 (f) is a TEM micrograph[73] taken from a specimen which received the same growth anneal as *BM-6*. It shows a cluster of oxide particles (and associated dislocations) similar to that imaged in Fig. 10.10 (a-e). Arrows indicate oxide particles which would be imaged by SIRM as bright spots, with some spots separate and others joined together.

To show how reflection confocal SIRM can image particles in specimens with different particle number densities, Fig. 10.11 is presented. Specimens (not listed in Table 10.1) were heat treated to produce a medium particle size with medium, high and very high particle number densities. Etching showed the densities to be $\sim 10^8$, $\sim 10^9$ and $\sim 10^{10}$ cm^{-3} respectively. These specimens were then assessed by reflection confocal SIRM to determine particle number density. In Fig. 10.11, image (a) corresponds to the medium, (b) to the high and (c) to the very high particle number density values. The scale bar corresponds to 80, 40 and 10 μm in (a), (b) and (c), respectively. These images have been image processed by means of a background subtraction. This image processing should change neither the particle size nor the particle number density. The particle number densities determined were 8.9×10^7, 9.1×10^8 and 8.7×10^9 cm^{-3}. It is clearly seen in Fig. 10.11(c) that even at the very high particle number density, the individual particles are imaged with good contrast. From such images it can be concluded that the highest oxide particle number density which can be detected in silicon with the reflection confocal SIRM (NA=0.85, $\lambda = 1.3\mu$m) is $\sim 10^{11}$ cm^{-3}, and the lowest number density is $\sim 10^6$

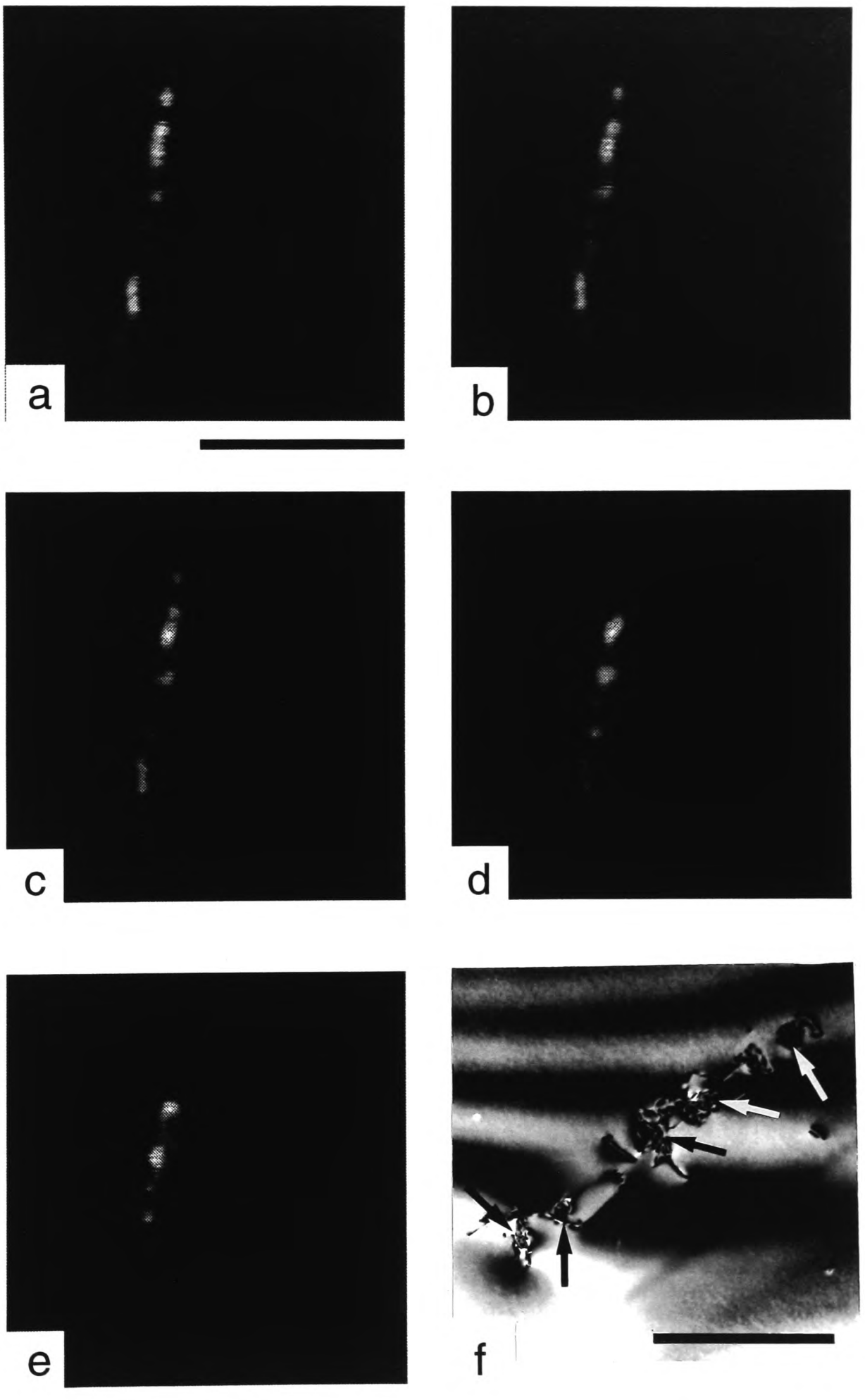


Figure 10.10: Focal series from specimen *BM-6* (a-e). The scale bar corresponds to 10 μm. The lateral spacing between images is $\sim$ 3.5 μm. Figure (f) is a plan-view TEM image (scale bar 2 μm) taken from a specimen which received the same growth anneal than *BM-6*. TEM image courtesy of Dr B. Pécz.

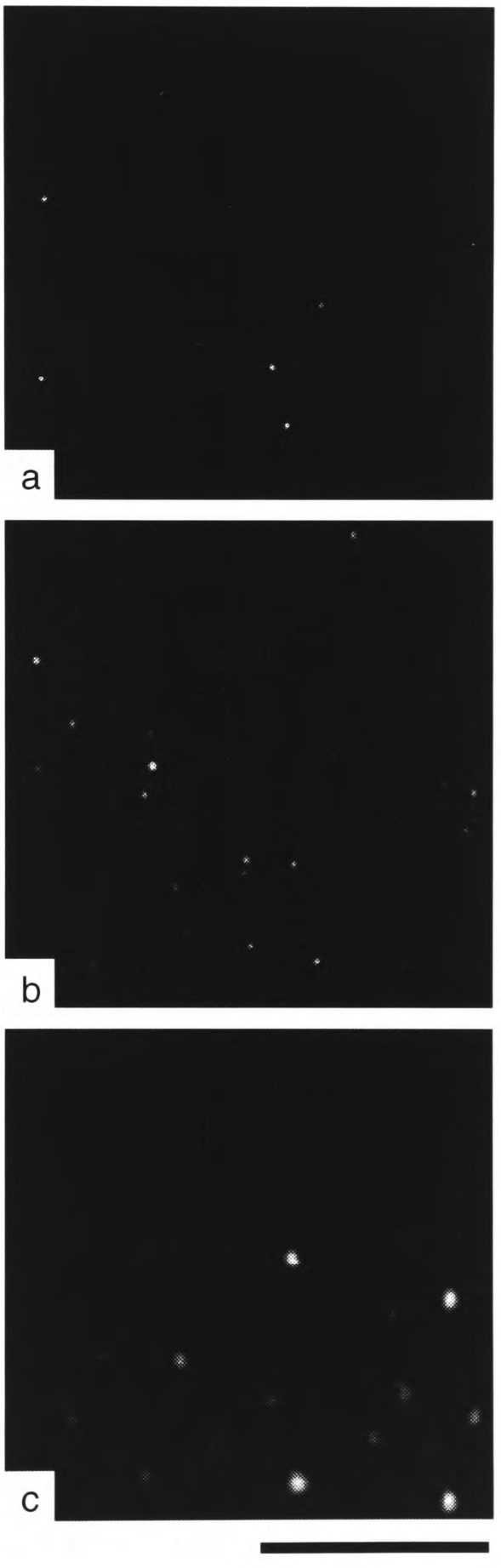


Figure 10.11: Images of specimens with different particle number densities. For details see text

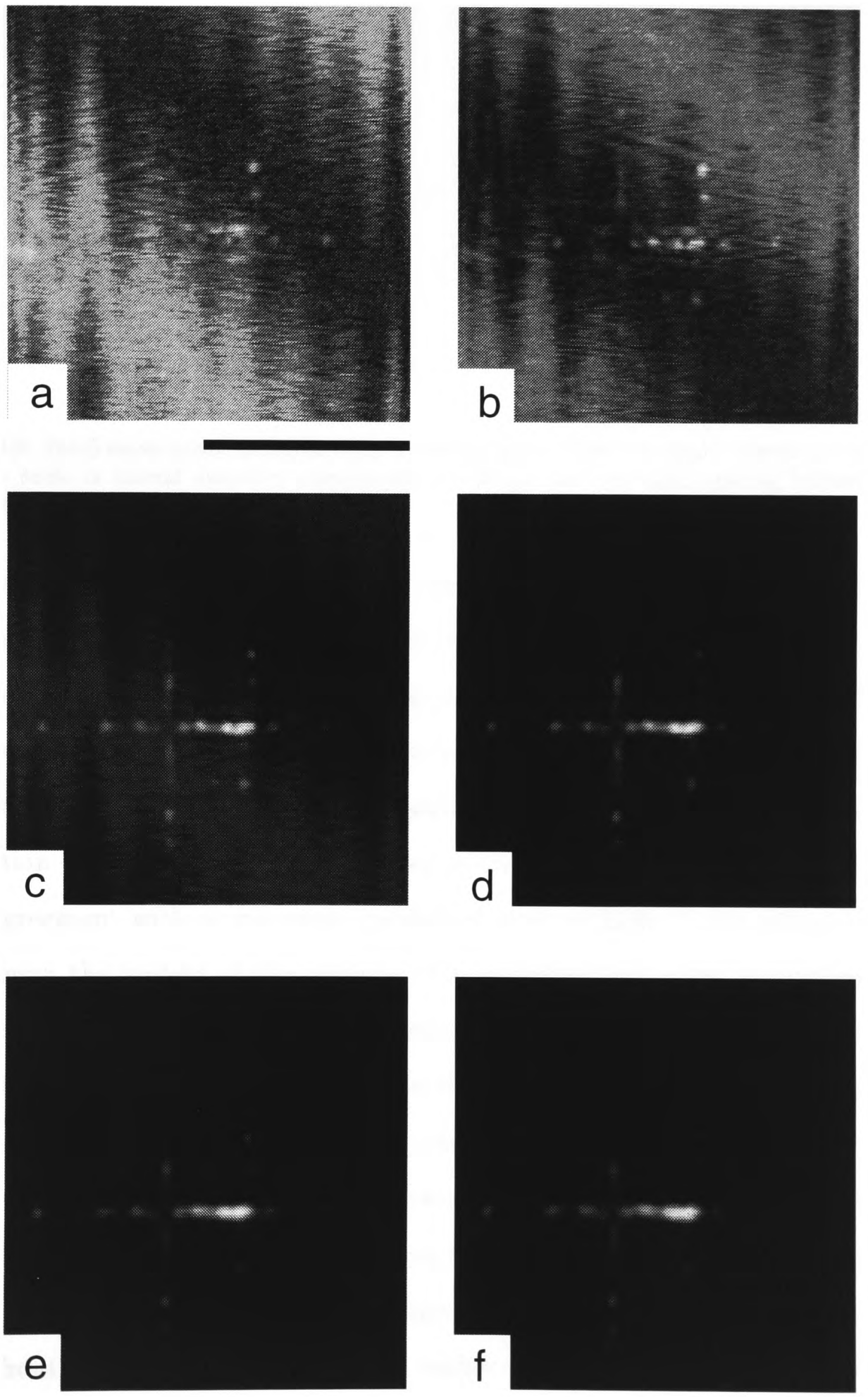
cm^{-3} (corresponding a 360×360 μm^2 field of view).

Spectacular images can be obtained by the reflection confocal SIRM from some types of specimen. In Fig. 10.12 a focal series is presented which was obtained from a specimen prepared as follows. Specimen *BM-0* was contaminated by scratching its back surface by a pure copper wire. Subsequently a heat treatment was given to the specimen at 700°C for 2 hours. During this anneal copper atoms diffused into the bulk. The specimen then was cooled down from 700 to 400°C with a cooling speed of 3°C/sec. During the cooling down copper atoms precipitated at oxide particles which were present in the specimen as a result of prior heat-treatments discussed above. Then specimens were assessed by the reflection confocal SIRM and it was found that copper precipitates at oxide particles and forms cross shaped clusters. A similar result was also obtained by transmission non-confocal SIRM examination prior to the reflection confocal measurements. It is seen in the images that, as the focussing depth progressively changes, different parts of the cluster is imaged so revealing its three-dimensional structure.

10.3.4 EFFECT OF THE SPHERICAL ABERRATION

To assess the effect of spherical aberration present in the system and introduced by the specimen, a series of measurements was performed. A particle at $\sim$ 300 μm depth in specimen *BM-4* was imaged in cross section. The correction collar of the microscope objective lens was set to zero and a cross-sectional image was recorded (Fig. 10.13(a)). In the images the scale bar corresponds to 20 μm along the horizontal direction and the total depth imaged is 50 μm.

Then the collar was turned to a correction of 0.1 mm and the next image was recorded (Fig. 10.13(b)). Subsequent images were taken with a correction setting of 0.2, 0.3, 0.4 and 0.5 mm and the corresponding images are shown in Fig. 10.13(c-f) respectively. Since the particle is situated at $\sim$ 300 μm depth, when the correction setting was 0, 0.1, 0.2, 0.3, 0.4 and 0.5 mm, the focussing depth equivalent of the spherical aberration was 300, 200, 100, 0, -100 and -200 μm respectively, where the negative values correspond to over-correction, and the positive values correspond to undercorrection. The elongated bright


a
b
c
d
e
f

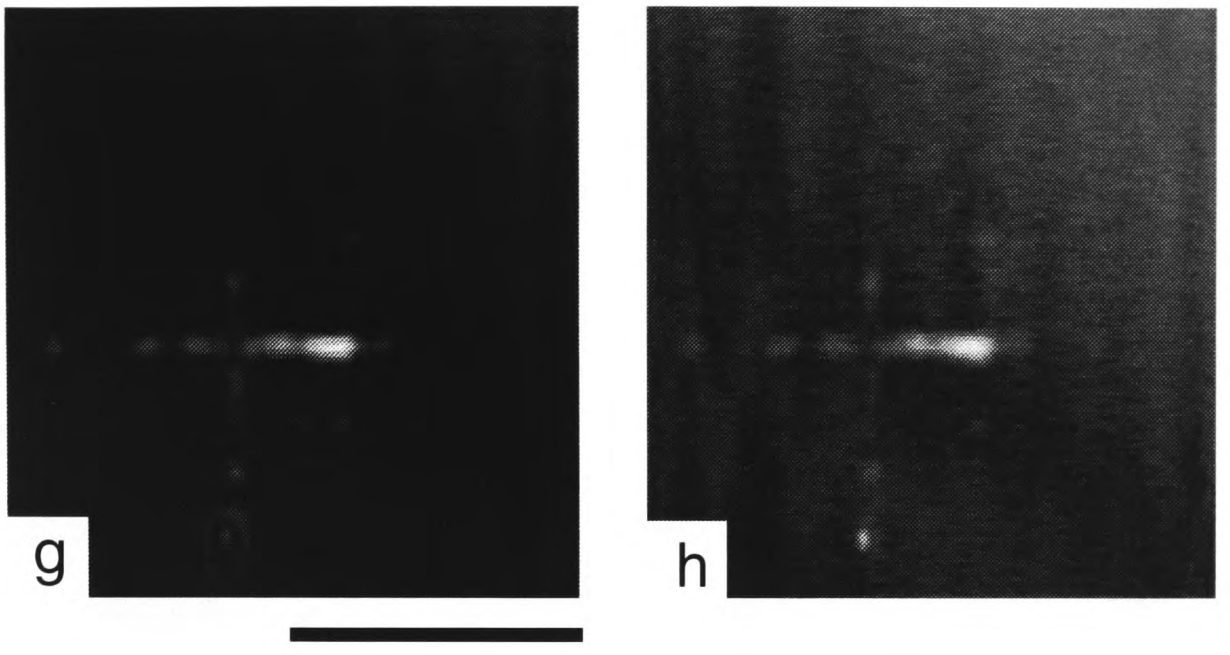


Figure 10.12: Focal series from specimen copper contaminated *BM-0* specimen showing a copper silicide cluster. The scale in lateral direction corresponds to 80 μm and the axial spacing between individual images is $\sim$3.5 μm.

'spot' corresponds to a single particle. The lateral FWHM of the particle (taken at intensity maximum) increases with the presence of different amount of spherical aberrations but this increase is not substantial. The axial FWHM of the particle (measured by the method described above) increases progressively as the spherical aberration increases in either direction. At negative spherical aberrations this increase is more significant, i.e. the overcorrection of spherical aberration causes serious loss in axial resolution. This finding is in full agreement with a previously published work of Keller[51] By taking vertical line traces through the middle of the particle, it is revealed that when no spherical aberration is present (d) the main peak is symmetrical with respect to the axial position of the highest intensity. When spherical aberration is present the line traces become asymmetrical. The axial image position of the particle is shifted as the spherical aberration varies, at positive aberration positive shift occurs, whilst at negative aberrations negative shift occurs. The horizontal fine dark lines in the particle are due to mechanical scanning noise. The horizontal dark and bright background fringes on some images are due to the specimen front surface and they are only visible when the intensity obtained from the particle is not sufficiently high to suppress the effect when the contrast normalisation is performed.

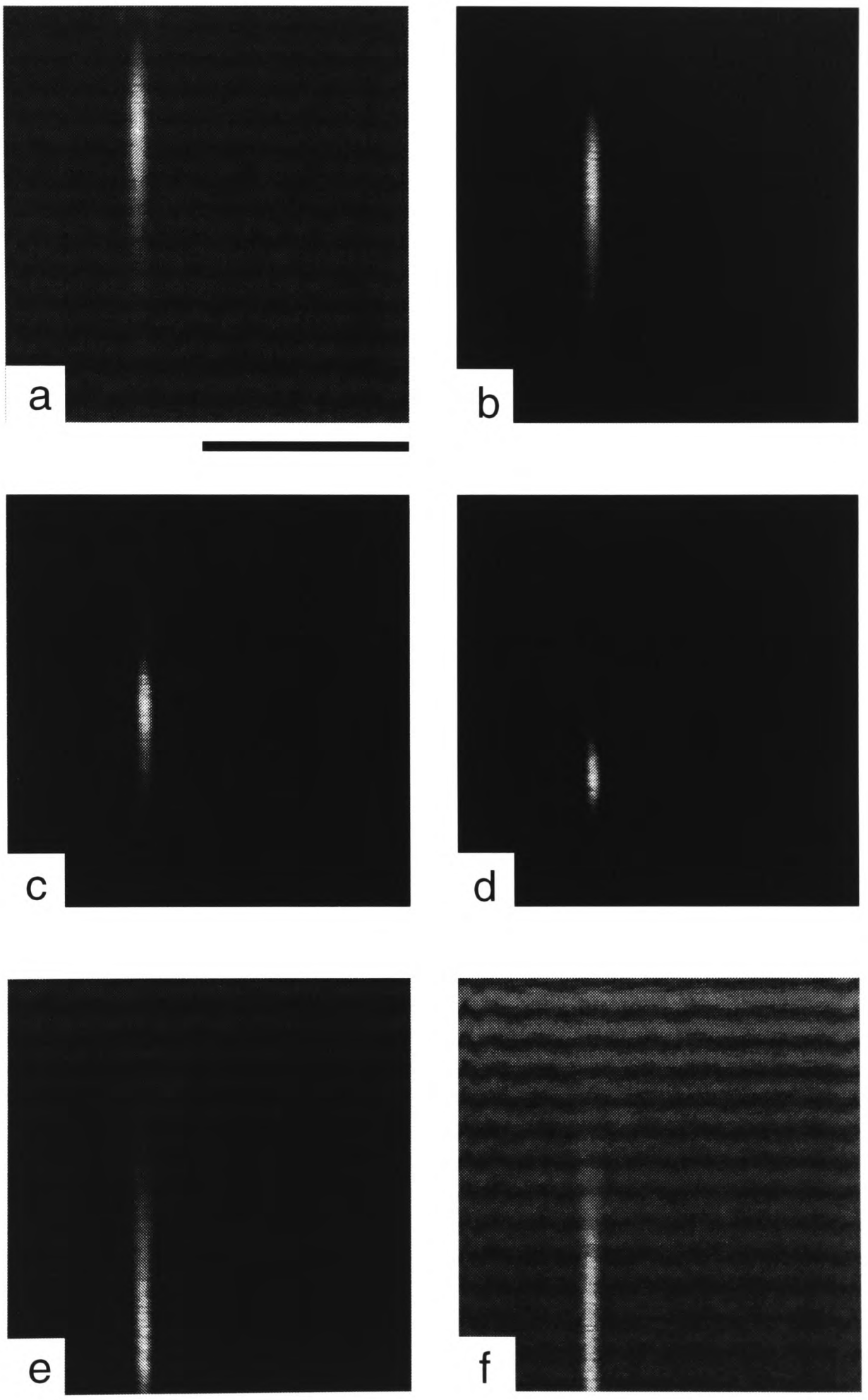


Figure 10.13: Cross sectional images from specimen *BM-4*. The focussing depth is 300 μm. Individual images (a to f) correspond to 0, 0.1, 0.2, 0.3, 0.4 and 0.5 mm correction setting respectively. The total depth imaged is 50 μm and the scale in lateral direction corresponds to 20 μm.

Mode	*BM-0*	*BM-1*	*BM-2*	*BM-3*	*BM-4*	*BM-5*	*BM-6*
reflection [10 mV]	-	~ 4	10.3	18.1	163.7	472.1	675.3
transmission [%]	-	-	-	1.9	3.9	7.7	7.9
scattering theory [a.u.]	-	~ 0.49	1.28	2.25	20.36	58.72	84.28
equivalent size [nm]	-	~ 63	88	97	182	273	325*

Table 10.2: Contrast values by reflection confocal and transmission non-confocal SIRM.

10.3.5 CONTRAST

The contrast values were measured by the reflection confocal SIRM as follows. The background in all of the specimens *BM-1* to *BM-6* were measured to be the same. This is partly due to the same depth at which the images were recorded and partly to the medium particle number density, i.e. it is not high enough to substantially contribute to the background signal by out-of-focus scattering. The contrast values obtained for each of specimens *BM-1* to *BM-6* are results of many averaged images, where the maximum intensity is measured on a focal series and intensity maxima are averaged for many individual particles.

The experimental contrast values are shown in Table 10.2 (first line) and for comparison we also give the contrast values measured previously on the same specimens by transmission non-confocal SIRM[57] (second line). If the reflection confocal SIRM contrast was plotted as a function of the equivalent particle diameter (size), then a comparison could be made between this experimental data and the theoretical data of Chapter 8. Unfortunately this could not be done since there are no reliable or accurate values from TEM examinations over the full specimen range for the equivalent particle diameters. Consequently, to interpret the experimental contrast values, the equivalent particle diameter for specimen *BM-6* was taken to be 325 nm based on TEM observations[57,73]. This particle size was used as a 'fixed value' to calibrate the scale of the measured contrast with the theoretical scale. The result of the calibration is the third line in Table 10.2. From the curve of Fig. 8.16 the equivalent particle size [nm] can be deduced which yields the fourth line. We note that the deduced equivalent particle size for specimen *BM-1* (63 nm)

is in good agreement with that based on TEM observations (70 nm). It can be seen from the above data that a sudden jump occurs in the reflection contrast between specimens *BM-3* and *BM-4*. Although the contrast value corresponding to specimen *BM-4* is $\sim 9\times$ greater than that of *BM-3*, this ratio for the equivalent diameter is only $\sim 1.9\times$ which finding is a consequence of the dependence of the scattered intensity on sphere diameter, as discussed in Chapter 8.

It can be briefly summarised that, depending on the particle which is being imaged with the reflection confocal SIRM (NA=0.85, $\lambda = 1.3$ μm), the best depth resolution that can be achieved is 5 to 7 μm and the best lateral resolution is ~ 0.7 μm and with these resolution values the detectable minimum and maximum particle number densities are 10^6 to 10^{11} cm^{-3} respectively. There are evidences from TEM examinations that reflection confocal SIRM examinations revealed some important features of the oxide particle formation.

10.4 Other imaging modes in the SIRM

When the SIRM is operating in the *transmission non-confocal* mode the optical set-up is highly simplified (Fig. 10.14). The microscope objective lens (d) images the object in the intermediate plane into the specimen (e) and the transmitted light is detected by a high numerical aperture detector (f) placed behind the specimen. Since this imaging mode has been the subject of many detailed investigations, images are not shown here. This imaging mode is the simplest, easiest to align and, when the specimen is double polished and the required resolution is not too high, this mode is the recommended way to obtain images provided that the difference between the transmission and reflection contrast mechanisms does not prevent the successful application. The best depth resolution that can be achieved by this mode is ~ 30 μm with a lateral resolution of ~ 1 μm, provided that the system is corrected for spherical aberration.

The SIRM can also operate in the *transmission confocal* mode. This set-up is shown in Fig. 10.15. The optics is now somewhat different from those previously described. The reason for this is that two infinite tube length corrected microscope objective lens (0.6

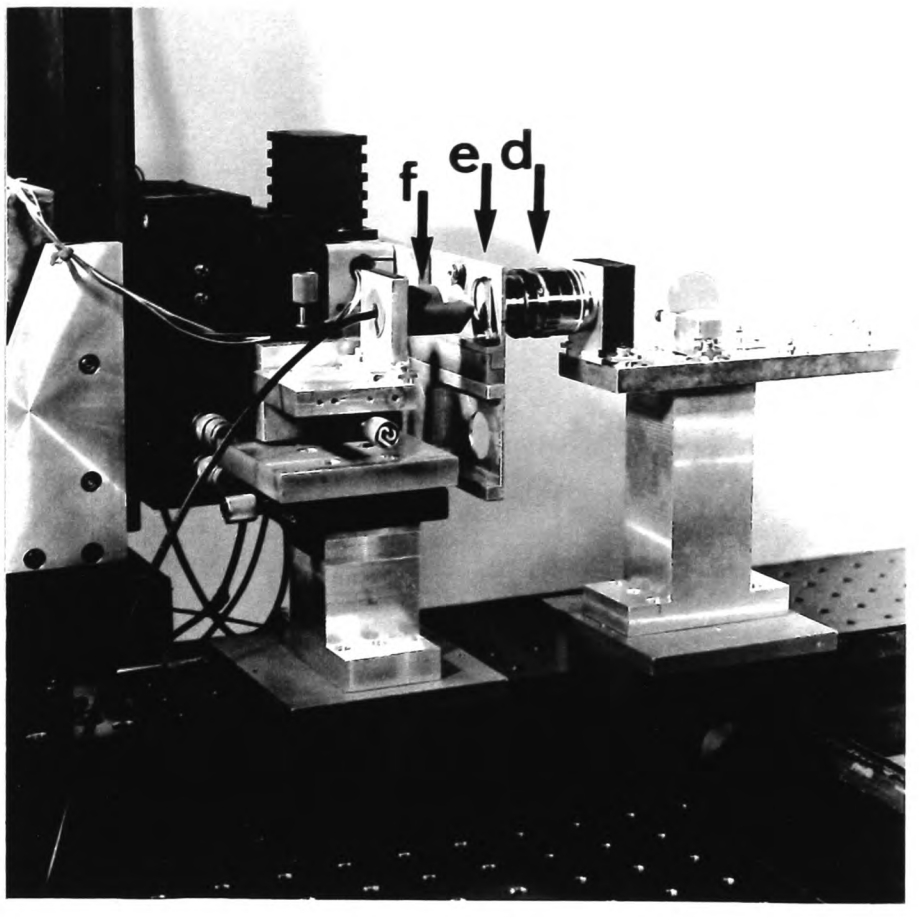


Figure 10.14: The transmission non-confocal SIRM.

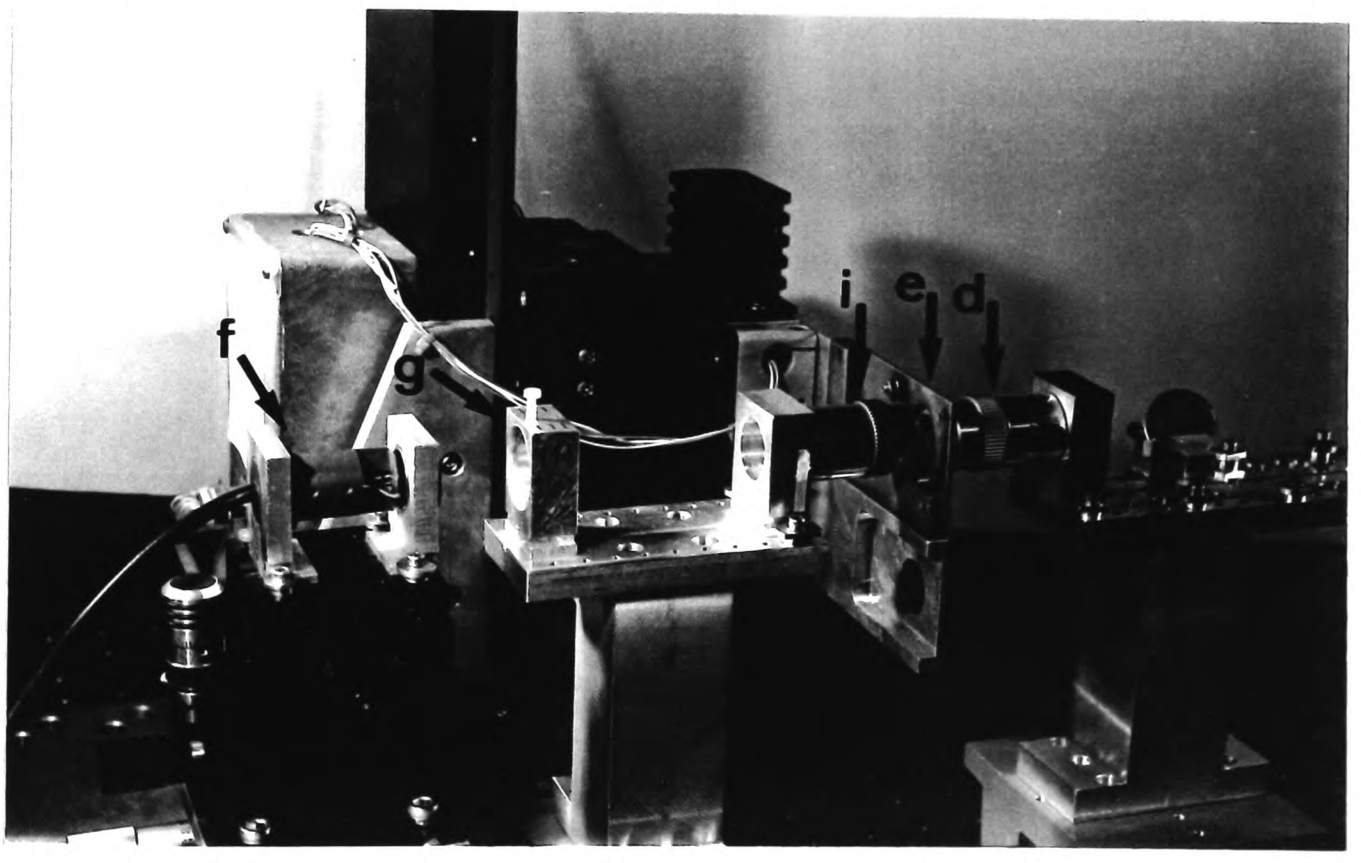


Figure 10.15: The transmission confocal SIRM.

numerical aperture, 0.5 mm cover-glass correction, 80 × magnification and 3 mm free working distance) are used, one of which produces the probe and the other which collects the transmitted light and re-collimates it. It is advantageous if these lenses possess the same numerical aperture. If the numerical aperture of the focussing lens (d) is smaller than that of the collector lens (i) then some loss occurs in resolution. In the opposite case, i.e. when the numerical aperture of the lens (d) is higher than that of lens (i) then some of the transmitted light is lost, i.e. not collected by the collector lens. These two objective lenses were selected for the transmission mode because of their high numerical aperture. When this arrangement is used then the collimated light after the second lens is re-focussed by the detector lens (g) to the pin-hole/detector complex (f). The focal length of the detector lens (g) with the focal length of the collector lens (i) determines the demagnification of the probe at the pin-hole. The size of the pin-hole can be calculated from the lateral size of the probe and the demagnification ratio. The best depth resolution which can be achieved by this is $\sim$ 7 to 9 μm with a best lateral resolution of $\sim$ 1 μm, provided that the system is corrected for spherical aberration. A typical transmission confocal SIRM image can be seen in Fig. 10.18 (a), where dark spots correspond to in-focus oxide particles. Out-of-focus particles can be seen as the larger bright blurred spots.

An analogous operation mode of the SIRM is the *transmission confocal differential phase contrast* (DPC) mode. This optical set-up is shown in Fig. 10.16. Light from the laser is collimated by a collimator lens (not shown in photograph) and is incident upon the focussing lens (d) which focusses the light into the specimen (e). The collector lens (i) is placed directly behind the specimen and it collects and re-collimates the transmitted light which is incident on a cube beam splitter (h). Half of the intensity is split towards the first detector arm (denoted by subscript 1) and the other half is transmitted towards the second detector arm (denoted by subscript 2). In the first detector arm a half stop is placed just before the detector lens (g_1) which stops the right half of the incident light. The other half is focussed by the detector lens and is incident on the pin-hole detector complex (f_1). The second detector arm is formed in the same manner, except that there the other half of the light is stopped. When the signals of the two detectors (f_1 and f_2)

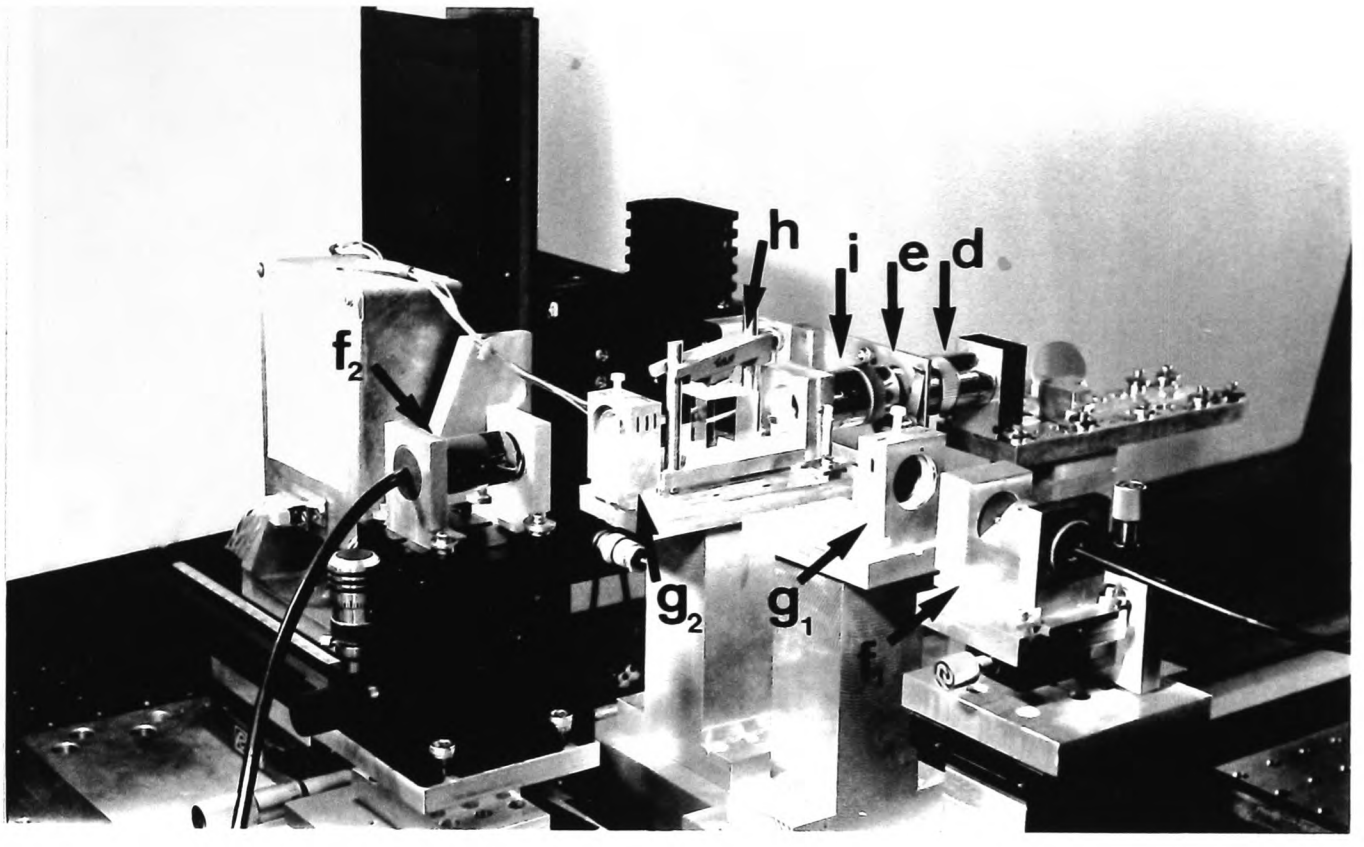


Figure 10.16: The transmission confocal DPC SIRM.

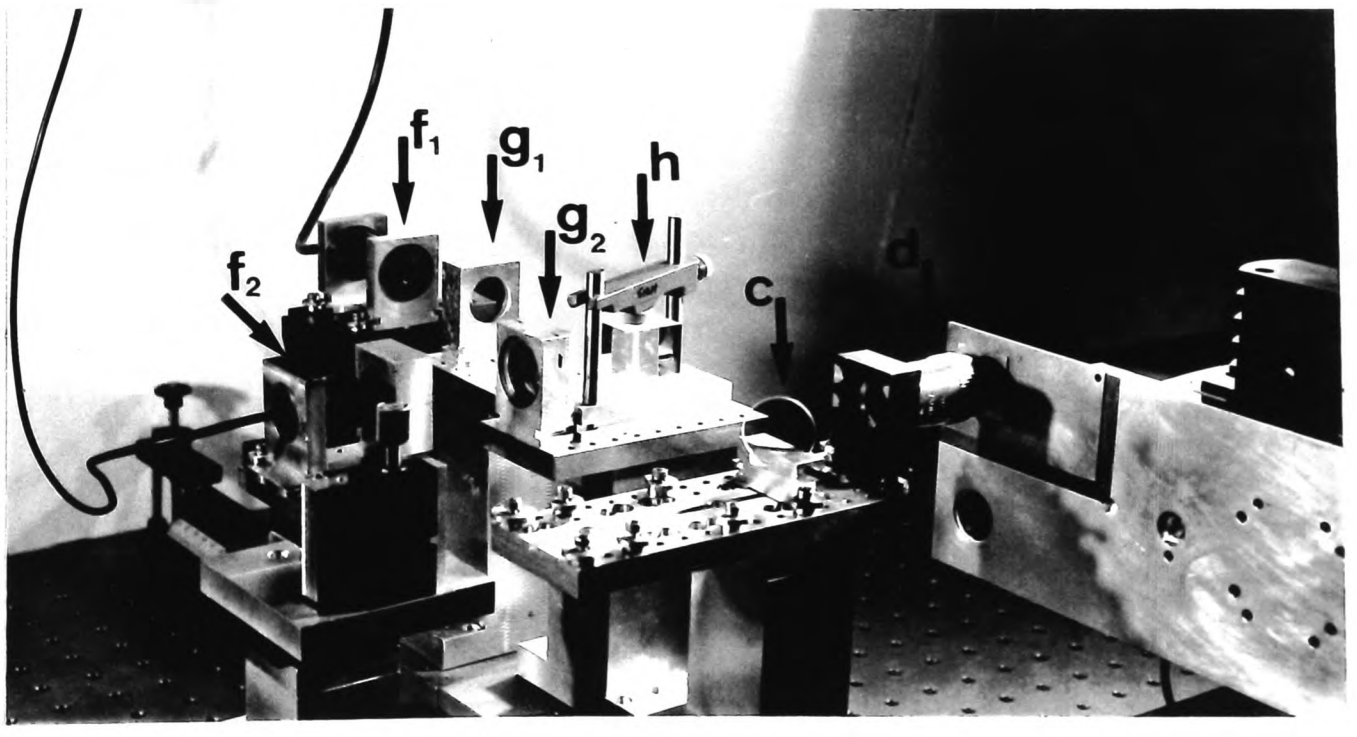


Figure 10.17: The reflection confocal DPC SIRM.

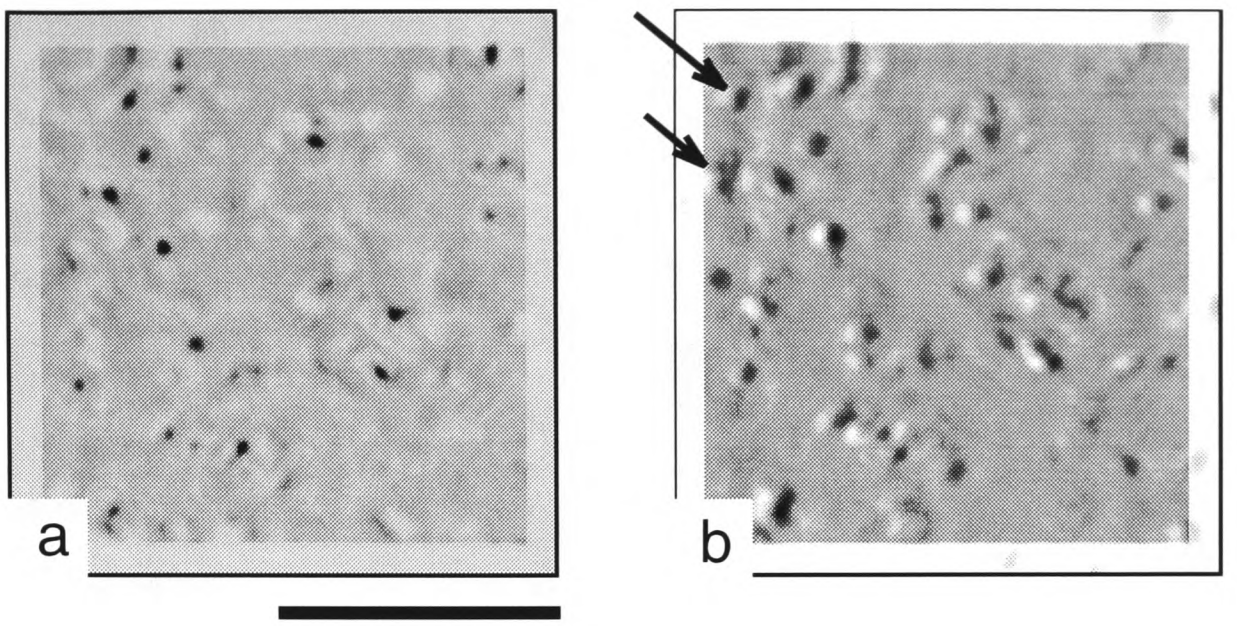


Figure 10.18: A pair of plan-view images from specimen *BM-4*. Figure (a) corresponds to a transmission confocal image and Figure (b) to a transmission confocal DPC image. Images are taken at 35 μm depth. The scale in lateral direction corresponds to 40 μm

are subtracted, in the case of a homogeneous specimen no signal is measured. When the signals of the two detectors (f_1 and f_2) are added, conventional confocal transmission imaging occurs. If however an inhomogeneity (e.g. a particle) is present in the probe, then on scanning the specimen orthogonally with respect to the direction of the two half-stops, DPC detection occurs. To illustrate the transmission confocal and transmission confocal DPC imaging modes a pair of images is shown in Fig. 10.18 obtained from specimen *BM-4* and from the same area.

The scale bar on the images correspond to 40 μm. It is important to note that some particles are revealed in Fig. 10.18(b), which corresponds to the DPC imaging mode, that were not visible with the conventional transmission confocal mode (two of such particles are indicated in Figure with arrows). These images were image processed to enhance contrast which should not change either the particle size or the number density.

The *reflection confocal DPC* mode was also built which set-up is shown in Fig. 10.17. This imaging mode operates in exactly the same manner as the transmission confocal DPC mode, except that here no transmitted light is detected. The focussing lens (d) collects the scattered light and re-collimates it. The collimated light is then incident on a cube beam splitter (h) after which the two detector arms are identical with the

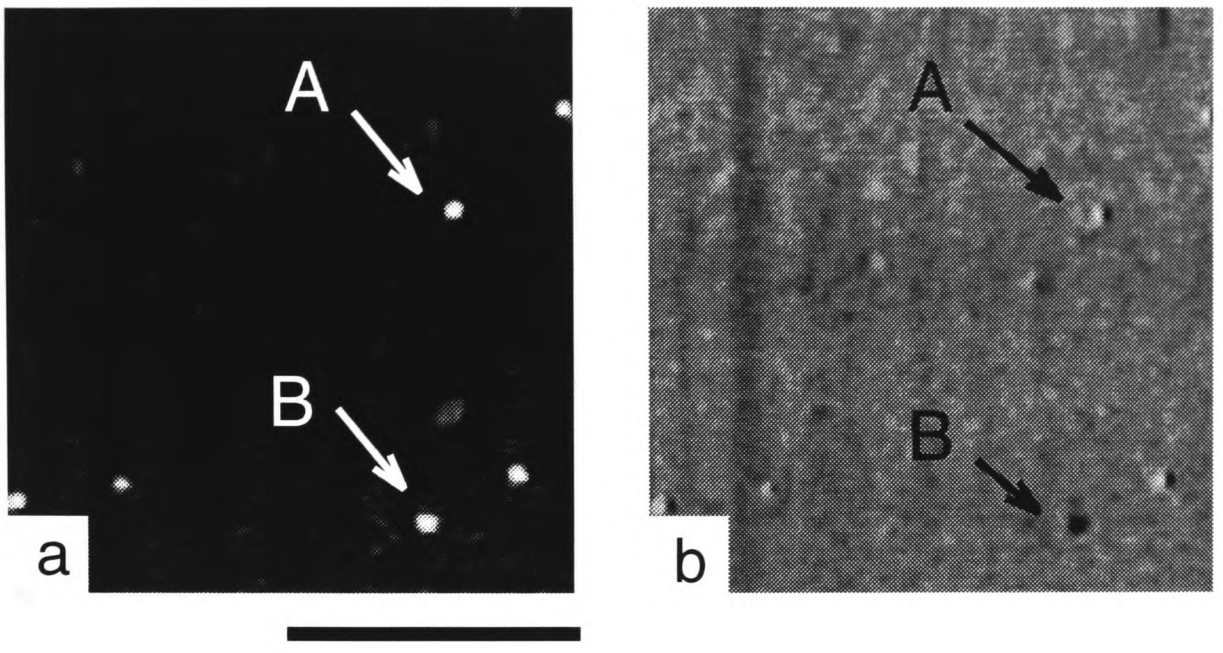


Figure 10.19: A pair of plan-view images from specimen *BB-5*. Figure (a) corresponds to a reflection confocal image and Figure (b) to a reflection confocal DPC image. Images are taken at 35 μm depth. The scale in lateral direction corresponds to 20 μm

corresponding transmission set-up. This imaging mode is of high importance since with perfect alignment of the optics it is expected to eliminate completely the surface reflection from the specimen. The second reason why this set-up might be favourable over the reflection confocal SIRM is shown in Fig. 10.19.

The specimen imaged was *BB-5* and the scale bar corresponds to 20 μm. In the figure, (a) and (b) show reflection confocal and reflection confocal DPC images respectively. The particle number density is $\sim 10^9$ cm^{-3}. Two particles are labelled in the figures. Both particle (A) and (B) are revealed in figure (a) with closely equal intensities. Particle (A) however in figure (b) exhibits a symmetrical contrast with equal but opposite sign intensities in the right and left hand side of the spot. Particle (B) in figure (b), unlike particle (A), exhibits only negative contrast. Subsequent cross-section images and focal series had shown that when a particle, measured in the DPC mode, exhibits symmetrical contrast, similar to that of particle (A) in figure (b), then that particle is in-focus and when the contrast loses the symmetry in any direction (positive or negative contrast), then the particle is out-of-focus. The depth sensitivity of this mode is far better than the depth resolution and is estimated to be ~ 1 μm.

Finally the *reflection confocal Nomarski differential interference contrast* (DIC) mode

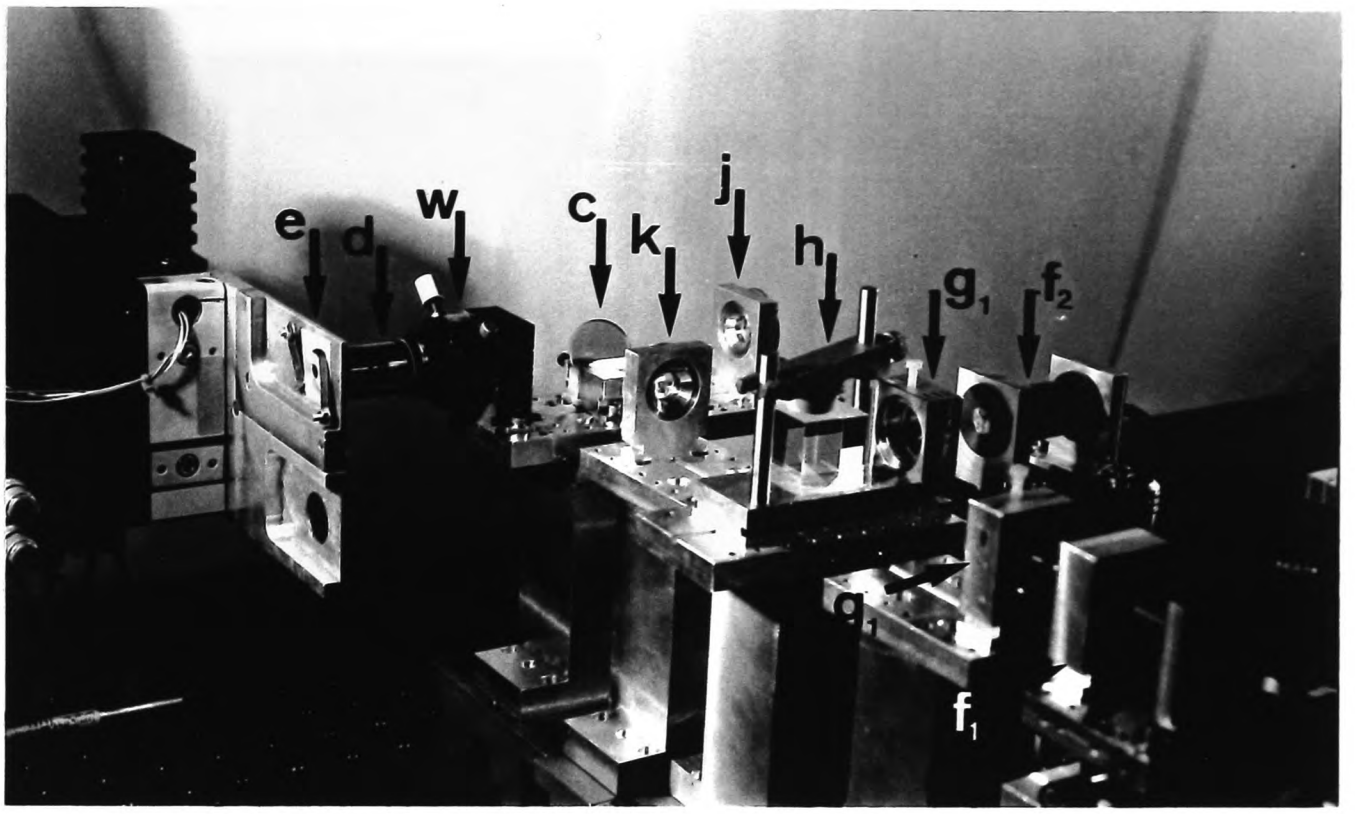


Figure 10.20: The reflection confocal DIC SIRM.

is shown in Fig. 10.20. Since the optical set-up and operation principle was described in great detail earlier (in Chapter 3), Here only the notation is given. The collimated and linearly polarised light of the laser traverses the quarter wave plate (j) and is incident on the Wollaston prism (w). This birefringent element separates the circularly polarised light into two, orthogonally polarised beams and with the objective lens (d) produces two laterally displaced probes. The scattered/reflected light is collected by the objective lens and recombined by the Wollaston prism. The light is then split by the non-polarising beam splitter (c) and is incident upon a half wave plate (k). The light after the wave plate is linearly polarised and incident on a polarising cube beam splitter (h). This acts as an analyser and splits the amplitude signal towards one detector arm and transmits the phase signal towards the other detector arm. Light is then focussed by the detector lenses (g_1 and g_2) to the pin-hole/detector complexes (f_1 and f_2). When the amplitude signal corresponds to the crossed polariser/analyser direction, then no specular reflection detection occurs. This provides an opportunity to examine specimens close to the front surface. It is also advantageous that in this operation mode amplitude detection occurs instead of intensity detection. Since the intensity signal is given by squaring the amplitude signal, when particle detection occurs the size range that can be covered by amplitude

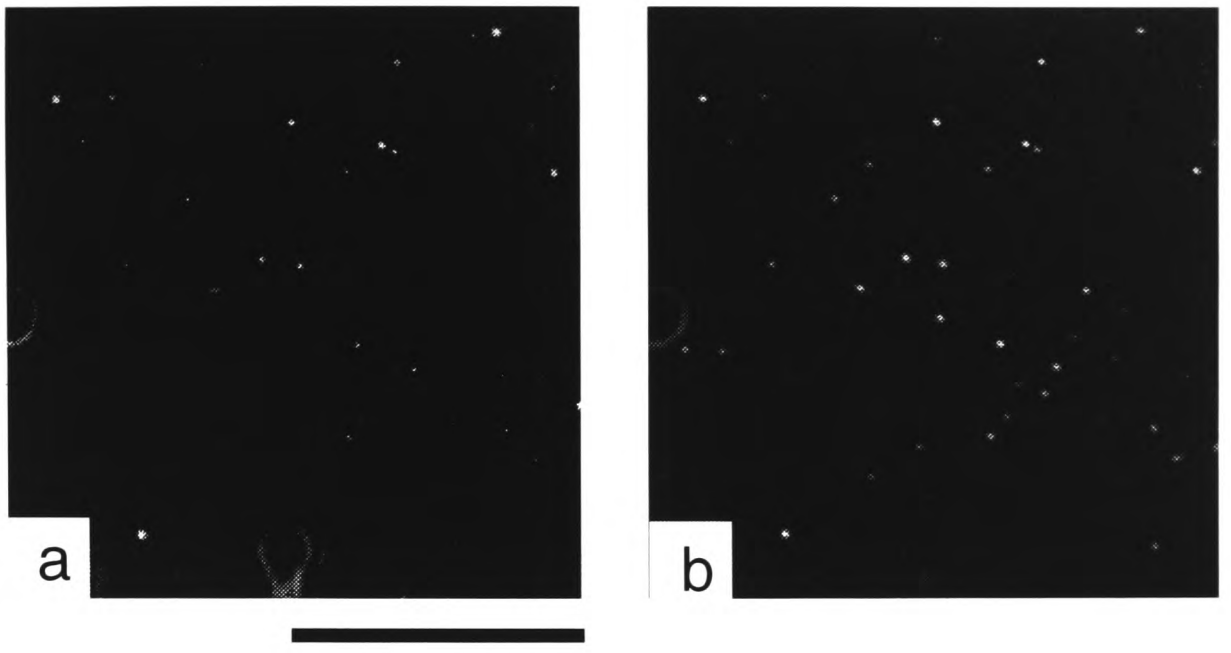


Figure 10.21: A pair of plan-view images from specimen BM-5. Figure (a) corresponds to a reflection confocal image and Figure (b) to a reflection confocal Nomarski DIC image. Images are taken at 30 μm depth. The scale in lateral direction corresponds to 80 μm.

detection is much wider than that covered by intensity detection.

Images taken with reflection confocal and reflection confocal DIC modes are shown in Fig. 10.21. The scale corresponds to 80 μm and in figure (a) the conventional reflection confocal image is shown and in (b) the confocal reflection DIC image is displayed from the same area. The images were recorded at 30 μm below the front surface of the specimen and when image (a) was recorded a substantial offset was subtracted from the image to avoid overload of the amplifier. At the same time when image (b) was recorded, no offset was subtracted nor subsequent image analysis was performed. Image (b) reveals some particles that are not shown in image (a), but most importantly the real difference between these imaging mode is that the specimen can be examined at extremely shallow depths.

For summary purposes in Table 10.3 we show for all the microscope modes the experimental resolution values [μm] and associated effective probe volumes [μm^3], the deduced maximum detectable particle number densities (PND) [cm^{-3}] and the depth values [μm] at which the shallowest particle imaging can occur with respect to the specimen front surface.

Mode	lat. res. [μm]	depth res. [μm]	probe volume [μm^3]	Max. PND [cm^{-3}]	Min. depth [μm]
RC	0.7	5-7	5.5	1.8×10^{11}	40 - 80
TNC	1-2	30	23.6	4.2×10^{10}	N/A
TC	0.7	7	5.5	1.5×10^{11}	N/A
TC-DPC	0.7	7	5.5	1.5×10^{11}	N/A
RC-DPC	0.7	5-7	5.5	1.8×10^{11}	~ 0
RC-DIC	0.7	5-7	5.5	1.8×10^{11}	~ 0

Table 10.3: Summary of different imaging modes in SIRM. Notations: RC - reflection confocal, TNC - transmission non-confocal, TC - transmission confocal, TC-DPC - transmission confocal differential phase contrast, RC-DPC - reflection confocal differential phase contrast, RC-DIC - reflection confocal differential interference contrast, N/A - not applicable

10.5 Transmission confocal SIRM

For the transmission confocal SIRM we measured the depth resolution dependence on the pin-hole size. Two optical set-ups were used, one with a probe-forming lens and a detector lens NA of 0.4 and 0.09 respectively, and the other with 0.85 and 0.05 respectively. The probe-forming lenses were infinite tube length corrected lenses designed for visible light and with no spherical aberration correction collar. The collector lenses were chosen to be the same as the corresponding probe-forming lenses. Particles were imaged in a silicon specimen heat treated to produce oxide particles with medium number density ($\sim 10^8 cm^{-3}$) and with medium size. Images were obtained in the middle of the specimen. Theoretical calculations were performed to compare our experimental results to those obtained previously by Wilson[113] and the resulting curves are plotted in Figs. 10.22 and 10.23. From these figures it can be concluded, that when low NA probe-forming and collector lenses were used, an excellent agreement is found between experimental and theoretical results. Experimental depth resolution values at pin-hole radii larger than 25 μm are constant because the radius of the detector used in these experiments was 25 μm and this then limits the effective pin-hole size. In Fig. 10.23 a less excellent agreement is found, i.e. the theoretical depth resolution values are better than those obtained from experiment. The most likely reason for this difference between the two curves is that due

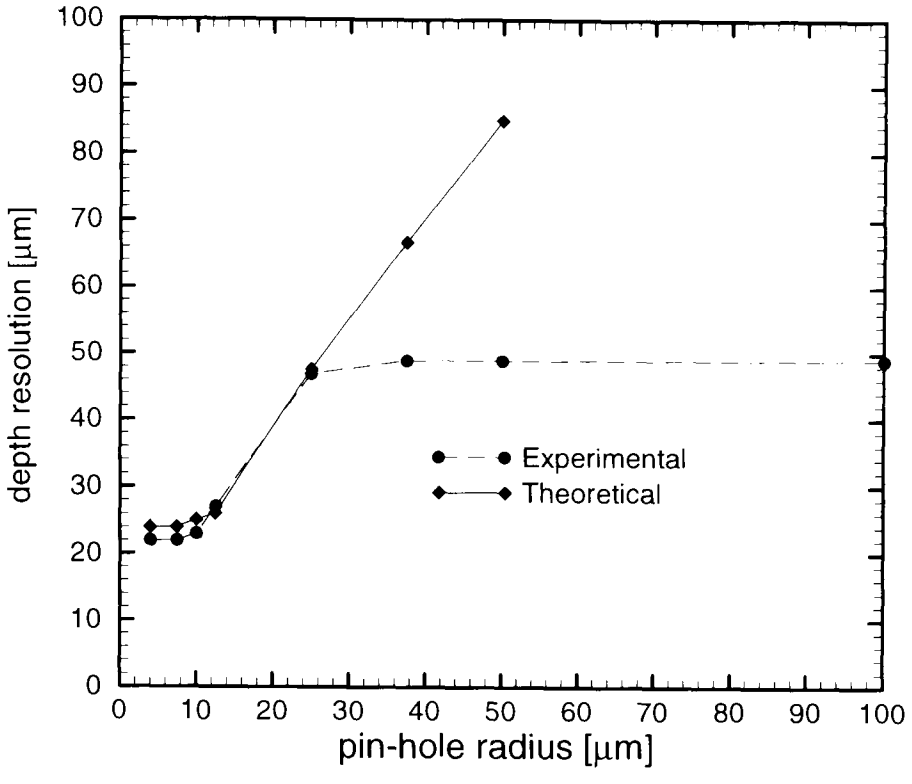


Figure 10.22: The depth resolution of the transmission confocal SIRM as the function of pin-hole size. The NAs of probe-forming and detector lenses are 0.4 and 0.09 respectively. Note that the 'saturation' effect on the experimental curve is caused by a limited detector size.

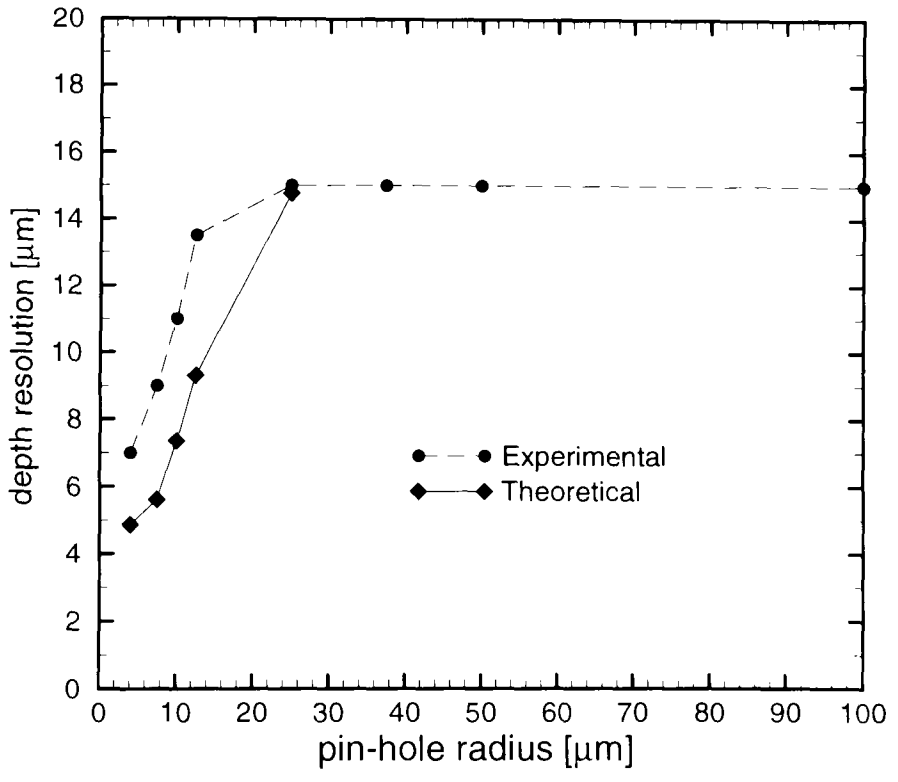


Figure 10.23: The depth resolution of the transmission confocal SIRM as the function of pin-hole size. The NAs of probe-forming and detector lenses are 0.85 and 0.05 respectively. Note that the 'saturation' effect on the experimental curve is caused by a limited detector size.

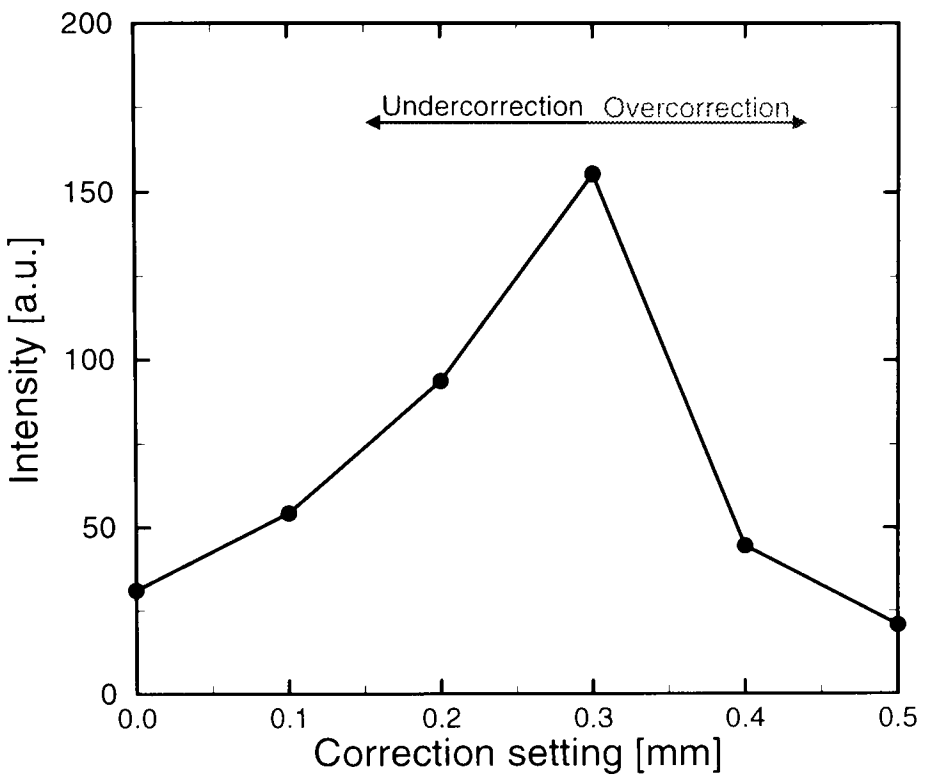


Figure 10.24: The effect of spherical aberration on the image quality in reflection confocal SIRM.

to the high NA probe-forming and collector lenses (NA=0.85), the spherical aberration due to the specimen worsens the depth resolution of the SIRM, while the theoretical curve was calculated for an aberration free case.

10.6 Interpretation of results

In this chapter the experimental aspects of the SIRM are examined. It was shown that the spherical aberration strongly effects the image quality and the maximum detectable particle number density. In particular, the presence of either positive or negative spherical aberration causes a substantial decrease in intensity of the main peak. By further examination of images shown in Fig. 10.13 it is found that the SIRM is more sensitive for spherical aberration overcorrection than for undercorrection. If the peak intensities from individual images of Fig. 10.13 are plotted against the correction, then we obtain Fig. 10.24. It is shown in the figure that when the system is undercorrected (i.e. less correction is applied than the optimum) the Strehl intensity is much less affected compared with that when higher corrections are applied (overcorrection). It is also shown in Fig. 10.13 that the axial resolution is worsened by spherical aberration when the correction setting

is different from the depth at which the particle is being imaged. This effect, for the examined range, can cause a depth resolution $\sim 3\times$ worse than the optimum resolution. These findings are in full agreement with the experimental results of Keller[51] who studied experimentally the effect of spherical aberrations when different cover-glass thicknesses were used with a lens compensated for a particular nominal thickness.

We also examined the possibility of near-surface imaging of particles with the reflection confocal (RC) SIRM. It was found that when the conventional RC set-up was applied, the surface reflection prevents the imaging of near-surface particles (in the 0 to 70 μm region). There are, however, some methods with which this problem can be overcome. The RC DPC or RC DIC methods can image these near-surface particles without any surface reflections as demonstrated in Fig. 10.21.

An important conclusion can be drawn from the measured contrast data. The contrast for the oxide particles is substantially different for the transmission non-confocal and the reflection confocal cases as shown by Table 10.3.5. In particular, in the reflection contrast a sudden jump occurs on going from specimen *BM-3* to *BM-4*. This jump was not observed by transmission SIRM which suggests that the contrast mechanisms of the two imaging modes are caused by different effects. In the reflection mode the contrast arises from scattering while in the transmission mode the contrast is likely due to a combined effect of scattering, absorption and blocking. Transmission electron microscopy examinations[73] of specimens described in Section 10.3.5 revealed that on going from specimen *BM-3* to *BM-4* a new group of particles appeared as the result of the longer heat treatments. The shape of these particles was platelet and their plane was often perpendicular to the incident beam direction. The presence of this new particle group could well be an explanation for the sudden jump in the reflection contrast. The change in orientation of the platelets from edge-on to face-on would significantly increase the equivalent particle size for scattering. Detailed RC SIRM examination of specimens *BM-2* to *BM-6* showed that while up to specimen *BM-5* individual particles were imaged, at specimen *BM-6* cluster formation occurred (*c.f.* Figs. 10.9 and 10.10). Although the spacing between some particles in the cluster was large enough to resolve them by RC SIRM as individual particles, some of the

particles were spaced closer than the resolution limit and these particles were detected as large but single particles.

The reflection confocal contrast obtained from the copper colonies (Fig. 10.12 - specimen *BM-0*) was comparable to the contrast obtained from the oxide particles of specimen *BM-2*. Transmission non-confocal SIRM examinations show that the transmission contrast from the same *BM-0* specimen on a similar copper colony was 1 to 1.3% compared with the transmission contrast from the oxide particles of specimen *BM-2* which was below noise level. This huge *relative* difference between transmission and reflection contrasts suggests again (together with theoretical considerations) that the contrast mechanism for the transmission and for the reflection modes is not the same. It seems likely that for the copper colonies in the transmission mode, absorption or possibly significant scattering in other than the forward or backward directions are mainly responsible for the contrast. This argument is supported by the fact that the refractive index of copper for a wavelength of 1.3 μm is not purely real, but has an imaginary part, and this means that copper is absorbent for this wavelength.

It can be concluded, that the reflection confocal SIRM is a non-destructive tool to measure particle number density and distribution in bulk silicon materials if the examination depth is large enough that surface reflection is not detected. It is shown that the depth resolution and therefore the maximum detectable particle number density are strongly influenced by the presence of spherical aberration. When the microscope is corrected for spherical aberration caused by the specimen, the maximum particle number density that can be detected is $\sim 10^{11}\text{cm}^{-3}$. When specimen regions close to the front surface are to be detected, the use of reflection confocal DPC or DIC SIRM could be more advantageous.

CHAPTER 11

Summary, suggestions for further work and conclusions

"Were I to await perfection, my book would never be finished."
Tai T'ung, 13th Century

Instrumentation

Based on the experience of the last ten years at the Department of Materials, University of Oxford a new SIRM was designed and built. It incorporates new imaging modes, including reflection and transmission confocal, reflection and transmission confocal differential phase contrast (DPC), and reflection confocal differential interference contrast (DIC). For the new imaging modes the SIRM was substantially re-designed mainly involving the electronic, optical and fine mechanical components. New electronics were designed and constructed for a piezo translator for fine positioning/scanning, lateral scanners to provide better mechanical movements, and a signal amplifier for differential detection. New mechanical elements were incorporated for better mechanical stability and versatility. The design involved the development of the optical set-ups for all of the new imaging modes and optimisation of the microscope for each. Confocal SIRM images were recorded to demonstrate the imaging capabilities of the new modes. By using optical signal processing additional information was obtained from oxide precipitate particles in silicon wafers, and surface reflections were eliminated without loss of resolution or light intensity. Specimens could be routinely examined without the need for microscope realignment after replacing the specimen.

In the future, the most promising imaging modes of the reflection SIRM, namely DPC and DIC, should be fully optimised and experimental work performed to understand better the type of information that can be obtained. The overall performance of the

reflection SIRM imaging modes could be improved by using a high power laser and a high performance cooled detector. It should be possible to deduce from the measurements the size, shape, orientation and material of the particles being imaged. This could be done for individual particles by making an angle dependent detection of the scattered light within a large solid angle or in the transmission DIC mode by measuring both the phase and amplitude of the light.

Theory

A new and rigorous theory was developed for high numerical aperture focussing of vector waves through a dielectric interface between mismatched refractive index materials so as to model the illuminating electromagnetic field, and particularly for the SIRM when its probe is focussed into silicon. The mathematical form of this theory although rigorous is simple and it can easily be used for numerical computation purposes. When the lens produces a perfect on-axis spherical wave, as with the present SIRM, the only aberration in the system is spherical aberration produced by the silicon specimen and this can be described by a closed analytical function. The theory gave back for special cases, e.g. high aperture focussing into a single medium or paraxial solution, the equations previously obtained by other workers. An analytical solution of the resulting diffraction integrals was obtained and a new method was proposed so that the solution could easily be computed numerically. These results agreed exactly with the numerical results previously obtained by others, e.g. for the case of no spherical aberration. Numerical data were also obtained, for the first time, for a focussing type optical system with high spherical aberration. These results showed a new behaviour of the focus shift and the Strehl intensity. Comprehensive results have been obtained (not included in this thesis) showing the effect of both spherical aberration and numerical aperture on the focus shift, Strehl intensity, and axial and lateral resolutions. The aberration function of the newly developed theory was related to the better known Zernike coefficients, numerical examples were given for the case of the air/silicon interface, and pictorial representations of Zernike's circular polynomials were presented.

The classical scattering theory (developed previously by others) was adapted to the case of the SIRM, a focussed and aberrated illumination, and a spherical oxide particle in silicon, and a formal solution was given for the back-scattered light corresponding to the confocal reflection SIRM. Based on numerical scattering calculations a contrast mechanism was proposed from which particle sizes could be deduced once the reflection confocal contrast was measured. These SIRM experimental data were compared to TEM observations and good agreement was found. Comparison with the corresponding non-confocal contrast data showed significantly different behaviours, especially for oxide particles with associated copper colonies, indicating that the reflection confocal and transmission non-confocal contrast mechanisms were significantly different. An explanation for the effect was given.

In the future, the primary aim should be to obtain the analytical solution for the backward propagation of light scattered by spherical oxide particles and so completing the quantitative model of the reflection confocal SIRM. With this model a better understanding of SIRM images should be possible. This model could also be applied to disciplines other than material science, e.g. biological specimens. For the description of oxide particles as scatterers, a more fully developed model is required incorporating platelet or other particle shapes, the strain fields caused by adjacent stresses in the crystal, and the occurrence of closely spaced particles. It seems likely that the operator method, proposed by other workers for confocal microscopy and already introduced to high numerical aperture vectorial focussing in this work, could simplify the mathematics for the analytical solution of the backward propagation.

From the practical point of view, the following conclusions are listed.

1. When imaging oxide particles in silicon wafers with the SIRM, the reflection mode has the advantage over the transmission mode and 90° scattering modes that the wafer does not have to be polished on the back surface or cleaved and polished, and so it is a truly non-destructive mode.

2. In order for the reflection SIRM to have high performance, it must be operated with

a high numerical aperture (NA) lens and in the confocal mode.

3. The combination of high NA and focussing deeply into a silicon wafer with high refractive index (3.5) produces high spherical aberration.

4. For a wavelength of 1.3 μm and NA=0.85, the best resolutions occur when there is no spherical aberration, the lateral resolution then being $\sim$ 0.7 μm and the depth resolution then being $\sim$ 5 μm.

5. These resolutions can be achieved in practice for any particular depth down to $\sim$ 500 μm by using a lens with a spherical aberration compensator adjustment collar and setting it to the particular depth value.

6. When focussing closer to the surface than 50 μm, light reflected from the surface 'swamps' the light scattered from the particles and degrades the detectability of the particles.

7. By using the reflection confocal DPC or DIC modes, the surface reflections can be completely eliminated and the above resolutions can be obtained right up to the surface, and so surface denuded zone (SDZ) measurements can be readily made.

8. With these resolutions, particle number densities in the range 10^6 to 10^{11} cm^{-3} can be measured and particle distributions determined.

9. A new rigorous quantitative theory has been described that enables the reflection confocal contrast for individual spherical oxide particles in silicon to be determined as a function of particle size and the predictions are in reasonable agreement with TEM observations. The results show that individual particles down to $\sim$ 90 nm in diameter can be imaged, and smaller particles would probably be imaged if a laser with higher intensity than 20 mW was used.

10. The contrast mechanisms when imaging oxide particles and also oxide particles with associated metal precipitate particles are different for the reflection confocal and

transmission non-confocal modes. For reflection, the contrast arises from light scattering only, while for transmission, the contrast arises from a combination of scattering and absorption. Consequently, the reflection contrast is less complex and is more readily interpretable.

11. The theory described shows that for small particles the reflection contrast varies as d^6 (where d is the particle diameter), and the exponent continually decreases as d increases until for large particles the contrast varies as $d^{1.2}$. The dynamic range is such that it should be possible for particles in the size range 50 to 500 nm in diameter to be simultaneously imaged.

APPENDIX A

Special functions and orthogonal polynomials

In this thesis special functions and orthogonal polynomials are used extensively. This Appendix summarises briefly some of the properties of these functions and polynomials. The relevant references are Abramowitz and Stegun[1], Arfken[3], Born and Wolf[14], Gradshteyn and Ryzhik[33] and Morse and Feshbach[66].

A.1 Bessel and associated functions

In this Section a summary is given on the Bessel, Neumann and Hankel functions.

A.1.1 The Bessel function

Probably one of the most commonly used functions is the Bessel function $J_n(x)$. Its generating function is given by:

$$e^{\frac{x}{2}\frac{t-1}{t}} = \sum_{n=-\infty}^{\infty} J_n(x)t^n, \tag{A.1}$$

and $J_n(x)$ is also defined by the infinite series

$$J_n(x) = \left(\frac{x}{2}\right)^n \sum_{s=0}^{\infty} \frac{(-1)^s}{s!(n+s)!}\left(\frac{x}{2}\right)^{2s}. \tag{A.2}$$

The recurrence relation with which the higher order functions can be generated from lower ones is:

$$J_{n-1}(x) + J_{n+1}(x) = \frac{2n}{x}J_n(x). \tag{A.3}$$

A.1.2 The Neumann function

The Neumann function is important as the second solution of Bessel's equation. Since it is closely related to the Bessel function, only its definition is given:

$$N_\nu(x) = \frac{\cos\nu\pi\ J_\nu(x) - J_{-\nu}}{\sin\nu\pi} \tag{A.4}$$

A.1.3 THE HANKEL FUNCTIONS

The Hankel functions are defined as a combination of the Bessel and Neumann functions:

$$H_\nu^{(1)}(x) = J_\nu(x) + \mathrm{i}N_\nu(x) \tag{A.5a}$$

$$H_\nu^{(2)}(x) = J_\nu(x) - \mathrm{i}N_\nu(x) \tag{A.5b}$$

For real arguments x, $H_\nu^{(1)}$ and $H_\nu^{(2)}$ are complex conjugates.

Approximations for the above three functions exist both for $x \ll 1$ and $x \gg 1$, but are rarely used due to their complexity.

A.2 The spherical Bessel and associated functions

In this Section the spherical Bessel, Neumann and Hankel functions are described. These functions are more frequently used in physics than those of the previous Section.

A.2.1 THE SPHERICAL BESSEL FUNCTION

The spherical Bessel function is defined in terms of the Bessel function as:

$$j_n(x) = \sqrt{\frac{\pi}{2x}}\, J_{n+\frac{1}{2}}(x) \tag{A.6a}$$

or as a power series:

$$j_n(x) = 2^n x^n \sum_{s=0}^{\infty} \frac{(-1)^s (s+n)!}{s!(2s+2n+1)!} x^{2s} \tag{A.6b}$$

The limiting cases are when the argument of the function is either very small or very large. For very small arguments ($x \ll 1$) we write:

$$j_n(x) \cong \frac{2^n n!}{(2n+1)!} x^n = \frac{x^n}{(2n+1)!!} \tag{A.7a}$$

or when $x \gg 1$

$$j_n(x) \approx \frac{1}{x} \sin\left(x - \frac{n\pi}{2}\right). \tag{A.7b}$$

A.2.2 THE SPHERICAL NEUMANN FUNCTION

The spherical Neumann function can be defined in terms of the Neumann or the Bessel functions:

$$n_n(x) = \sqrt{\frac{\pi}{2x}}\, N_{n+\frac{1}{2}}(x) = (-1)^{n+1}\sqrt{\frac{\pi}{2x}}\, J_{-n-\frac{1}{2}}(x) \tag{A.8a}$$

or as a power series:

$$n_n(x) = (-1)^{n+1}\frac{2^n\pi^{1/2}}{x^{n+1}}\sum_{s=0}^{\infty}\frac{(-1)^s}{s!(s-n-\frac{1}{2})!}\left(\frac{x}{2}\right)^{2s}. \tag{A.8b}$$

The limiting cases are ($x \ll 1$):

$$n_n(x) \cong -\frac{(2n)!}{2^n n!}\, x^{-n-1} = -(2n-1)!!\, x^{-n-1} \tag{A.9a}$$

or when $x \gg 1$:

$$n_n(x) \approx -\frac{1}{x}\cos\left(x - \frac{n\pi}{2}\right). \tag{A.9b}$$

A.2.3 THE SPHERICAL HANKEL FUNCTIONS

The spherical Hankel functions can either be defined from the corresponding Hankel functions or from the spherical Bessel and Neumann functions. Their definition is given by:

$$h_n^{(1)}(x) = \sqrt{\frac{\pi}{2x}}\, H^{(1)}_{n+\frac{1}{2}}(x) = j_n(x) + \mathrm{i}n_n(x) \tag{A.10a}$$

and

$$h_n^{(2)}(x) = \sqrt{\frac{\pi}{2x}}\, H^{(2)}_{n+\frac{1}{2}}(x) = j_n(x) - \mathrm{i}n_n(x) \tag{A.10b}$$

or the definition of the first spherical Hankel function as a power series:

$$h_n^{(1)}(x) = (-\mathrm{i})^{n+1}\frac{\mathrm{e}^{\mathrm{i}x}}{x}\sum_{s=0}^{n}\frac{\mathrm{i}^s}{s!(2x)^s}\frac{(n+s)!}{(n-s)!} \tag{A.10c}$$

The limiting cases are ($x \gg 1$):

$$h_n^{(1)}(x) \approx (-\mathrm{i})^{n+1}\frac{\mathrm{e}^{\mathrm{i}x}}{x} \tag{A.11a}$$

and

$$h_n^{(2)}(x) \approx \mathrm{i}^{n+1}\frac{\mathrm{e}^{\mathrm{i}x}}{x} \tag{A.11b}$$

It is interesting to note that the asymptotic approximation of $h_n^{(1)}$ represents, on a substitution of $x = kr$, a spherical wave as it diverges from the origin, whilst $h_n^{(2)}$ represents a convergent spherical wave. For scattering purposes, the first approximation is used when the scattered field described is far from the scatterer. This is because of the spherical wave property of the scattered far-field distribution (see Eq. (8.32)).

The recurrence formula that can be applied to any of the three spherical functions is given by:

$$f_{n-1}(x) + f_{n+1}(x) = \frac{2n+1}{x} f_n(x) \tag{A.12a}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}x}\left[x^{n+1} f_n(x)\right] = x^{n+1} f_{n-1}(x) \tag{A.12b}$$

where f_n can correspond to j_n, n_n, $h_n^{(1)}$, $h_n^{(2)}$.

A.2.4 THE RICCATI-BESSEL FUNCTIONS

These functions are widely used in mathematical physics, since they are the solutions of the radial equation (Eq. (8.19a)) that occurs when the Maxwell's equations are solved in spherical polar co-ordinates.

Their definition is given by:

$$\psi_n(x) = \sqrt{\frac{\pi x}{2}}\, J_{n+\frac{1}{2}}(x) = x\, j_n(x) \tag{A.13a}$$

$$\chi_n(x) = -\sqrt{\frac{\pi x}{2}}\, N_{n+\frac{1}{2}}(x) = x\, n_n(x) \tag{A.13b}$$

and

$$\zeta_n^{(1)}(x) = \psi_n(x) - \mathrm{i}\chi_n(x) = \sqrt{\frac{\pi x}{2}}\, H^{(1)}_{n+\frac{1}{2}}(x) = x\, h_n^{(1)}(x) \tag{A.13c}$$

$$\zeta_n^{(2)}(x) = \psi_n(x) + \mathrm{i}\chi_n(x) = \sqrt{\frac{\pi x}{2}}\, H^{(2)}_{n+\frac{1}{2}}(x) = x\, h_n^{(2)}(x) \tag{A.13d}$$

It should be noted that when the solution of the radial equation is considered at or close to the origin (i.e. $x = 0$), the functions χ_n must be omitted from the solution as when the Neumann function is considered.

The asymptotic expansions of the Riccati-Bessel functions are:

$$\psi_n(x) \approx \frac{1}{2}\left[\mathrm{i}^{n+1}\mathrm{e}^{-\mathrm{i}x} + (-\mathrm{i})^{n+1}\mathrm{e}^{\mathrm{i}x}\right] \tag{A.14a}$$

$$\zeta_n^{(1)}(x) \approx (-\mathrm{i})^{n+1}\mathrm{e}^{\mathrm{i}x} \tag{A.14b}$$

$$\zeta_n^{(2)}(x) \approx \mathrm{i}^{n+1}\mathrm{e}^{-\mathrm{i}x} \tag{A.14c}$$

and the first derivatives of the above functions are needed for the theory of scattering:

$$\psi_n'(x) \approx \frac{1}{2}\left[\mathrm{i}^{n}\mathrm{e}^{-\mathrm{i}x} + (-\mathrm{i})^{n}\mathrm{e}^{\mathrm{i}x}\right] \tag{A.14d}$$

$$\left[\zeta_n^{(1)}(x)\right]' \approx (-\mathrm{i})^{n}\mathrm{e}^{\mathrm{i}x} \tag{A.14e}$$

$$\left[\zeta_n^{(2)}(x)\right]' \approx \mathrm{i}^{n}\mathrm{e}^{-\mathrm{i}x} \tag{A.14f}$$

Since the Riccati-Bessel functions are not the most commonly used functions in Fig. A-1, A-2, A-3 and A-4 the functions ψ_n, χ_n, $\left|\zeta_n^{(1)}\right|$ and $\arg\left(\zeta_n^{(1)}\right)$ are plotted for $n = 1 \ldots 9$ at the end of the Appendix.

A.3 The Legendre and associated functions

The Legendre functions are the solutions of Legendre's equation and they are frequently used to expand functions in terms of Legendre polynomials.

A.3.1 THE LEGENDRE FUNCTIONS

The definition of the Legendre functions is given by:

$$P_n(x) = \sum_{k=0}^{[n/2]} (-1)^k \frac{(2n-2k)!}{2^n k!(n-k)!(n-2k)!} x^{n-2k} \tag{A.15a}$$

or by the Rodrigues' formula:

$$P_n(x) = \frac{1}{2^n n!}\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^n \left(x^2 - 1\right)^n \tag{A.15b}$$

The generation function is given by the formula:

$$\left(1 - 2xt + t^2\right)^{-1/2} = \sum_{n=0}^{\infty} P_n(x)t^n |t| < 1 \tag{A.16}$$

The polynomials can be computed from the recurrence:

$$P_{n+1}(x) = 2xP_n(x) - P_{n-1}(x) - \frac{xP_n(x) - P_{n-1}(x)}{n+1} \tag{A.17}$$

A.3.2 THE ASSOCIATED LEGENDRE FUNCTIONS

The associated Legendre functions are the solutions of the associated Legendre equation (Eq. (8.19b)). Their definition is given in terms of the Legendre functions by:

$$P_n^{(m)}(x) = \left(1 - x^2\right)^{m/2} \frac{\mathrm{d}^m}{\mathrm{d}x^m} P_n(x) \tag{A.18a}$$

or

$$P_n^{(m)}(\cos\theta) = (\sin\theta)^{m/2} \frac{\mathrm{d}^m}{\mathrm{d}(\cos\theta)^m} P_n(\cos\theta). \tag{A.18b}$$

The generation function is given by the formula:

$$\frac{(2m)!\,(1 - x^2)^{m/2}}{2^m m!\,(1 - 2xt + t^2)^{m+\frac{1}{2}}} = \sum_{s=0}^{\infty} P_{s+m}^{(m)}(x)t^s \tag{A.19}$$

The polynomials can be computed from the recurrence:

$$(2n+1)xP_n^{(m)} = (n+m)P_{n-1}^{(m)} + (n-m+1)P_{n+1}^{(m)} \tag{A.20}$$

A.3.3 THE τ_n AND π_n FUNCTIONS

The angle dependent functions of the scattering are defined as:

$$\pi_n(\cos\theta) = \frac{P_n^{(1)}(\cos\theta)}{\sin\theta}, \qquad \tau_n(\cos\theta) = \frac{\mathrm{d}}{\mathrm{d}(\theta)} P_n^{(1)}(\cos\theta) \tag{A.21}$$

The recurrence relation is

$$\pi_n = \left[\frac{2n-1}{n-1}\cos\theta\right]\pi_n - \frac{n}{n-1}\pi_{n-2} \tag{A.22a}$$

$$\tau_n = (n\cos\theta)\pi_n - (n+1)\pi_{n-1} \tag{A.22b}$$

which is very stable as also follows from Eq. (A.18).

A.4 The Gegenbauer polynomials

The Gegenbauer polynomials are rare functions in mathematical physics. Their application in this work arises from the analytical solution of the diffraction integrals obtained in Chapter 5.

The generating function is defined as:

$$\left(1-2xt+t^2\right)^{-\left(\alpha+\frac{1}{2}\right)} = \sum_{n=0}^{\infty} C_n^{\left(\alpha+\frac{1}{2}\right)}(x)t^n \tag{A.23}$$

The definition of the Gegenbauer polynomials is:

$$C_n^{\left(\alpha+\frac{1}{2}\right)}(x) = \frac{\alpha!(2\alpha+n)!}{(-2)^n n!(2\alpha)!(\alpha+n)!}\frac{1}{(1-x^2)^\alpha}\frac{\mathrm{d}^n\left(1-x^2\right)^{\alpha+n}}{\mathrm{d}x^n}. \tag{A.24}$$

These polynomials can be calculated from the stable recurrence relation:

$$(n+1)C_{n+1}^{\left(\alpha+\frac{1}{2}\right)}(x) = (2n+2\alpha+1)xC_n^{\left(\alpha+\frac{1}{2}\right)}(x) - (n+2\alpha)C_{n-1}^{\left(\alpha+\frac{1}{2}\right)}(x). \tag{A.25}$$

As it is shown by the definition Eq. (A.24) and on comparison with Eq. (A.15b)

$$C_n^{(1/2)}(x) = P_n(x) \tag{A.26}$$

A.5 Orthogonal polynomials

The expansion of functions in mathematical physics is a frequently used tool to approximate functions. Any well behaved, i.e. at least piecewise continuous, function $F(x)$ can be approximated to any accuracy by a series of complete orthogonal polynomials*:

$$F(x) = \sum_{n=0}^{\infty} a_n\varphi_n(x). \tag{A.27}$$

The set of polynomials $\varphi_n(x)$ is called *complete* on an interval $[a,b]$ if the mean square error vanishes:

$$\lim_{m\to\infty}\int_a^b\left[F(x)-\sum_{n=0}^{\infty}a_n\varphi_n(x)\right]^2 w(x)\mathrm{d}x = 0 \tag{A.28}$$

*More stringent, sufficient but not necessary, conditions for the convergence of the expansion to the function $F(x)$ are the Dirichlet conditions.

where $w(x)$ is the weighting function. A system of polynomials is called *orthogonal* in $[a, b]$ if:

$$\int_a^b w(x)\varphi_n(x)\varphi_m(x)\mathrm{d}x = \delta_{nm} \tag{A.29}$$

where δ_{nm} is the Kronecker symbol. A complete set of orthogonal functions is said to be *orthonormal* if:

$$\int_a^b w(x)\varphi_n^2(x)\mathrm{d}x = h_n \tag{A.30}$$

where h_n is the normalisation factor. The coefficients a_m can be computed as:

$$a_m = \int_a^b F(x)\varphi_m(x)w(x)\mathrm{d}x \tag{A.31}$$

which follows from Eq. (A.27) on multiplying it by $\varphi_m(x)w(x)$ and integrating, and also from the definition of orthogonality.

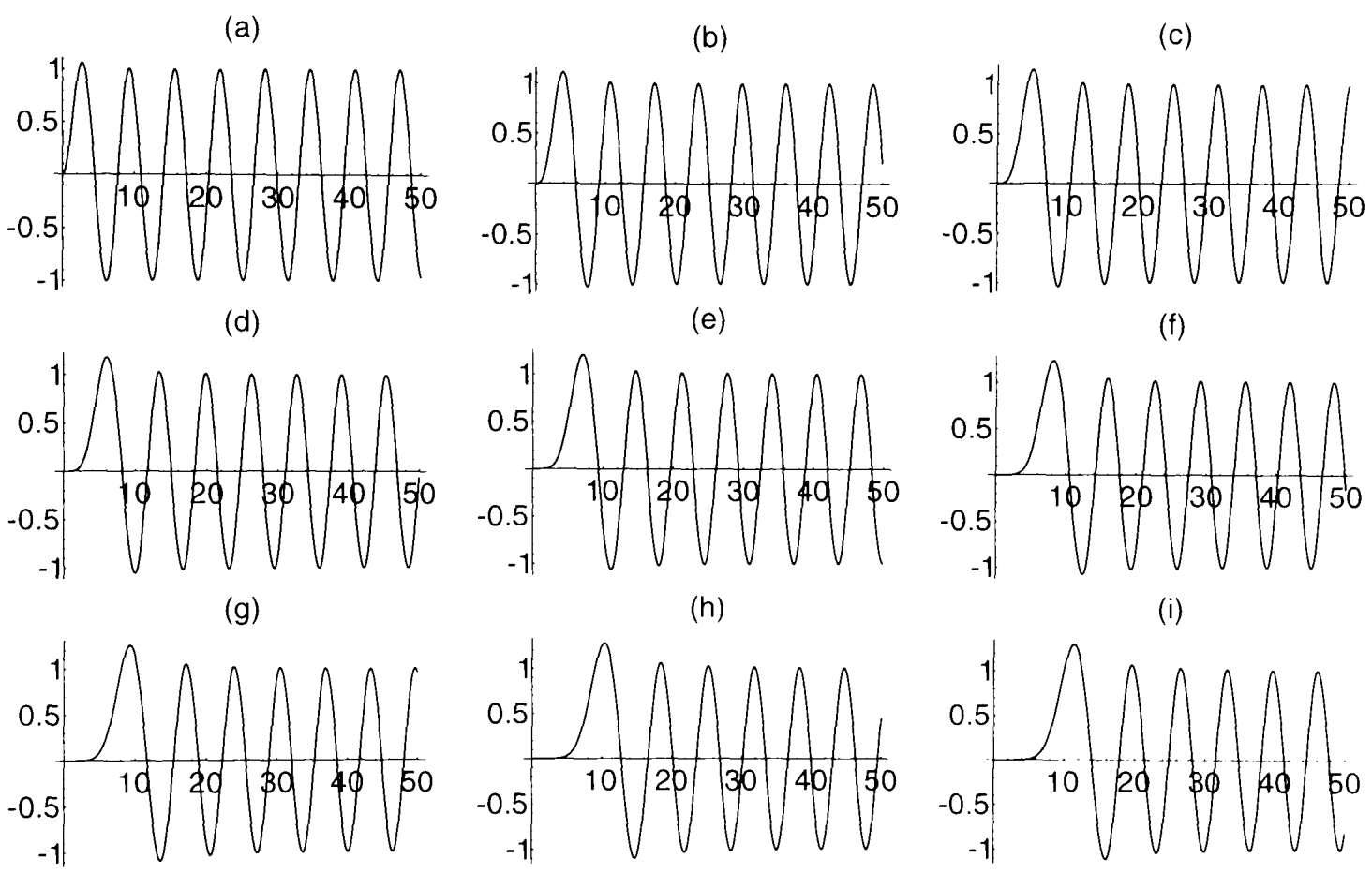


Figure A-1: Functions ψ_n. Lettering (a)...(i) corresponds to $n = 1 \ldots 9$.

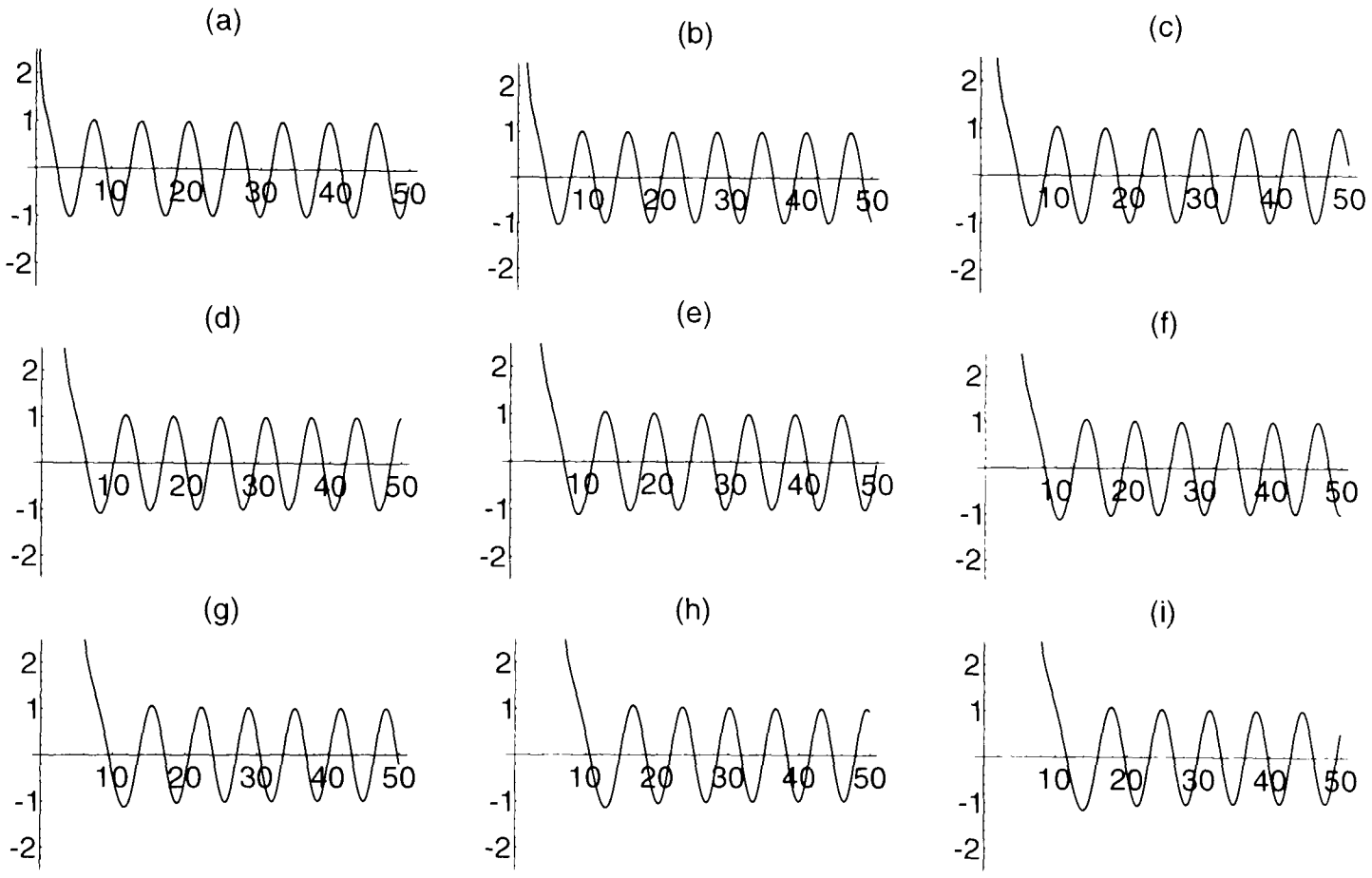


Figure A-2: Functions χ_n. Lettering (a)...(i) corresponds to $n = 1 \ldots 9$.

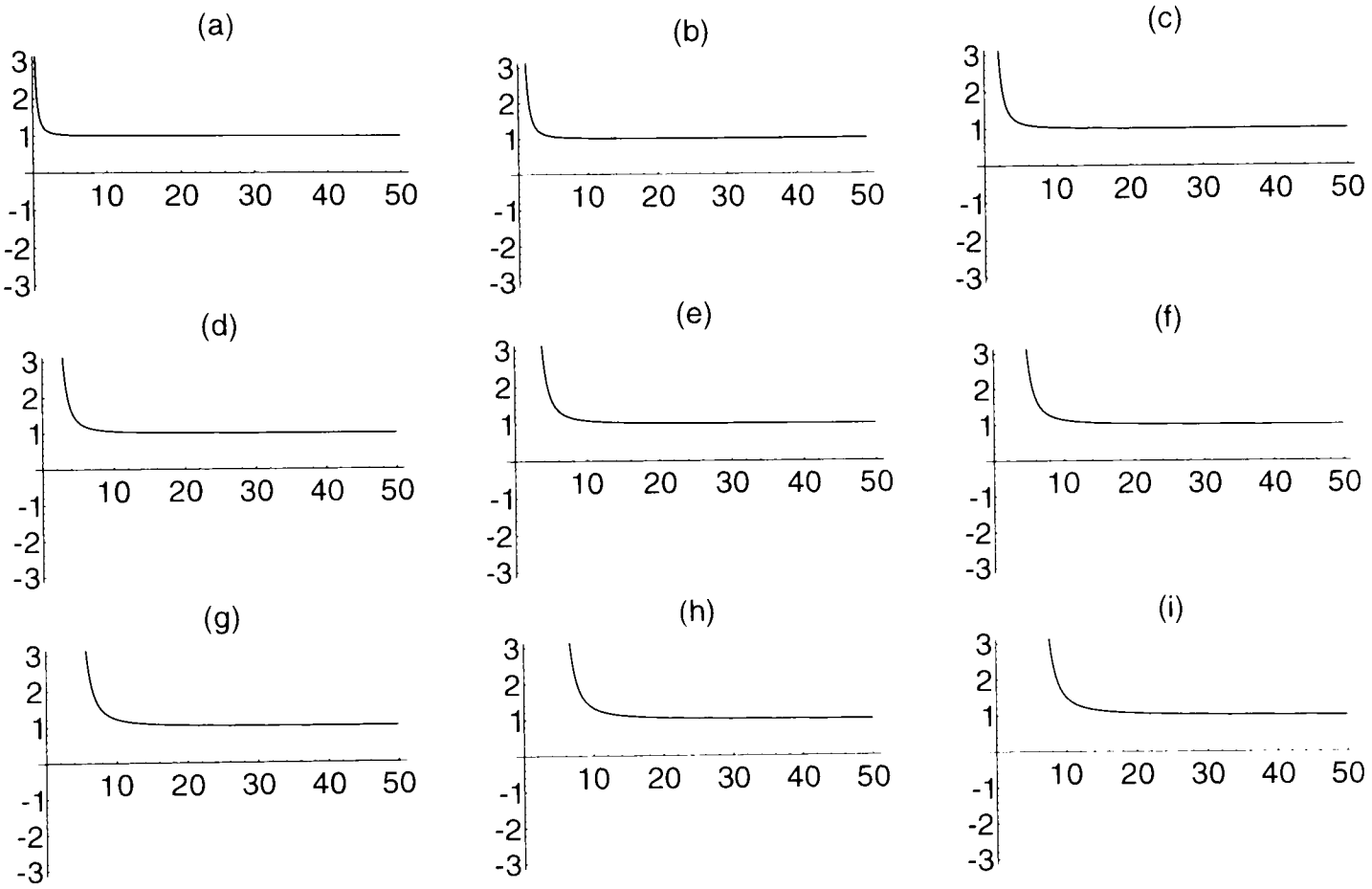


Figure A-3: Functions $\left|\zeta_n^{(1)}\right|$. Lettering (a)...(i) corresponds to $n = 1 \ldots 9$.

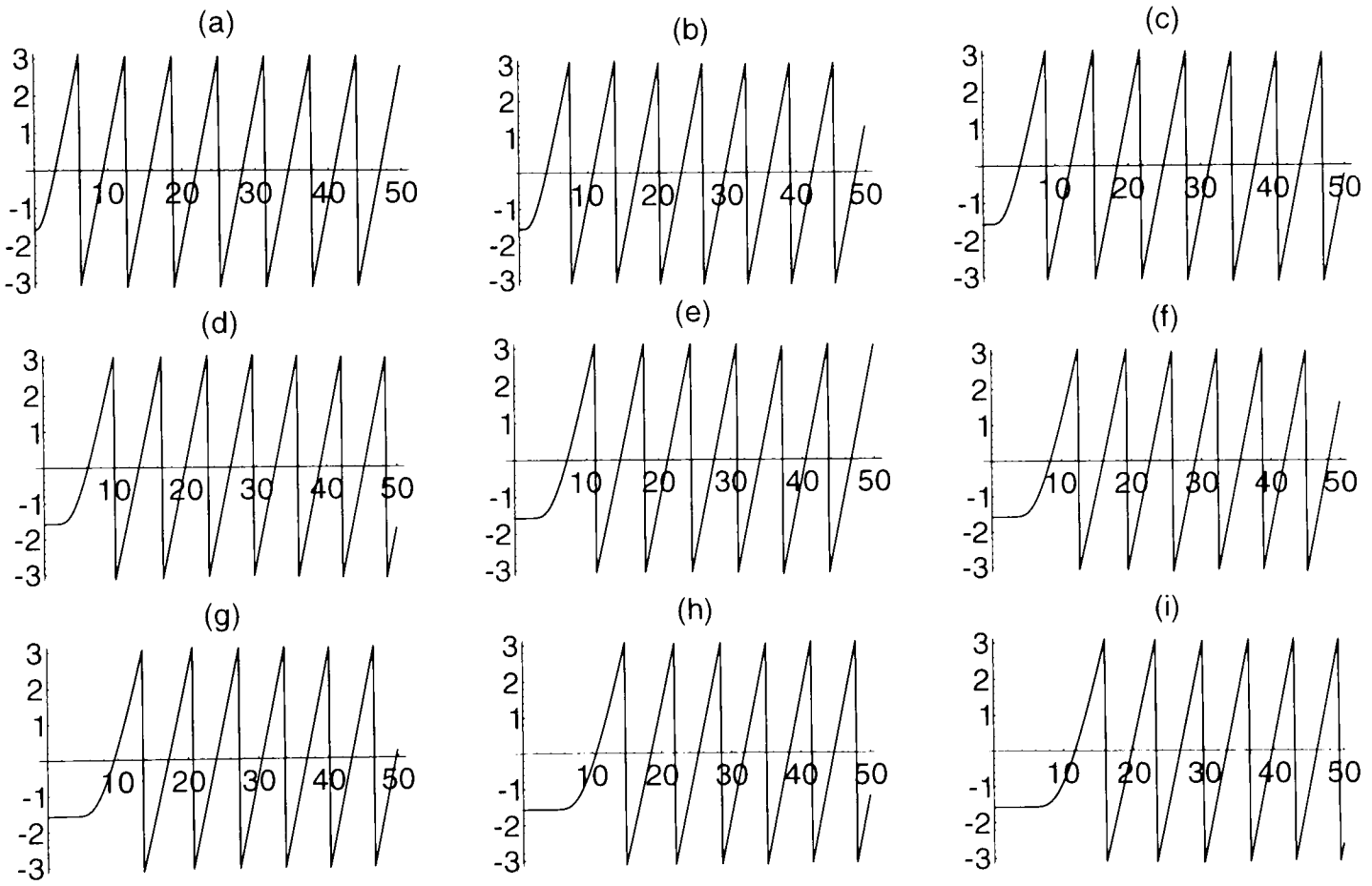


Figure A-4: Functions $\arg\left(\zeta_n^{(1)}\right)$. Lettering (a)...(i) corresponds to $n = 1 \ldots 9$.

APPENDIX B

Proof of two solutions

This Appendix presents proof for the solutions of Bessel's differential equation and that of the associated Legendre equation.

B.1 The solution of Bessel's differential equation

Consider the recurrence relation of Eq. (A.3) which also can be written in a different form:

$$J_{n-1}(x) - J_{n+1}(x) = 2J'_n(x). \tag{B.1}$$

From Eqs. (A.3) and (B.1) we can write:

$$J_{n-1}(x) = \frac{n}{x}J_n(x) + J'_n(x). \tag{B.2}$$

If we now consider any set of any functions $(Z_n(x))$that satisfy the recurrence relations Eqs. (A.3) and (B.1) and by letting $\nu \to n$:

$$xZ'_\nu(x) = xZ_{\nu-1}(x) - \nu Z_\nu(x) \tag{B.3}$$

Differentiating with respect to x we obtain:

$$xZ''_\nu(x) + (\nu+1)Z'_\nu(x) - xZ'_{\nu-1}(x) - Z_{\nu-1}(x) = 0$$

Multiplying this equation by x and subtracting it from Eq. (B.3) yields:

$$x^2Z''_\nu(x) + xZ'_\nu(x) - \nu^2 Z_\nu(x) + (\nu-1)xZ_{\nu-1}(x) - x^2Z'_{\nu-1}(x) = 0 \tag{B.4}$$

Now Eq. (B.2) gives on substituting by $n = \nu - 1$:

$$xZ'_{\nu-1}(x) = (\nu-1)Z_{\nu-1}(x) - xZ_\nu(x) \tag{B.5}$$

This equation can be used the eliminate $Z_{\nu-1}$ from Eq. (B.4) and this gives:

$$x^2 Z_\nu''(x) + x Z_\nu'(x) + \left(x^2 - r^2\right) Z_\nu(x) = 0 \tag{B.6}$$

On dividing the above equation by x^2 and substituting $\nu = l + 1/2$ we obtain:

$$Z''_{l+\frac{1}{2}}(x) + \frac{1}{x} Z'_{l+\frac{1}{2}}(x) + \left[1 - \frac{\left(l+\frac{1}{2}\right)^2}{x^2}\right] Z_{l+\frac{1}{2}}(x) = 0 \tag{B.7}$$

which is the essentially of the same form as Eq. (8.19a) and is Bessel's differential equation. Since the present derivation started with the recurrence form, Eq. (A.3), on substituting by $\nu = l + 1/2$ and $f_n = Z_{l+\frac{1}{2}}$ we obtain:

$$f_{n-1}(x) + f_{n+1}(x) = \frac{2n+1}{x} f_n(x)$$

which recurrence formula is identical to Eq. (A.12a). Thus we have proved that either the j_n, n_n, $h_n^{(1)}$, $h_n^{(2)}$ or the Riccati-Bessel functions are solutions of Bessel's differential equation.

B.2 The associated Legendre equation

To prove that the associated Legendre functions are solutions of the associated Legendre equation is slightly more difficult, but essentially involves the same principles. We start from the recurrence relation for the Legendre polynomials (Eq. (A.17)). Differentiation of the generation function gives:

$$P'_{n+1}(x) + P'_{n-1}(x) = 2x P'_n(x) + P_n(x). \tag{B.8}$$

On differentiating the recurrence relation Eq. (A.17) and multiplying Eq. (B.8) by $(2n+1)$ and adding the two equations, we obtain:

$$P'_{n+1}(x) + P'_{n-1}(x) = (2n+1) P_n(x) \tag{B.9}$$

Now from Eqs. (B.8) and (B.9) we get the following equations:

$$P'_{n-1}(x) = -n P_n(x) + x P'_n(x) \tag{B.10a}$$

$$\left(1 - x^2\right) P'_n(x) = n P_{n-1}(x) - n x P_n(x) \tag{B.10b}$$

and by differentiating Eq. (B.10b) and using Eq. (B.10a):

$$\left(1-x^2\right)P_n''(x)-2xP_n'(x)+n(n+1)P_n(x)=0 \tag{B.10c}$$

With the help of the Leibnitz's formula:

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}\left[A(x)\,B(x)\right]=\sum_{s=0}^{\infty}\binom{n}{s}\frac{\mathrm{d}^{n-s}}{\mathrm{d}x^{n-s}}A(x)\frac{\mathrm{d}^s}{\mathrm{d}x^s}B(x)$$

the m times differentiation of Eq. (B.10c) gives:

$$\left(1-x^2\right)f''-2x(m+1)f'+(n-m)(n+m+1)f=0 \tag{B.11}$$

where

$$f=\frac{\mathrm{d}^m}{\mathrm{d}x^m}P_n(x).$$

By introducing:

$$v(x)=\left(1-x^2\right)^{m/2}f(x)=\left(1-x^2\right)^{m/2}\frac{\mathrm{d}^m}{\mathrm{d}x^m}P_n(x)$$

we obtain:

$$f'=\left(v'+\frac{mxv}{1-x^2}\right)\left(1-x^2\right)^{-m/2} \tag{B.12a}$$

$$f''=\left[v''+\frac{2mxv'}{1-x^2}+\frac{mv}{1-x^2}+\frac{m(m+2)x^2v}{\left(1-x^2\right)^2}\right]\left(1-x^2\right)^{-m/2} \tag{B.12b}$$

On substituting into Eq. (B.11) we obtain:

$$(1-x^2)v''-2xv'+\left[n(n+1)-\frac{m^2}{1-x^2}\right]v=0 \tag{B.13a}$$

which in spherical polar co-ordinate system gives:

$$\frac{1}{\sin\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin\theta\frac{\mathrm{d}v}{\mathrm{d}\theta}\right)+\left[n(n+1)-\frac{m^2}{\sin^2\theta}\right]v=0 \tag{B.13b}$$

for which the solution is:

$$v=\left(1-x^2\right)^{-m/2}\frac{\mathrm{d}^m}{\mathrm{d}x^m}P_n(x)\equiv P_n^{(m)}(x) \tag{B.13c}$$

with which the proof is completed.

REFERENCES

1. M. Abramowitz and I.A. Stegun, *Handbook of mathematical functions*, 1st ed. (Dover, New York, 1965).

2. M. Adachi, K. Yasaka, Appl. Opt. **25**, 764 (1986).

3. G. Arfken, *Mathematical methods for physicists*, 3rd ed. (Academic Press, London, 1985).

4. M.R. Atkinson and A.E. Dixon, Appl. Opt. **33**, 641 (1994).

5. B.B. Baker and E.T. Copson, *The mathematical theory of Huygens' principle*, 2nd ed. (Clarendon Press, Oxford, 1953).

6. P. Barber and C. Yeh, Appl. Opt. **14**, 2864 (1975).

7. J.P.H. Benschop, Proc. SPIE **809**, 90 (1987).

8. S.N. Bezdid'ko, Sov. J. Opt. Technol. **41**, 425 (1975).

9. C.F. Bohren and D.R. Huffman, *Absorption and scattering of light by small particles*, 1st ed. (John Wiley, New York, 1993).

10. A. Boivin and E. Wolf, Phys. Rev. B **138**, 1561-1565 (1965).

11. A. Boivin, J. Dow and E. Wolf, J. Opt. Soc. Am. **57**, 1171-1175 (1967).

12. G.R. Booker, Z. Laczik and R. Falster, European Pattent Application, (1990).

13. G.R. Booker, Z. Laczik and P. Kidd, Semicond. Sci. Technol. **7**, A110 (1992).

14. M. Born and E. Wolf, *Principles of optics*, 4th ed. (Pergamon, Oxford, 1970).

15. G.J. Brakenhoff, P. Blom and P. Barends, J. Microsc. **117**, 219 (1979).

16. G.J. Brakenhoff, J. Microsc. **117**, 233 (1979).

17. G.J. Brakenhoff, J.S. Binnerts and C.L. Woldringh, in *Scanned Image Microscopy*, 183 (1980).

18. M.R. Brozel, I. Grant, R.M. Ware and D.J. Stirland, Appl. Phys. Lett. **42**, 610 (1983).

19. A.R. Carlini and T. Wilson, Proc. SPIE **809**, 97 (1987).

20. X.Z. Cao and A.F. Witt, J. Cryst. Growth **112**, 838 (1991).

21. K. Carlsson, J. Microsc. **157**, 21 (1990).

22. C.J. Cogswell, C.J.R. Sheppard and M.C. Moss, J. Microsc. **158**, 177 (1990).

23. C.J. Cogswell and C.J.R. Sheppard, J. Microsc. **165**, 81 (1992).

24. P. Debye, Ann. Physik (4) **30**, 57 (1909). (referenced in [50])

25. P. Debye, Ann. Phys. **30**, 755 (1909).

26. N.H. Dekkers and H. de Lang, Optik **41**, 452 (1974).

27. M.D. Egger and M. Petrañ, Science **157**, 305 (1967). referred in [82]

28. T.J. Fawcett and M.R. Brozel, J. Cryst. Growth **103**, 78 (1990).

29. J.P. Fillard, J. Cryst. Growth **103**, 71 (1990).

30. J. Gasper, G.C. Sherman and J.J. Stamnes, J. Opt. Soc. Am. **66**, 955 (1976).

31. M. Glass and T. Dabbs, J. Microsc. **164**, 153 (1991).

32. J.W. Goodman, *Introduction to Fourier optics*, 1st ed. (McGraw - Hill, San Francisco, 1968).

33. I.S. Gradshteyn and I.M. Ryzhik, *Tables of integrals, series and products*, 5th ed.

(Academic Press, London, 1994).

34. D.K. Hamilton and T. Wilson, J. Phys. E **19**, 52 (1986).

35. D.K. Hamilton and C.J.R. Sheppard, Optica Acta **29**, 1573 (1982).

36. D.K. Hamilton and T. Wilson, J. Appl. Phys. **53**, 5320 (1982).

37. D.K. Hamilton, C.J.R. Sheppard and T. Wilson, J. Microsc. **135**, 275 (1984).

38. D.K. Hamilton and T. Wilson, J. Opt. Soc. Am. A **1**, 322 (1984).

39. D.K. Hamilton and T. Wilson, Appl. Opt. **23**, 348 (1984).

40. D.K. Hamilton and T. Wilson, IEE Proc. **134**, 85 (1987).

41. D.K. Hamilton and T. Wilson, Micron and Micr. Acta **18**, 77 (1987).

42. A. Hardy and D. Treves, J. Opt. Soc. Am. **63**, 85-90 (1973).

43. S. Hell, G. Reiner, C. Cremer and E.H.K. Stelzer, J. Microsc. **169**, 391 (1993).

44. T. Hellmuth, P. Seidel and A. Seigel, Proc. of SPIE, Scanning Imaging **1028**, 28 (1988).

45. H.H. Hopkins and M.J. Yzuel, Optica Acta **17**, 157-182 (1970).

46. J.D. Jackson, *Classical electrodynamics*, 2nd ed. (John Wiley, New York, 1975).

47. N.Y. Jin, G.R. Booker and I.R. Grant, Mat. Sci. Eng. **B20**, 94 (1993).

48. R. Kant, J. Mod. Opt. **40**, 337-347 (1993).

49. R. Kant, J. Mod. Opt. **40**, 2293 (1993).

50. M. Kerker, *The scattering of light and other electromagnetic radiation*, 1st ed. (Academic Press, London, 1969).

51. H.E. Keller, in *Handbook of Biol. Conf. Microscopy*, 77 (1990).

52. P. Kidd, G.R. Booker and D.J. Stirland, Appl. Phys. Lett. **51**, 1331 (1987).

53. P. Kidd, G.R. Booker and D.J. Stirland, Inst. of Phys. Conf. Ser. **87**. 275 (1987).

54. P. Kidd, D. Phil. Thesis, Oxford (1987).

55. Z. Laczik, G. R. Booker, W. Bergholz and R. Falster, Appl. Phys. Lett. **55** 2625 (1989).

56. Z. Laczik, G. R. Booker, R. Falster and N. Shaw, Inst. of Phys. Conf. Ser. **100** 807 (1989).

57. Z. Laczik, D. Phil. Thesis, Oxford (1992).

58. H. Ling and S-W. Lee, J. Opt. Soc. Am. A **1**, 965 (1984).

59. W. Lukosz, J. Opt. Soc. Am. **56**, 1463 (1966).

60. R.K. Luneburg, *Mathematical theory of optics*, 2nd ed. (University of California Press, Berkeley and Los Angeles, 1966)

61. E.W. Marchand and E. Wolf, J. Opt. Soc. Am. **56**, 1712 (1966).

62. J.C. Maxwell, Phil. Trans. **155**, 459 (1865). (referenced in [50])

63. G. Mie, Ann. Physik **3**, 377 (1908).

64. P.C. Montgomery and J.P. Fillard, Meas. Sci. Technol. **1**, 120 (1990).

65. K. Moriya and T. Ogawa, J. Cryst. Growth **44**, 53 (1978).

66. P.M. Morse and H. Feshbach, *Methods of theoretical physics*, 1st ed. (McGraw-Hill, Tokyo, 1953).

67. K. Nienhuis and B.R.A. Nijboer, Physica **14**, 590 (1949).

68. B.R.A. Nijboer, Physica **10**, 679 (1943).

69. B.R.A. Nijboer, Physica **13**, 605 (1943).

70. M.A. Nokes and P.G. Borden, High Yield Technology Report, (1991).

71. T. Ogawa, J. Cryst. Growth, **88**, 332 (1988).

72. H. Osterberg and L.W. Smith, J. Opt. Soc. Am. **51**, 1050-1054 (1961).

73. B. Pecz, Private communication, Budapest and Oxford (1993).

74. N. Petford and J.A. Miller, J. Microsc. **170**, 201 (1993)

75. B. Richards and E. Wolf, Proc. R. Soc. London Ser. A **253**, 358 (1959).

76. Lord Rayleigh, Nature **3**, 234, 264, 265 (1871). (referenced in [50])

77. O. Seger and R. Lenz, Proc. SPIE, **1660**, 233 (1992).

78. O. Seger, *Model building and restoration with applications in confocal microscopy*, Linköping Studies in Science and Technology, Dissertations No. 301, Linköping (1993).

79. C.J.R. Sheppard and A. Choudhury, Optica Acta **24**, 1051 (1977).

80. C.J.R. Sheppard and T. Wilson Appl. Opt. **18**, 1058 (1979).

81. C.J.R. Sheppard, in *Scanned Image Microscopy*, 201 (1980).

82. C.J.R. Sheppard and T. Wilson, J. Microsc. **124**, 107 (1981).

83. C.J.R. Sheppard and H.J. Matthews, J. Opt. Soc. Am. A **4**, 1354 (1987).

84. C.J.R. Sheppard, Appl. Opt. **27**, 4782 (1988).

85. C.J.R. Sheppard and X.Q. Mao, J. Opt. Soc. Am. A **6**, 1260 (1989).

86. C.J.R. Sheppard and M. Gu, Appl. Opt. **30**, 3563 (1991).

87. C.J.R. Sheppard, M. Gu and M. Roy, J. Microsc. **168**, 209 (1992).

88. C.J.R. Sheppard, M. Gu, K. Brain and H. Zhou, Appl. Opt. **33**, 616 (1994).

89. M.S. Skolnick, Conf. Proc. of DRIP (ed. J.P. Fillard), 165 (Elsevier, Amsterdam, 1985).

90. M.S. Skolnick, M.R. Brozel, L.J. Reed, I. Grant, D.J. Stirland and R.M. Ware, J. Elec. Materials **13**, 107 (1984).

91. G.E. Sommargren, Appl. Opt. **20**, 610 (1981).

92. A. Sommerfeld, *Optics*, 1st ed. (Academic Press, New York, 1954).

93. J.J. Stamnes and G.C. Sherman, J. Opt. Soc. Am. **66**, 780 (1976).

94. J.J. Stamnes and G.C. Sherman, J. Opt. Soc. Am. **68**, 502 (1978).

95. J.J. Stamnes, *Waves in focal regions*, 1st ed. (Adam Hilger, Bristol and Boston, 1986).

96. E.H.K. Stelzer,H.J.B. Marsman and R.W. Wijnaendts van Restandt, Optik **73**, 30 (1986).

97. L. Taijing, K. Toyoda, N. Nango and T. Ogawa, J. Cryst. Growth **108**, 482 (1991).

98. L. Taijing, K. Toyoda, N. Nango and T. Ogawa, J. Cryst. Growth **114**, 64 (1991).

99. M. Vaez Iravani, Electronics Letters **22**, 103 (1986).

100. T.D. Visser and S.H. Wiersma, J. Opt. Soc. Am. A **8**, 1404 (1991).

101. van de Hulst, *Principles of optics*, 4th ed. (Pergamon, Oxford, 1970).

102. H.T.M. van der Voort and G.J. Brakenhoff, J. Microsc. **158**, 43 (1990).

103. J.Y. Wang and D.E. Silva, Appl. Opt. **19**, 1510 (1980).

104. G.N. Watson, *A treatise on the theory of Bessel functions*, 2nd ed. (Cambridge University Press, 1966).

105. R.W. Wijnaendts van Restandt,H.J.B. Marsman,R. Kaplan, J. Davoust, E.H.K. Stelzer and R. Stricker, J. Microsc. **138**, 29 (1985).

106. T. Wilson, W.R. Osicki, J.N. Gannaway, G.R. Booker, J. Mater. Science **14**, 961 (1979).

107. T. Wilson,J.N. Gannaway and P. Johnson, J. Microsc. **118**, 309 (1980).

108. T. Wilson, J.N. Gannaway and C.J.R. Sheppard, in *Scanned Image Microscopy*, 227 (1980).

109. T. Wilson and D.K. Hamilton, J. Microsc. **128**, 139 (1982).

110. T. Wilson and D.K. Hamilton, Optik **71**, 23 (1985).

111. T. Wilson and A.R. Carlini, Appl. Opt. **27**, 3791 (1988).

112. T. Wilson and C.J.R. Sheppard, *Theory and practice of scanning microscopy*, 1st ed. (Academic Press, London, 1989).

113. T. Wilson (ed.), *Confocal microscopy*, 1st ed. (Academic Press, London, 1992).

114. E. Wolf, Rep. Prog. Phys. **14**, 95 (1951).

115. E. Wolf, Proc. R. Soc. London Ser. A **253**, 349 (1959).

116. E. Wolf and Y. Li, Opt. Commun. **39**, 205 (1981).

117. C. Yeh, S. Colak and P. Barber, Appl. Opt. **21**, 4426 (1982).

118. F. Zernike, Physica **1**, 689 (1934).