Peter Török
Fu-Jen Kao
Editors

SPRINGER SERIES IN OPTICAL SCIENCES 87

Optical Imaging and Microscopy

Techniques and Advanced Systems

Second Edition

Springer

Springer Series in
OPTICAL SCIENCES 87

founded by H.K.V. Lotsch

Springer Series in

OPTICAL SCIENCES

The Springer Series in Optical Sciences, under the leadership of Editor-in-Chief *William T. Rhodes*, Georgia Institute of Technology, USA, provides an expanding selection of research monographs in all major areas of optics: lasers and quantum optics, ultrafast phenomena, optical spectroscopy techniques, optoelectronics, quantum information, information optics, applied laser technology, industrial applications, and other topics of contemporary interest.
With this broad coverage of topics, the series is of use to all research scientists and engineers who need up-to-date reference books.

The editors encourage prospective authors to correspond with them in advance of submitting a manuscript. Submission of manuscripts should be made to the Editor-in-Chief or one of the Editors. See also http://www.springer.com/series/624

Peter Török Fu-Jen Kao

Optical Imaging and Microscopy

Techniques and Advanced Systems

2nd Revised Edition

With 328 Figures

Springer

Peter Török

Imperial College London
Department of Physics
Prince Consort Road
London SW7 2BW, UK
Email: peter.torok@imperial.ac.uk

Fu-Jen Kao

National Yang-Ming University
Institute of Biophotonics Engineering
Taipei 112, Taiwan
Email: fjkao@ym.edu.tw

ISSN 0342-4111

ISBN 978-3-540-69563-9 2nd Edition Springer Berlin Heidelberg New York
ISBN 978-3-540-43493-1 1st Edition Springer Berlin Heidelberg New York

Typesetting and production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: eStudio Calamar S. L., F. Steinen-Broo, Girona, Spain

SPIN 11965817 57/3180/YL - 5 4 3 2 1 0 Printed on acid-free paper

Preface to the Second Edition

It has been four years since the publication of the first edition of this book. During these four years optical imaging has advanced considerably. Nevertheless, when we were asked to compile a second edition of this book the first question we asked was: are there any topics that were missed out from the first edition? Even though there are a fairly large number of subjects that we could potentially have considered it was decided to extend the book by two new chapters. The first one is on the use of diffractive optical elements in high quality objective lenses by Drs Brunner and Dobschal from Carl Zeiss AG. It is particularly exciting to have this chapter in the book because it is likely that diffractive elements will play an increasingly important role in high precision optical instrumentation in the future. The second new chapter is by Daniel Axelrod of University of Michigan on total internal reflection fluorescence (TIRF). We believe that many readers will find the chapter from Prof. Axelrod, who developed TIRF about 20 years ago, insightful and very interesting.

In addition to the new chapters some chapters have been revised and considerably expanded. These revisions are primarily there to include the most up-to-date material and recent advances of the field.

Special note of thanks must go to Drs Claus Ascheron and Angela Lahee of Springer-Verlag for their encouragement and help. Most importantly, we would like to thank all the contributors of this book and wish our readers enjoyable and profitable browsing.

London and Taipei,
June 2007

Peter Török
Fu-Jen Kao

Preface

The motivation to collect contributions from a wide variety of subjects in contemporary optics, centered around optical imaging, originates from two ideas. First, it is easy to recognise that certain fields of contemporary optics have been developing in quite a parallel manner. Sometimes workers of the different fields discover each other's contributions, but mostly they do not. One of our major goals is to show how closely these contributions are inter-related. Such an example is the development of scanning/confocal optical microscopy and optical data storage. In the former, imaging in the classical sense, occurs by scanning a tightly focused laser beam over the sample. In optical data storage imaging does not occur as the aim is to *detect* pits, rather than to image them. Nevertheless, the optical systems of these two arrangements have striking resemblance and hence their governing equations are practically the same. The second motivation of this book is to collect contributions from imaging related subjects that were not previously published in this form or they are difficult to access. Such examples are a chapter on white light interferometry, surface plasmon microscopy or the characterisation of high numerical aperture microscope objective lenses.

We are extremely pleased that we have contributions in this book from the international leaders of individual fields. It has been our privilege to work with these authors and we would like to take this opportunity to thank them all.

It now remains to acknowledge those who contribution made the publication of this book possible. First and foremost we wish to thank Miss Janey Lin and Mr Eoin Phillips for their work on the manuscript. They spent endless hours correcting typographical errors, and putting the manuscripts into a format that was suitable for publication. We are also grateful to Mr Sebastian Rahtz who answered all our questions regarding LaTeX; without his help this book would have looked fairly different. Dr Angela Lahee of Springer Verlag was kind enough to offer her help throughout this project that we all but overused. We would like to acknowledge the National Science Council of Taiwan for generous support towards the publication of this book.

The Editors hope that the Reader will derive as much joy from reading the contributions in this book as we did while working on it.

London and Kaohsiung,
January 2003

Peter Török
Fu-Jen Kao

Contents

List of Contributors

M.R. Arnison
University of Sydney
Department of Physical Optics
School of Physics A28
NSW 2006
Australia
mra@physics.usyd.edu.au

D. Axelrod
University of Michigan
Ann Arbor, MI 48109
USA
daxelrod@umich.edu

J.J.M. Braat
Optics Research Group
Faculty of Applied Sciences
Delft University of Technology
Lorentzweg 1
NL 2627 CA Delft
The Netherlands
j.j.m.braat@tnw.tudelft.nll

G.J. Brakenhoff
Biophysics and Microscopy Group
Swammerdam Institute for Life Sciences
University of Amsterdam
P.O. Box 94062, 1090 GB Amsterdam
The Netherlands
brakenhoff@science.uva.nl

R. Brunner
Carl Zeiss AG, Central Research and Technology
Carl-Zeiss-Promenade 10
D-07745 Jena
Germany

S.A. Boppart
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
405 North Mathews Avenue
Urbana, IL 61801
USA
boppart@uiuc.edu

W.T. Cathey
Electrical and Computer Engineering Department
University of Colorado at Boulder,
Boulder, CO 80309-0425
USA
cathey@schof.colorado.edu

D. Chana
King's College London
Department of Physics
Strand
London WC2R 2LS
UK
deeph.chana@kcl.ac.uk

C.J. Cogswell
Electrical and Computer Engineering Department
University of Colorado at Boulder,
Boulder, CO 80309-0425
USA
carol.cogswell@schof.colorado.edu

J.C. Dainty
Dept. of Experimental Physics
National University of Ireland, Galway
University Road, Galway
Ireland
c.dainty@nuigalway.ie

P. Dirksen
Philips Research Laboratories
Building WA
Professor Holstlaan 4
NL - 5656 AA Eindhoven
The Netherlands
peter.dirksen@philips.com

H.J. Dobschal
Carl Zeiss AG, Central Research and Technology
Carl-Zeiss-Promenade 10
D-07745 Jena
Germany

J. Huisken
MBL-Heidelberg
Light Microscopy Group
Cell Biophysics Programme
Meyerhofstrasse 1
D-69117 Heidelberg
Germany
jan.huisken@EMBL-Heidelberg.de

S. Inoué
Marine Biological Laboratory
7 MBL Street
Woods Hole, MA 02543-1015
USA
jmacneil@mbl.edu

A.J.E.M. Janssen
Philips Research Laboratories
Building WA
Professor Holstlaan 4
NL - 5656 AA Eindhoven
The Netherlands
peter.dirksen@philips.com

S.-h. Jiang
King's College London
Department of Physics
Strand
London WC2R 2LS
UK

R. Juškaitis
University of Oxford
Department of Engineering Science
Parks Rd
Oxford OX1 3PJ
UK
rimas.juskaitis@eng.ox.ac.uk

F. Lagugné Labarthet
Institut des Sciences Molculaires
UMR 5255 CNRS
Universit Bordeaux 1
351 Cours de la Libration
33405 Talence Cedex
France
f.lagugne@ism.u-bordeaux1.fr

M. Müller
Biophysics and Microscopy Group
Swammerdam Institute for Life Sciences
University of Amsterdam
P.O. Box 94062, 1090 GB Amsterdam
The Netherlands
muller@bio.uva.nl

P. Neocleous
King's College London
Department of Physics
Strand
London WC2R 2LS
UK

M. Ohtsu
Tokyo Institute of Technology
Interdisciplinary Graduate School of
Science and Engineering,
4259 Nagatsuta, Midori-ku,
Yokohama 226-8502
Japan
ohtsu@ae.titech.ac.jp

E.R. Pike
King's College London
Department of Physics
Strand
London WC2R 2LS
UK
erp@maxwell.ph.kcl.ac.uk

A. Rohrbach
Institute of Microsystem Technology
Albert-Ludwigs-University of Freiburg
Georges-Koehler-Allee 102
79110 Freiburg
Germany
rohrbach@imtek.uni-freiburg.de

M. Roy
University of Sydney
Department of Physical Optics
School of Physics A28
NSW 2006
Australia
m.roy@physics.usyd.edu.au

Y.R. Shen
University of California, Berkeley
Department of Physics
Berkeley, CA 94720-7300
USA
shenyr@physics.berkeley.edu

C.J.R. Sheppard
Division of Bioengineering
National University of Singapore
9 Engineering Drive 1
Singapore 117576
colin@nus.edu.sg

S.A. Sherif
Institute for Microstructural Sciences
National Research Council Canada
1200 Montreal Rd, M-36, Room 135
Ottawa ON K1A 0R6
Canada
Sherif.Sherif@nrc-cnrc.gc.ca

M.G. Somekh
School of Electrical and Electronic
Engineering
University Park
Nottingham NG8 2RD
UK
mike.somekh@nottingham.ac.uk

E.H.K. Stelzer
MBL-Heidelberg
Light Microscopy Group
Cell Biophysics Programme
Meyerhofstrasse 1
D-69117 Heidelberg
Germany
Ernst.Stelzer@EMBL-Heidelberg.de

C.-K. Sun
National Taiwan University
Institute of Electro-Optical Engineering
and Department of Electrical
Engineering
Taipei 10617
Taiwan
sun@cc.ee.ntu.edu.tw

P. Török
Blackett Laboratory
Imperial College
Prince Consort Rd.
London SW7 2BW
UK
peter.torok@imperial.ac.uk

Part I

High Aperture Optical Systems and Super-Resolution

1 Exploring Living Cells and Molecular Dynamics with Polarized Light Microscopy

S. Inoué

1.1 Introduction

Here, I would like to discuss the reasons, and methods, for using polarized light microscopy for studying the organization, and function, of living cells. In illustrating these points, I will be relying on some biological examples that I am particularly familiar with, namely, those related to the events of mitosis and cell division, which represent some of the more fundamental and critical events in the life of all living organisms.

Now, why should we be interested in polarized light microscopy for the study of living cells? What is special about living cells and why is polarization microscopy particularly useful for their studies?

As biologists have learned over the last half century, the cell is highly organized not only at the microscopically resolvable micrometer level, but also at diverse ultrastructural levels, that is, down at nanometer levels, many times smaller than the classical limit of resolution of the light microscope. Furthermore, the structures in the living cells are dynamic and their molecular organizations and locations are often changing rapidly with time.

In order to follow these *dynamic* events, the light microscope is the instrument of choice so long as we can penetrate and gain the desired information at the submicroscopic levels. Fortunately, many structures inside living cells are organized into local para-crystalline arrays, and show a weak but definite optical anisotropy [1,2]. The optical anisotropy, such as birefringence can be detected and measured with polarized light microscopy, if examined with sufficiently high sensitivity, and spatial and temporal resolution.

What I and others have been able to do over the last half century is to improve the sensitivity for detecting the optical anisotropy, especially the weak birefringence retardation exhibited by structures inside functioning living cells, and to make such detection and measurement possible at reasonably high spatial and temporal resolutions. The birefringence that is detected non-invasively, in turn, signals the events taking place at the submicroscopic, molecular, and even submolecular levels.

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

1.2 Equipment Requirement

As the pioneer W. J. Schmidt pointed out in the 1930's, the combined use of high extinction polars (polarizers), strain-free optics, a Brace-Koehler compensator, and a bright light source is indispensable for using a polarizing microscope to detect the very low birefr+ingence retardation (BR) exhibited by many organelles in living cells (e.g., [2]). In fact, the BR of many components of interest in the living cell typically ranges from a few nm to a small fraction of an nm, making them detectable only with very high extinction polarization optics [3,4].

With the small dimensions of many of the structures of interest, it is also necessary to improve the extinction factor ($I_{\parallel}/I_{\perp}$) for microscope condenser and objective lenses that provide high numerical aperture (NA), so that one can combine the high sensitivity for detecting and measuring small BR with high spatial resolution. Even using carefully selected strain-free lenses, however, the extinction factor between crossed polars generally drops exponentially as the NA is increased, primarily due to the differential transmittance of the parallel and perpendicular vectorial components

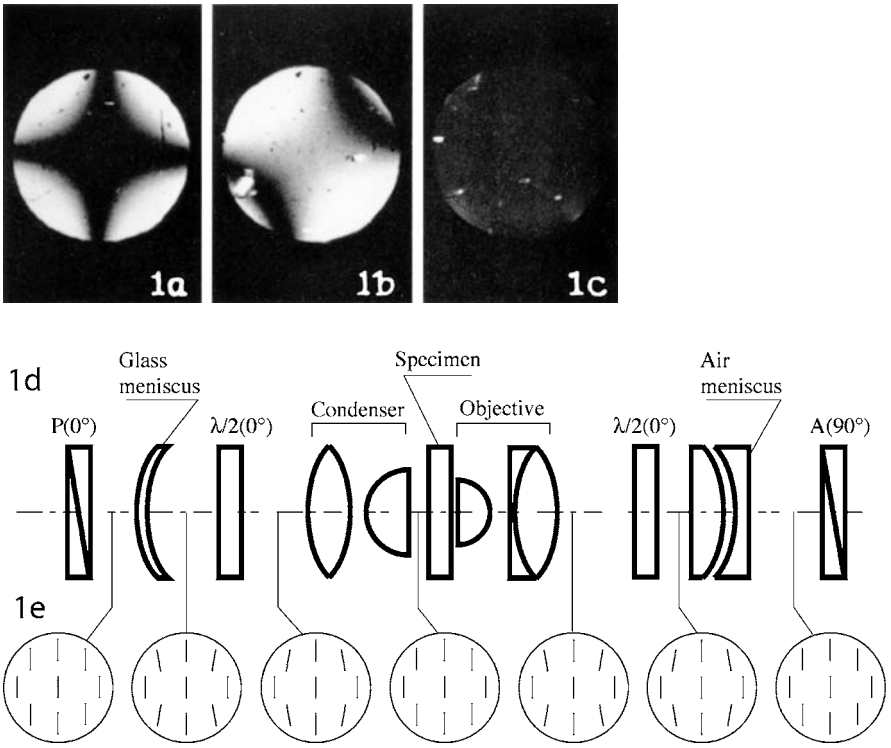


Fig. 1.1. Back aperture of 1.25 NA oil-immersion objective lens and matched NA condenser. (**a**) between crossed polars. (**b**) with analyzer rotated clockwise 3 degrees. (**c**) with crossed polars after rectification. (**d**) schematic of microscope with meniscus rectifiers. (**e**) orientation of polarized light at indicated planes. From [5]

of polarized light as the rays pass through oblique air-glass interfaces. This results in the "Maltese cross" seen at the back aperture of the objective lens, whereby up to several percent of the light passing between the arms of the cross escapes through the analyzer and drastically reduces the extinction factor at high NAs (Figs. 1.1 (a), 1.1 (b)). We were able to overcome this problem and gain the required high extinction factor, as well as a uniformly extinguished back aperture of high NA lenses, by introducing the polarization rectifier (Figs. 1.1 (c), 1.1 (d), 1.1 (e)) [5]. Rectification also eliminates the diffraction anomaly that, in non-rectified lenses, distorts the unit diffraction pattern (of weakly birefringent point sources) from an Airy disc to a four-leaf clover pattern (Figs. 1.2 (a), 1.2 (b)) [6,7]. An example of the sensitivity and resolution gained by the use of rectified high NA lenses is shown in Fig. 1.3. These panels display the arrangement and distribution of DNA bundles in the living sperm head of an insect, cave cricket. The images taken under three compensator settings show, at very high optical resolution and veritable contrast, the helical arrangement of the negatively birefringent DNA bundles, as well as the tandem packing arrangement of the chromosomes. These features of living sperm structure had never before been seen in sperm of any species by any mode of microscopy.

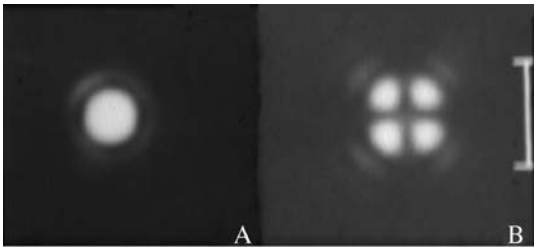


Fig. 1.2. Diffraction image of pinhole observed with 1.25 NA objective lens and condenser. (**a**) after rectification. (**b**) in the absence of rectifiers. Scale bar = 1.0 μm. From Inoué and Kubota, 1958 [6]

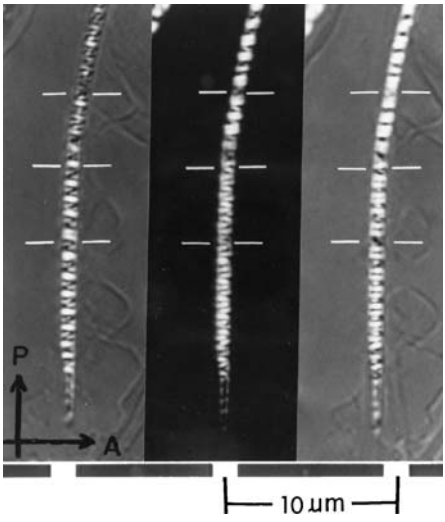


Fig. 1.3. Images of live sperm head of cave cricket. The sperm was immersed in dimethyl sulfoxide and observed between crossed polars at three compensator settings. From [8]

More recently, as an extension of the LC Pol-Scope (that I will touch on a little later), we developed another type of rectifier (Fig. 1.4). These new rectifiers reduce the ellipticity introduced by the anti-reflection coatings on the lens surfaces in addition to the differential attenuation (di-attenuation) of the s- and p-components that we primarily compensated for with the earlier rectifiers.

When a microscope is equipped with oppositely rotating circular polars instead of crossed linear polars, the back aperture of non-rectified objective and condenser lenses shows a dark center surrounded by a brighter ring (Fig. 1.5). The light from the outer zone of the aperture arises from a combination of di-attenuation of the s- and p-components and elliptical polarization. The elliptical polarization for modern microscope lenses with multi-layer anti-reflection coatings tend to be considerably larger than in earlier lenses with single MgF coatings.

Both the ellipticity and di-attenuation at high NA lens surfaces can be reduced quite effectively by using the new universal rectifier made of computer-driven sectored liquid crystal elements and $\lambda/2$ liquid-crystal element (Fig. 1.5). Figures 1.6 and 1.7 show an optical-bench-inverted polarizing microscope that we developed

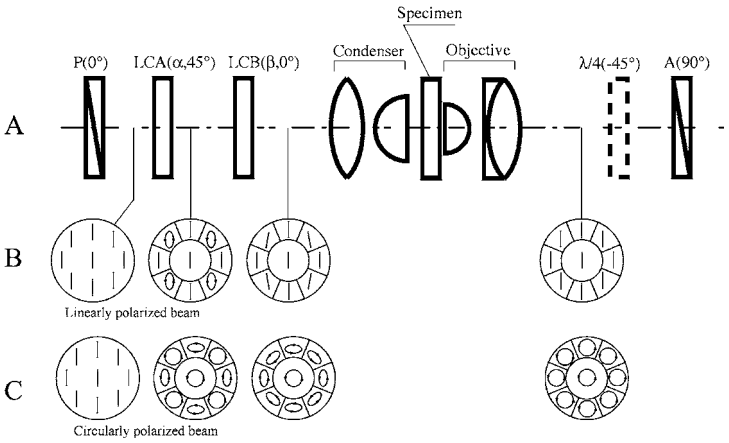


Fig. 1.4. Rectifier using sectored liquid crystal. (**a**) schematic of optical system. (**b**), (**c**) states of polarization at planes indicated. From [9]

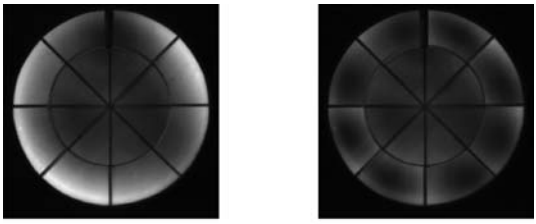


Fig. 1.5. Back aperture of 40×/0.85 NA lens. (**a**) observed between left and right circular polars without rectification. (**b**) in presence of sectored LC-rectifier. From [9]

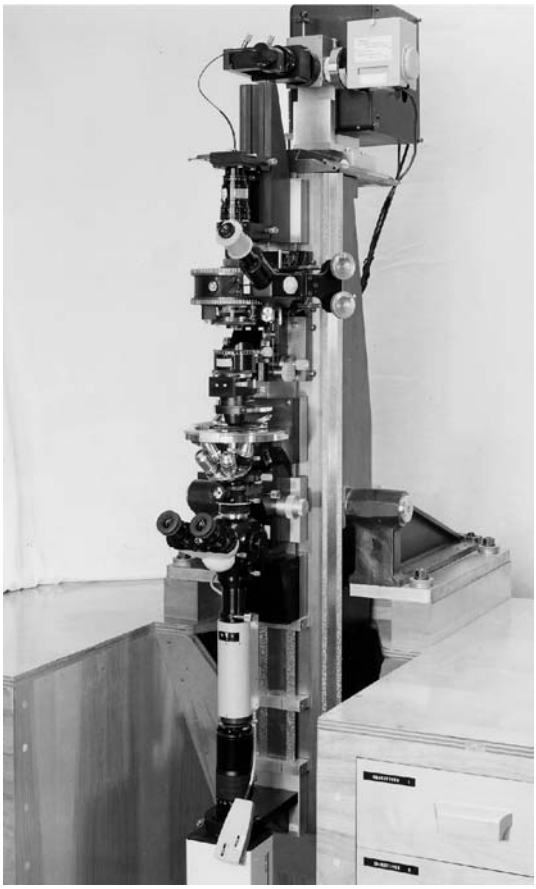


Fig. 1.6. Optical bench universal microscope. From [10]

for our biological and optical studies (see, e.g., pp. 158–162 in [10]). In addition to meeting the optical requirements discussed above, we prefer to use monochromatic green light (546 nm) from a 100 Watt mercury arc lamp. That satisfies the need to gain high field luminance (required since the analyzer removes much of the light) and to do so in the green spectral region, where least damage is introduced to living cells in general, despite the high intensity of illumination that the specimen receives. The uneven, very small image of the concentrated mercury arc is passed through a fiber-optic light scrambler in order to gain uniform field, and uniform full aperture, illumination [11]. (Light scramblers optimized for various microscopes are available from Technical Video, Ltd., PO Box 693, Woods Hole, MA 02543, USA; rknudson@tiac.net.)

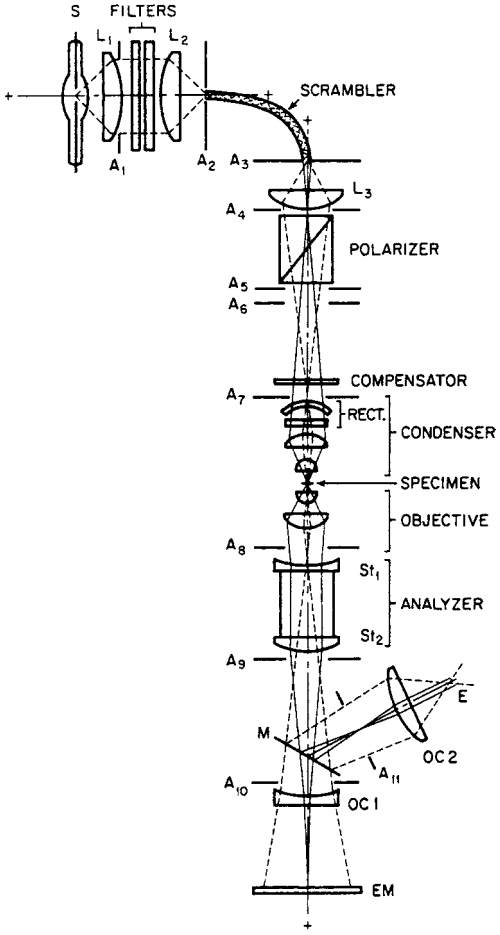


Fig. 1.7. Schematic optical path for microscope in Fig. 1.6. From [10]

1.3 Biological Examples

Before continuing with more instrumentation issues, let me introduce some examples of dynamic biological events. The first scene shows the segregation of chromosomes in a living plant cell of *Haemanthus*, the African blood lily (Fig. 1.8 and video). These scenes were taken in phase contrast by A. S. Bajer and J. Mole-Bajer in the 1950's (for on-line video see [12]), soon after Zernike's phase contrast microscope became available commercially.

In these video copies of the time-lapsed sequences, originally captured on 16 mm movie film, phase contrast clearly brings out the structure and behavior of the nucleolus and chromosomes, whose refractive indexes are somewhat greater than that of the surrounding nucleoplasm and cytoplasm. We clearly see the dissolution of the nucleoli, the condensation, splitting, and anaphase movement of the chromosomes, as well as the appearance of the cell plate that separates the cell

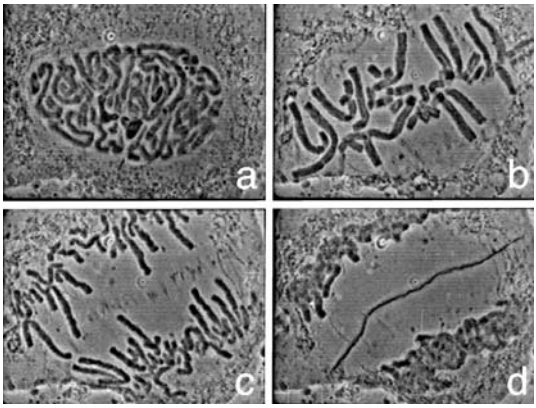


Fig. 1.8. Mitosis and cell plate formation in plant cell observed with phase contrast microscopy. After [13,14]. See also on-line version at www.molbiolcell.org. From [12]

between the two daughter nuclei. The machinery responsible for the alignment and segregation of the chromosomes, and the deposition of the cell plate, were, however, undetectable with bright field or phase contrast microscopy in the living cells. The filaments that were reputed to be attached to the chromosomes, and those that deposit the cell plate, were visible only after the cells were killed with acidic fixatives and not in healthy living cells. Such observations had led to many theories over half a century as to how chromosomes move, including their autonomous motility. Since it has long been known that the proper segregation of chromosomes is essential for proper bipartitioning of hereditary factors, and errors in proper bipartitioning lead to developmental anomalies as well as cancer and other critical diseases, the mechanism of mitosis has received much attention over the years.

In 1951, using a sensitive hand-built polarizing microscope (Fig. 1.9), I was finally able to show the fibrous elements of the mitotic spindle that bring the chromosomes to the cell equator and then move them apart to the two spindle poles. In the sensitive polarizing microscope, the dynamic appearance, disappearance, and changes of the fibers (made up of weakly birefringent parallel fibrils, now known as microtubules) were clearly visible as they aligned the chromosomes to the metaphase plate and then brought them to the spindle poles (Fig. 1.10 and video). These observations on the low levels of birefringence , made directly on healthy living cells from several species of plants and animals, finally settled a 50 year controversy as to the reality of those fibers and fibrils and opened up a new era of study on the mechanism of mitosis, chromosome and organellar transport, and the establishment of cell polarity. In addition, the fluctuating birefringence that demonstrated the dynamic nature of the protein filaments opened up a new concept that those filaments (microtubules) and their subunit molecules (tubulin) were in a dynamic equilibrium (Fig. 1.11 and video). We showed not only that microtubule-based structures could be assembled in the living cell reversibly by polymerization of tubulin and then be taken apart again after finishing a particular physiological task, but also that the microtubules could be depolymerized reversibly by cold and hydrostatic pressure, in other words, thermodynamically review in [16]. In the

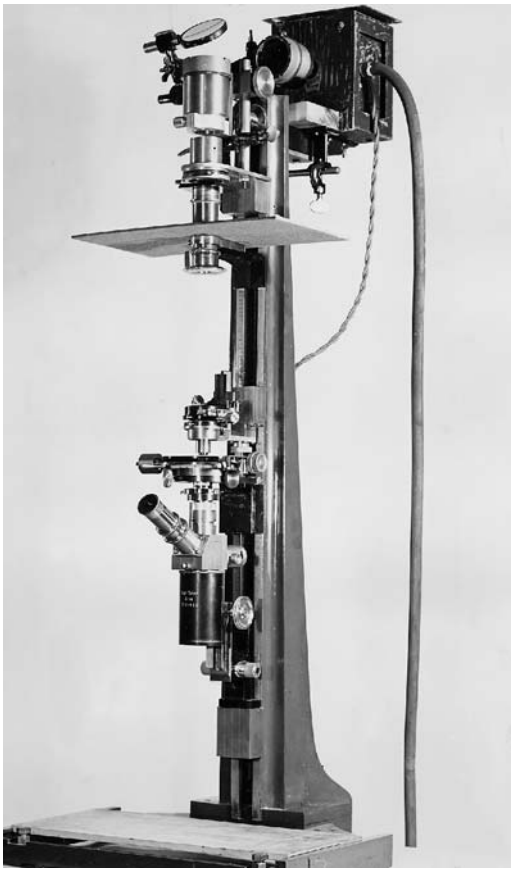


Fig. 1.9. Universal, inverted polarizing microscope designed by author in 1949. From [15]

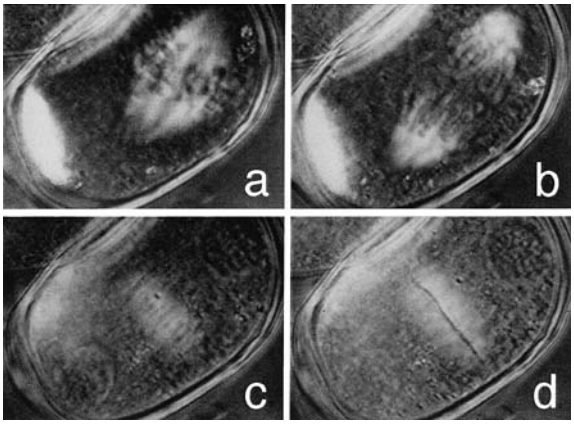


Fig. 1.10. Mitosis and cell plate formation in pollen mother cell of Easter Lily observed with sensitive polarizing microscope shown in Fig. 1.9. From [19]

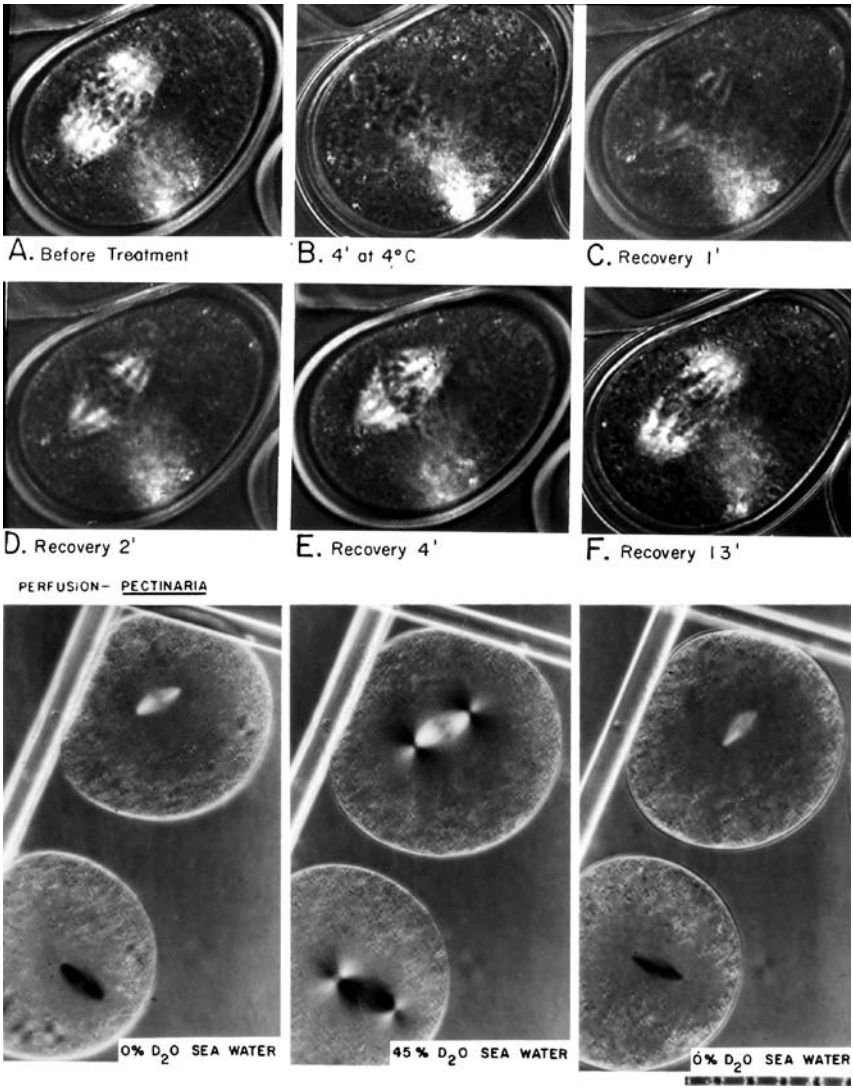


Fig. 1.11. Top. Reversible disassembly of spindle fiber microtubules induced by cold. (**a**) before chilling. (**b**) after 4 min at $-4°$C. (**c**)–(**f**): recovery at room temperature. From [19]. Bottom. Reversible assembly of spindle fiber microtubules by D_2O. Scale 10 μm. From [17]

meanwhile, we posited that the self-assembling microtubules could push and position cell organelles such as chromosomes, and by the very act of depolymerizing, pull objects to their destinations (review in [17,18]). Such heretic notions, that we derived by observing the birefringence changes in living cells, took many years to be fully accepted. Eventually, however, they were proven to be sound when isolated microtubules in media free of Ca^{2+} ions were disassembled reversibly by exposure

to cold in the test tube [20,21]. In fact, chilling cell extracts to depolymerize microtubules, centrifuging to remove formed particles, and then warming the clear supernatant to polymerize microtubules has become the standard method for separating tubulin from other cell ingredients and for obtaining purified microtubules. By the late 1980's, a number of investigators have further been able to show that single microtubules isolated or assembled outside of the living cell could remain attached to and pull loads, such as chromosomes, to the microtubule anchoring point, as the microtubule itself was made to disassemble, even in the absence of common chemical energy sources such as ATP review in [18].

1.4 Video-Enhanced Microscopy

Returning to advances in light microscopy, one of the developments over the last couple of decades that made it possible to shed considerably more light on molecular behavior in living cells and functional cell extracts is video-enhanced microscopy. Surprisingly, there turned out to be much useful information in the microscope image that had been hidden and unexplored in the past but could now be brought forth by video enhancement. For example, the next video-enhanced DIC (Nomarski) sequence shows, in real time, the energetic growth of the acrosomal process from the tip of a sperm head triggered by its close approach to the egg (Fig. 1.12 and

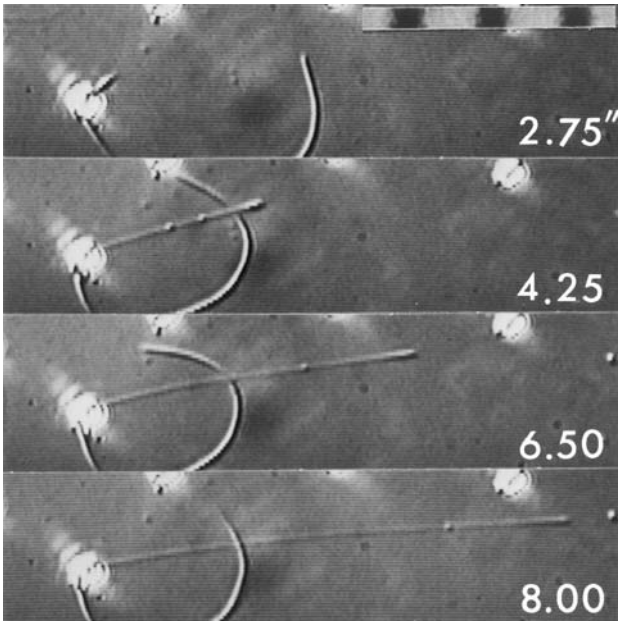


Fig. 1.12. Growth of 65 nm diameter acrosomal process from sperm of a sea cucumber. Video-enhanced DIC, elapsed time in seconds, scale bar = 10 μm. After [22]

video). The growth of this thin process, which in the electron microscope measures only 65 nm in diameter, is propelled by an explosive polymerization of actin molecules within the process, and is essential for successful fertilization in almost all animal species. In order to bring out the features previously hidden, all that was needed was to use a suitable video camera and circuits, to subtract away the stationary background optical noise (inhomogeneity), and to raise the image contrast by boosting the amplifier gain [23,24]. Thus, the faint diffraction image of single microtubules, only 25 nm wide, could now be clearly visualized with the light microscope by video-enhanced DIC. Furthermore, the gliding behavior of individual microtubules relative to the motor proteins attached to the coverslip, as well as the polarized growth, shortening, and distortion of the microtubules themselves, could be followed and quantitated at video rate. Such observations have led to the discovery and characterization of many new motor proteins and other molecules that associate with microtubules and transport organelles, as well as the microtubules themselves, in the living cell.

It should be noted in passing that, in contrast to fluorescence or dark-field microscopy, video-enhanced DIC and polarization microscopy yield rather shallow depths of field and, therefore, opportunities for serial optical sectioning and three-dimensional image construction even in the absence of confocal optics. Examples are shown in the following two video sequences. The first sequence shows real-time through-focal optical sections of live sea urchin embryo in video-enhanced DIC [25]. The second sequence shows real-time through-focal optical sections in video-enhanced polarization microscopy of a developing sea cucumber embryo (S. Inoué, unpublished data).

In addition to DIC microscopy, electronic imaging and processing in many forms are now playing important roles in confocal, fluorescence, and polarization microscopy as well. Among some of the more recent developments in polarization microscopy, I would like to touch on some of our own further contributions that include: a new type of polarized light microscope called the LC Pol-Scope, the centrifuge polarizing microscope, and observations of fluorescence polarization.

1.5 The LC Pol-Scope

The LC Pol-Scope was devised by Rudolf Oldenbourg in our Program at the Marine Biological Laboratory in Woods Hole, Massachusetts, to make possible the simultaneous imaging of weakly birefringent objects regardless of their axis orientation, and without the need for any mechanical adjustments of the specimen orientation or of the polarization optical components. With conventional polarization microscopy, the contrast of a birefringent object varies as the $\sin^2(2\times$ orientation angle), so that the specimen contrast drops to zero when the optical axes of the specimen come to lie parallel to the transmission axes of the crossed polars (polarizer and analyzer). With the LC Pol-Scope, not only is this angle dependence eliminated, but the brightness of the image is now directly proportional to the BR, rather than to $\sin^2$(BR), and thus directly reflects, e.g., the number of unit filaments that are packed into a fiber

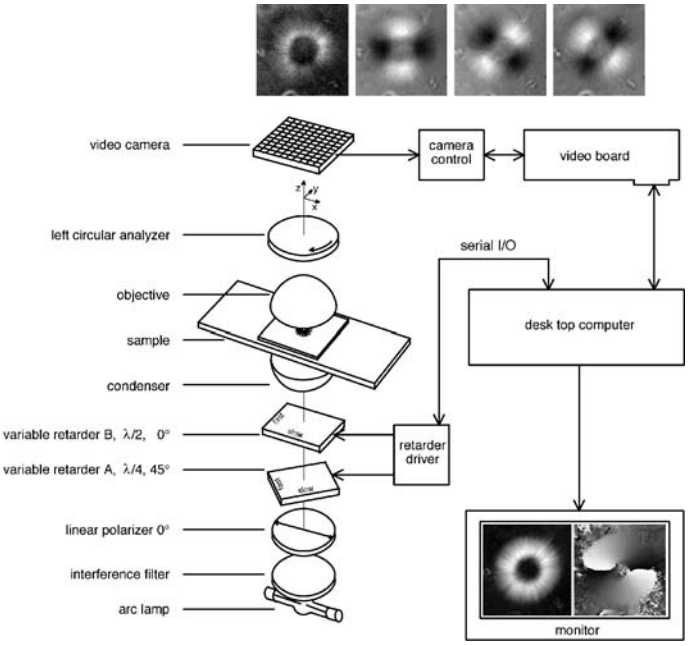


Fig. 1.13. Schematic of Oldenbourg's LC Pol-Scope. Top row: four images sent to computer. Bottom left: retardance image. Bottom right: gray scale image of azimuth angle. From [26]

whose diameter is less than the microscope's unit diffraction width [26]. As shown in the schematic of the system (Fig. 1.13), the LC Pol-Scope uses two liquid crystal (LC) retarders in place of a regular (mechanically adjusted) compensator, and a circular analyzer in place of a linear polarizing analyzer. A computer calculates and quickly displays the retardance distribution image from the four images captured by a digital camera at preset LC voltages. The next videos (Fig. 1.14 and videos) show examples of the dynamic retardance images. Figure 1.15 illustrates the sensitivity of the LC Pol-Scope for measuring the retardance of submicroscopic filaments.

1.6 The Centrifuge Polarizing Microscope

When living cells suspended in a density gradient are exposed to high acceleration in a centrifuge, the cell contents gradually stratify within the cell according to their density differences. For example, in an invertebrate egg cell, the heavy pigment and yolk granules sink to the bottom, while the lighter lipid granules and the nucleus accumulate at the top. By developing a centrifuge microscope, which allows the use of polarized light for observations of the weak cellular BR, we have been able to observe the alignment of fine structures in living cells associated with stratification of their organelles. As the organelles stratify, some membrane components become regularly oriented and exhibit birefringence (Fig. 1.16 (a), 1.16 (b), and video). Sur-

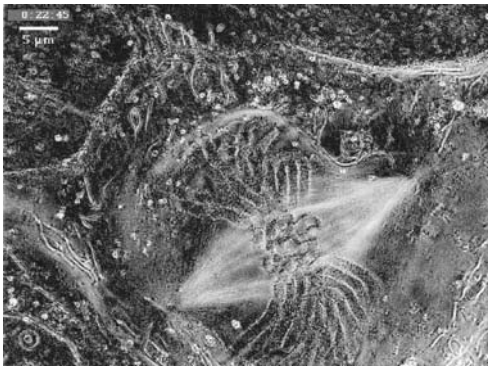


Fig. 1.14. Newt lung epithelial cell undergoing mitosis, observed with the LC Pol-Scope. See also on-line version at www.molbiolcell.org. From [26]

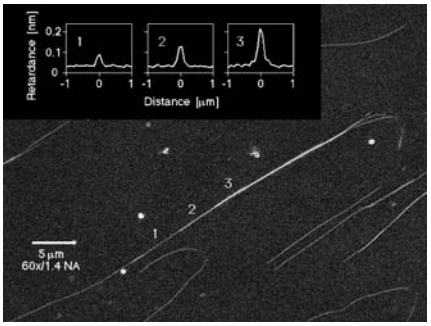


Fig. 1.15. Birefringence distribution of microtubules recorded with the LC Pol-Scope. Insets show retardance versus distance plots measured across single 1, 2, and 3 microtubules. From [27]

prisingly, the birefringence disappears in a few seconds when the egg is activated, e.g., by the introduction of sperm or influx of Ca^{2+} ions into the cell followed by assembly of the meiotic spindle (Fig. 1.16 (c), 1.16 (d), and video). These changes that signal an early response of the egg cell to activation, as well as the emergence of polymerizing microtubules that generate the meiotic spindle in the centrifuged egg fragment, are illustrated in the video taken with the centrifuge polarizing microscope. This centrifuge polarizing microscope, which we developed in collaboration with Hamamatsu Photonics and Olympus Optical, required several technical innovations (Fig. 1.17). In brief, the specimen is carried in a special leak-proof chamber with low stress-birefringence windows; the chamber, supported in the rotor mounted on an air-spindle motor, transects the optical axes of the microscope between the objective and condenser lenses; and a frequency-doubled Nd:YAG laser fires a 6 nsec pulse exactly as the specimen comes in alignment with the microscope axis so that the image is frozen and stationary to better than 1.0 µm (Fig. 1.18 and video). Speckles that arise from the highly monochromatic, brief laser flashes are eliminated by using multiple-length optical fibers in the illumination path, and the development of interference patterns at or near the camera face plate is prevented by introducing a fiber bundle plate in front of the CCD detector. Different regions of the spinning specimen chamber are viewed by remotely driving the microscope, mounted on an *x*, *y*, *z* set of sleeve bearings. The polarization optics and interference fringe-free video camera provide a 1 nm retardance sensitivity for measuring birefringence, as

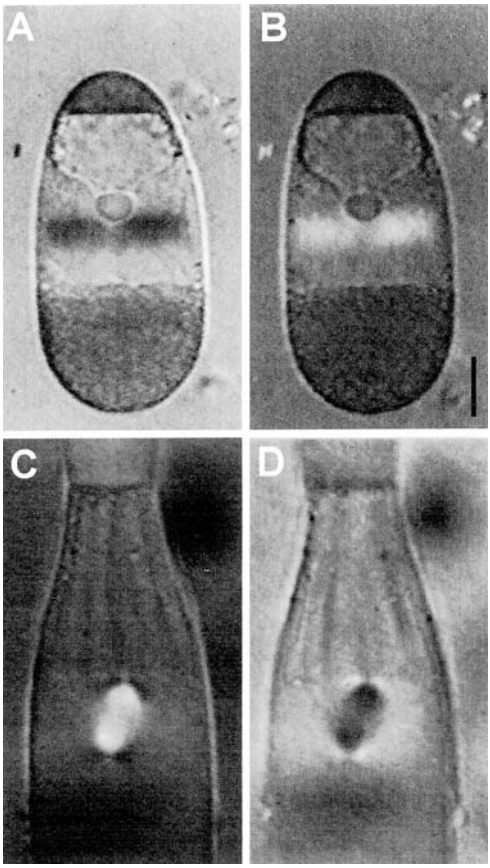


Fig. 1.16. Parchment worm egg observed in the centrifuge polarizing microscope. (**a**), (**b**) stratified cell with oil cap, very large nucleus with protruding nucleolus layered above negatively birefringent membrane layers that are oriented vertically. (**c**), (**d**) ca. 10 minutes after activation. The nuclear envelope and membrane birefringence are gone and replaced by the positively birefringent meiotic spindle. (**a**), (**c**) with compensator slow axis oriented vertically. (**b**), (**d**) horizontally. From [28]

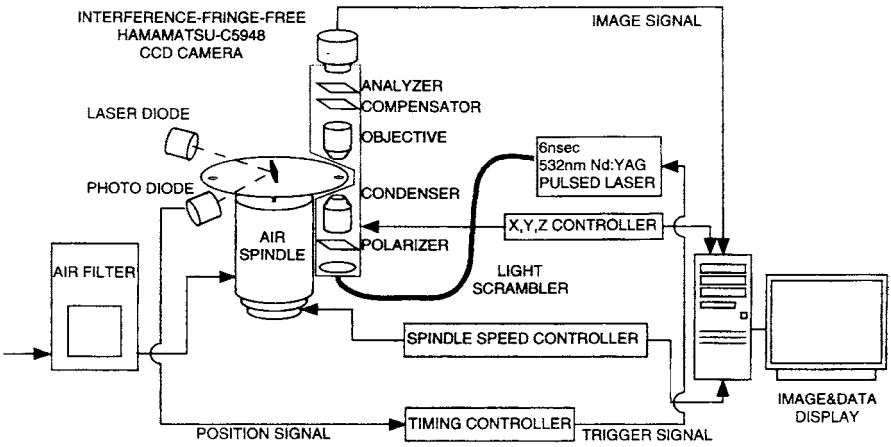


Fig. 1.17. Schematic of centrifuge polarizing microscope. From [29]

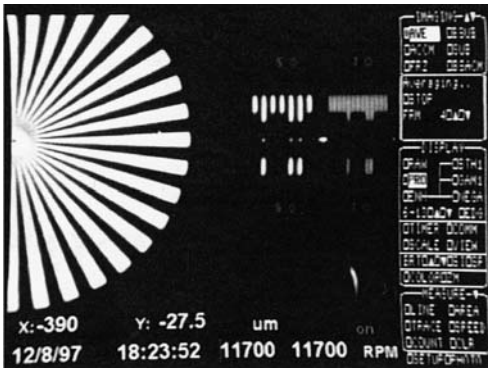


Fig. 1.18. Image of MBL-NNF test target recorded with centrifuge polarizing microscope rotating at 17000 rpm (10500× Earth's gravitational acceleration). 1 and 2 μm rulings show to the right of the Siemens test star. From [29]

well as capability for fluorescence (532 nm excitation) and DIC observations. The system and chamber design are detailed in a recent article [29], together with examples of several applications in the study of living cells and related objects [30].

1.7 Polarized Fluorescence of Green Fluorescent Protein

Finally, I would like to describe some fluorescence polarization properties of green fluorescent protein (GFP) crystals that we have been studying since last summer. GFP is a protein extracted from the light-producing organ of a jellyfish, *Aequorea* sp., and is widely used today by biologists as a sensitive fluorescent reporter [31]. GFP is a remarkably stable fluorescent protein whose gene can be expressed by living cells together with other selected gene products without interfering with either gene function. It is thus used for signaling the expression, location, and interaction of a variety of proteins of interest in living cells and tissues.

Our own interest in the fluorescence of this material was awakened when we examined the fluorescence emitted by crystals of purified native GFP under the polarizing microscope (Fig. 1.19). When illuminated, or examined, through a linear polar, the crystals exhibited a maximum-to-minimum ratio of fluorescence as high as 6:1 (Fig. 1.20 (a), 1.20 (b)). Furthermore, between a pair of polars *oriented parallel to each other*, the maximum-to-minimum fluorescence ratio rose to an incredibly high ratio of 30:1 (Fig. 1.20 (c)). Unexpectedly, the latter is nearly the product of the ratios measured with a polarizer or an analyzer alone. In fact, as the specimen is rotated on the microscope's revolving stage, we find that the fluorescence intensity follows a function that includes a cosine 4^{th} power term (square of cosine squared terms) [32]. We believe these observations re-awaken the utility of polarized light microscopy for studying the arrangement of chromophores in a crystal or semi-crystalline material. Furthermore, the use of parallel polars (in other words, using both a polarizer and analyzer on a polarizing microscope, but with their axes oriented parallel rather than perpendicular to each other) should substantially improve the sensitivity for dynamically following the changes in orientation even of

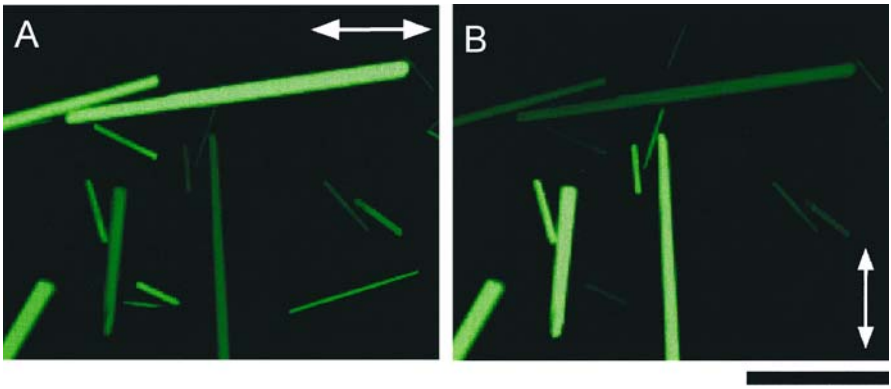


Fig. 1.19. Fluorescence anisotropy by crystals of native GFP. (**a**) polarizer transmission axis oriented horizontally. (**b**) vertically. Scale bar = 20 μm. From [32]

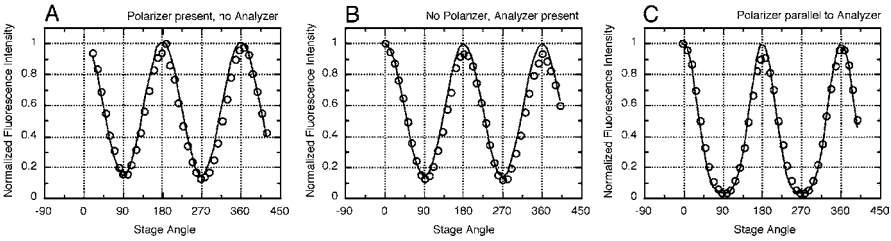


Fig. 1.20. Normalized fluorescence intensity versus GFP crystal orientation. (**a**) with polarizer but no analyzer. (**b**) with analyzer but no polarizer. (**c**) with parallel polarizer and analyzer. O: measured points; solid-line curves calculated from equations. From [32]

single fluorophores. For example, Kinosita [33] has used the orientation-dependent changes in polarized fluorescence of attached rhodamine to demonstrate the gyration of a single actin filament (ca. 12 nm in diameter) gliding on a lawn of myosin molecules. The increased sensitivity achieved by use of parallel polars instead of a single polarizing element could also prove useful for following conformational changes in portions of a single molecule tagged with appropriate fluorophores.

1.8 Concluding Remarks

In summary, while polarized light microscopy has still not been as widely used in biology as its potential indicates, it has played some critical roles in deciphering the molecular events underlying the enigmatic, yet vitally important, processes in living cells such as mitosis and the dynamic organizational roles played by microtubules. With expanded interest in biophysics and improved collaboration between optical physicists, electrical engineers, material scientists, and biologists, we look forward to further non-invasive explorations of living cells by polarized light microscopy,

and to the unveiling of hidden dynamic molecular and submolecular events that underlie the physical basis of life.

References

1. H. Ambronn, A Frey: *Das Polarisationsmikroskop, seine Anwedung in der Kolloidforschung in der Farbarei* (Akademische Verlag, Leipzig 1926)
2. W.J. Schmidt: 'Die Doppelbrechung von Zytoplasma, Karyoplasma und Metaplasma', In: *Protoplasma Monographien*, Vol. 11 (Borntraeger, Berlin 1937)
3. M.M. Swann, M. Mitchison: J. Exp. Biol. **27**, 226 (1950)
4. S. Inoué, K. Dan: J. Morphol. **89**, 423 (1951)
5. S. Inoué, W.L. Hyde: J. Biophys. & Biochem. Cytol. **3**, 831 (1957)
6. S. Inoué, H. Kubota: Nature **182**, 1725 (1958)
7. H. Kubota, S. Inoué: J. Opt. Soc. Am. **49**, 191 (1959)
8. S. Inoué, H. Sato: 'Deoxyribonucleic Acid Arrangement in Living Sperm', In: *Molecular Architecture in Cell Physiology*, ed. by T. Hayashi, A. Szent-Györgyi (Prentice Hall, New Jersey 1966) pp. 209–248
9. M. Shribak, S. Inoué, R. Oldenbourg: Opt. Eng. **43**, 943 (2002)
10. S. Inoué, K.R. Spring: *Video Microscopy – The Fundamentals* (Plenum Press, New York 1997)
11. G.W. Ellis: J. Cell Biol. **101**, 83a (1985)
12. S. Inoué, R. Oldenbourg: Mol. Biol. Cell. **9**, 1603 (1998)
13. A. Bajer, J. Molè–Bajer: Chromosoma **7**, 558 (1956)
14. A. Bajer, J. Molè–Bajer: J. Cell Biol. **102**, 263 (1986)
15. S. Inoué: *Studies of the Structure of the Mitotic Spindle in Living Cells with an Improved Polarization Microscope PhD Thesis* (Princeton University, New Jersey 1951)
16. S. Inoué: J. Cell Biol. **91**, 131s (1981b)
17. S. Inoué, H. Sato: J. Gen. Physiol. **50**, 259 (1967)
18. S. Inoué, E.D. Salmon: Mol. Biol. Cell. **6**, 1619 (1995)
19. S. Inoué: 'Organization and Function of the Mitotic Spindle', In: *Primitive Motile Systems in Cell Biology*, ed. by R.D. Allen, N. Kamiya (Academic Press, New York 1964) pp. 549–598
20. R. Weisenberg: Science **177**, 1196 (1972)
21. J.B. Olmsted, G.G. Borisy: Biochemistry **14**, 2996 (1975)
22. L.G. Tilney, S. Inoué: J. Cell Biol. **93**, 820 (1982)
23. R.D. Allen, J.L. Travis, N.S. Allen, H. Yilmas: Cell Motil. **1**, 275 (1981)
24. S. Inoué: J. Cell Biol. **89**, 346 (1981)
25. S. Inoué, T.D. Inoué: Biol. Bull. **187**, 232 (1994)
26. R. Oldenbourg: Nature **381**, 811 (1996)
27. R. Oldenbourg, E.D. Salmon, P.T. Tran: Biophys. J. **74**, 645 (1998)
28. S. Inoué: FASEB J. **13**, S185 (1999)
29. S. Inoué, R.A. Knudson, M. Goda, K. Suzuki, C. Nagano, N. Okada, H. Takahashi, K. Ichie, M. Iida,K. Yamanaka: J. Microsc. – Oxford **201**, 341 (2001)
30. S. Inoué, M. Goda, R.A. Knudson: J. Microsc. – Oxford **201**, 357 (2001)
31. S. Inoué, M. Goda: Biol. Bull. **201**, 231 (2001)
32. S. Inoué, O. Shimomura, M. Goda, M. Shribak, P.T. Tran: P. Natl. Acad. Sci. USA **99**, 4272 (2002)
33. K. Kinosita Jr.: FASEB J. **13**, S201 (1999)

2 Characterizing High Numerical Aperture Microscope Objective Lenses

Rimas Juškaitis

2.1 Introduction

Testing and characterization of high quality lenses have been perfected into fine art with the advent of lasers, phase-shifting interferometers, CCD cameras and computers. A bewildering array of techniques is described in Malacara's classical reference book on the subject [1]. Several of these techniques, in particular the Twyman–Green interferometer and the star test, are applicable to testing of microscope objective lens.

Characterizing high-numerical aperture (NA) objective lenses presents unique challenges. Many of the standard approaches, including Twyman–Green interferometry, are in fact comparative techniques. They require a reference object – an objective or a concave reflective surface – of the same or larger numerical aperture and of perfect (comparatively speaking) optical quality. This is problematic. Even if two lenses of the same type are available a problem still remains of appropriating the measured aberrations to the individual lenses. The star test, which is absolute, hits a similar problem in that the Airy disc produced by the lens being tested is impossible to observe directly, and hence it has to be magnified by a lens with a similar or better resolution, i.e. higher NA. Immersion lenses create further complications. All tests described in this Chapter are free from these problems. They are absolute and use a small point scatterer or a flat mirror to create a reference wavefront against which the lens aberrations are checked. Together with advanced interferometric techniques and processing algorithms this results in a range of techniques suitable for routine characterization of all available microscope objective lenses.

2.1.1 Disclaimer

A few words have to be said regarding identity of the lenses used throughout this chapter. Just as in a movie, which despite being a work of fiction uses real life people as actors, I had to use flesh-and-blood microscope objective lenses in order to verify the techniques and gather some "typical" data. The choice of lenses was typically dictated by what was available in our lab at a particular moment. Inevitably these were not necessarily the top-of-the-range and most up-to-date specimens. Being acutely aware of how sensitive the lens manufacturers can be about publishing such data – positive or otherwise – I feel it prudent to precede this Chapter with the following disclaimers:

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

All characters in this chapter are entirely fictitious. Any resemblance to a living lens is purely coincidental. Any attempts to match data published here with real lenses and to infer any generalizations about their respective manufacturers are to be undertaken entirely at readers' risk.

and also:

No lenses were harmed during these experiments.

2.1.2 Objective Lens Basics

Before describing specific lens testing techniques it might be useful to repeat here a few basic facts about microscope objective lenses in general. Modern objective lenses are invariably designed for infinite conjugate ratio, i.e. the object of observation is placed in the front focal plane and its image is formed at infinity. In order to obtain a real intermediate image a separate lens, called the tube lens, is used. The focal length of this lens F (which ranges from 165 mm for Zeiss and 180 mm for Olympus to 200 mm for Leica and Nikon) together with the magnification of the objective M gives the focal length of the objective $f = F/M$.

One of the basic postulates of aberration-free lens design in that it has to obey the Abbe's sine condition. For a microscope objective treated as a thick lens this can be interpreted as the fact that its front principal surface is actually a sphere of radius f centered at the front focus. Any ray leaving the focus at an angle α to the optical axis is intercepted by this surface at the height $d = f \sin \alpha$ and emerges from the back focal plane parallel to the axis at the same height, as shown in Fig. 2.1. For immersion lenses this has to be multiplied by the refractive index of the immersion fluid n.

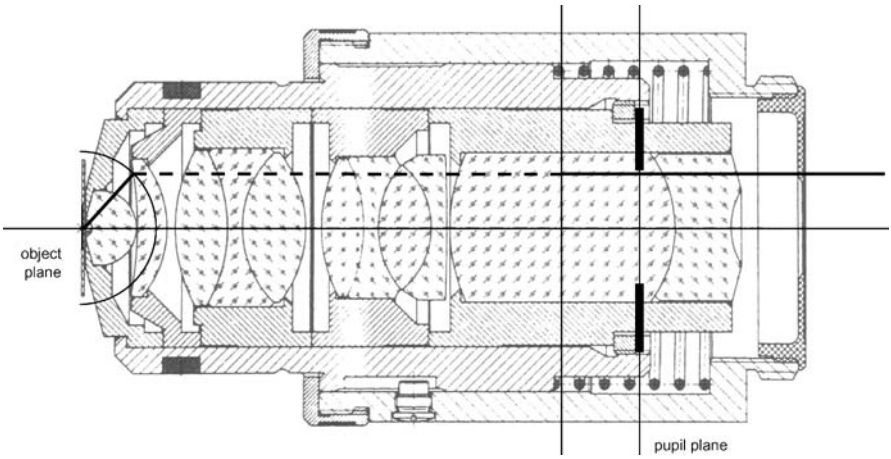


Fig. 2.1. Schematic diagram of a typical high NA planapochromat objective lens. Principal surfaces, aperture stop and marginal ray are indicated

In most high NA objective lenses the back focal plane, also called the pupil plane, is located inside the lens and is not therefore physically accessible. Fortunately lens designers tend to put an aperture stop as close to this plane as possible, which greatly simplifies the task of identifying the pupil plane when reimaging it using an auxiliary lens. Any extra elements, such as phase rings in phase contrast lenses or variable aperture iris diaphragms, will also be located in the back focal plane.

The physical aperture of an objective lens D is related to its numerical aperture $n \sin \alpha$ via

$$D = \frac{2Fn \sin \alpha}{M}. \tag{2.1}$$

Ultimately it is limited by the size of the objective thread. For a modern low magnification high NA immersion lens with, say, $n \sin \alpha = 1$ and $M = 20$, D can be as large as 20 mm. This is one of the reasons why some lens manufacturers (notably Leica and Nikon) have now abandoned the former golden standard of RMS thread and moved to larger thread sizes.

2.2 Point Spread Function

A perfect lens transforms a plane wave front into a converging spherical wave. Any deviations from this ideal behaviour, if they are not too dramatic, can be described by introducing a complex Pupil function $P(\rho, \theta)$, where ρ is the normalized radial coordinate in the pupil plane and θ is the azimuthal angle in the same plane. Both amplitude and phase aberrations can be present in a lens, but it is the latter that usually play the dominant role. The amplitude aberrations are typically limited to some apodization towards the edge of the pupil; these are discussed in more detail in Sect. 2.5.2.

The optical field distribution produced by this (possibly aberrated) converging wave is termed the Point Spread Function, or PSF, of the lens. This distribution can be obtained in its most elegant form if dimensionless optical coordinates in lateral

$$v = \frac{2\pi}{\lambda} n \sin \alpha \sqrt{x^2 + y^2} \tag{2.2}$$

and axial

$$u = \frac{8\pi}{\lambda} n \sin^2 \frac{\alpha}{2} z \tag{2.3}$$

directions are used. In these coordinates the intensity distribution in the PSF is independent of the NA of the lens, and the surface $u = v$ corresponds to the edge of the geometric shadow. The actual focal field distribution in these newly defined

cylindrical coordinates is given by [2]:

$$h(u,v,\psi) = A\exp\left[\frac{\mathrm{i}u}{4\sin^2\frac{\alpha}{2}}\right]\int_0^1\int_0^{2\pi} P(\rho,\theta) \times \exp\left\{-\mathrm{i}\left[v\rho\cos(\theta-\psi)+\frac{u\rho^2}{2}\right]\right\}\rho\,\mathrm{d}\rho\,\mathrm{d}\theta. \tag{2.4}$$

The exponential term in front of the integral is nothing else than a standard phase factor of a plane wave $2\pi nz/\lambda$.

For the aberration-free case $P = 1$ and the integral over θ can be calculated analytically to give $2\pi \mathrm{J}_0(v\rho)$. Equation (2.4) now simplifies to

$$h(u,v) = 2\pi A\exp\left[\frac{\mathrm{i}u}{4\sin^2\frac{\alpha}{2}}\right]\int_0^1 \exp\left(-\frac{\mathrm{i}u\rho^2}{2}\right)\mathrm{J}_0(v\rho)\,\rho\,\mathrm{d}\rho. \tag{2.5}$$

This equation is readily calculated either numerically, or using Lommel functions. The resulting intensity distributions are well know and can be found, for instance, in [2]. Not only that, but also PSFs subjected to various aberrations have been calculated countless times in the past and are instantly recognizable to most microscopists. It is precisely for this reason that a relatively straightforward measurement of the PSF can frequently provide an instant indication of what is wrong with a particular objective lens.

Equations (2.4) and (2.5) are, of course, scalar approximations. This approximation works remarkably well up to angular apertures of about 60°. Even above these angles the scalar approximation can be safely used as a qualitative tool. For those few lenses that seem to be beyond the scalar approach (and for the rigorous purists) there is always an option to use a well developed vectorial theory [3].

2.2.1 Fibre-Optic Interferometer

Requirement to measure both amplitude and phase of the PSF calls for an interferometer-based setup. The Fibre-optic interferometer, Fig. 2.2, that was eventually chosen for the task has several important advantages. It is an almost common-path system, which dramatically improves long-term stability. It is also a self-aligning system: light coupled from the fibre to the lens and scattered in the focal region is coupled back into the fibre with the same efficiency. For convenience the whole setup is built around a single-mode fibre-optic beam splitter, the second output of which is index-matched in order to remove the unwanted reflection. A He-Ne laser operating at 633 nm is used as a light source. The whole setup bears cunning resemblance to a confocal microscope. In fact, *it is* a confocal microscope, or at least can be used as such [4]! Provided that light emerging from the fibre overfills the pupil of the objective lens the former acts as an effective pinhole ensuring spatial filtering of the backscattered light. Thus if the object can be regarded as a point scatterer, then the amplitude of the optical signal captured by the fibre

$$R \sim h^2 = |h|^2\exp\left[2\mathrm{i}\arg(h)\right], \tag{2.6}$$

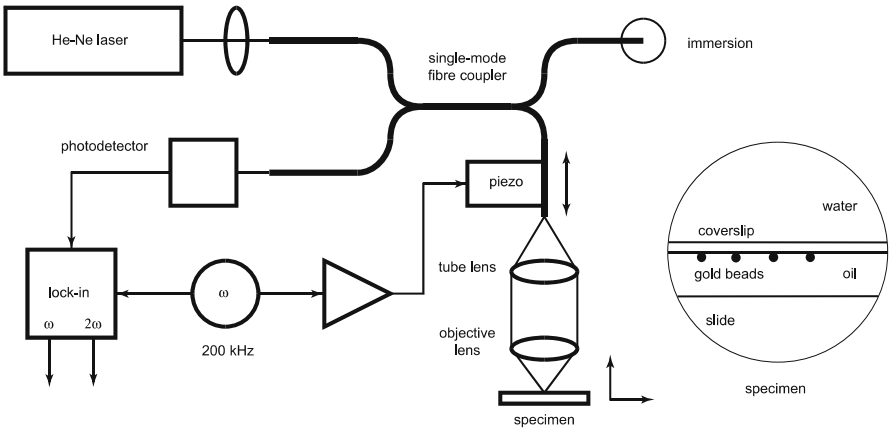


Fig. 2.2. Schematic diagram of the fibre-optic interferometer based setup for measuring objective PSFs

i.e. its magnitude is equal to the intensity of PSF whereas the phase is twice the phase of the PSF. In order to measure both these parameters light reflected back along the fibre from the tip is used as a reference beam r. Wavefronts of both signal and reference beams are perfectly matched in a single mode fibre – this is the beauty of fibre-optic interferometer. Furthermore, the fibre tip is dithered to introduce a time-varying phase shift between the interfering beams $\phi(t) = \phi_0 \cos \omega t$. The interference signal reaching the photodetector is now given by

$$I = |r + R|^2 = r^2 + |R|^2 + 2r\left[\mathrm{Re}(R)\cos(\phi_0 \cos \omega t) - \mathrm{Im}(R)\cos(\phi_0 \cos \omega t)\right], \quad (2.7)$$

were r was assumed to be real for simplicity. It is now a simple matter to extract both $\mathrm{Re}(R)$ and $\mathrm{Im}(R)$ from this signal by using lock-in detection technique. The signal of (2.7) is multiplied by $\cos \omega t$ and the result is low-pass filtered to give

$$I_1 = r\mathrm{J}_1(\phi_0)\mathrm{Im}(R), \quad (2.8)$$

whereas synchronous demodulation with $\cos 2\omega t$ yields

$$I_2 = r\mathrm{J}_0(\phi_0)\mathrm{Re}(R). \quad (2.9)$$

By appropriately adjusting the modulation amplitude ϕ_0 it is easy to achieve $\mathrm{J}_1(\phi_0) = \mathrm{J}_2(\phi_0)$ and, by substituting (2.6), to calculate

$$h \sim \sqrt{I_1^2 + I_2^2}\exp\left(\frac{\mathrm{i}}{2}\arctan\frac{I_1}{I_2}\right). \quad (2.10)$$

Thus the goal of obtaining both the amplitude and phase of the PSF of the objective lens has been achieved. Of course, in order to obtain full 2- (2-D) or 3-dimensional (3-D) PSF corresponding scanning of the object, the point scatterer, is still required.

2.2.2 PSF Measurements

In order to demonstrate the effectiveness of this method in detecting small amounts of aberrations it was tested on an special kind of objective lens. This 60× 1.2 NA water immersion plan-apochromat was developed for deconvolution applications and hence was specifically designed to have a well-corrected PSF. It was also equipped with a correction collar to compensate for cover glass thicknesses in the range 0.14–0.21 mm. 100 nm colloidal gold beads mounted beneath a Nr 1.5 coverslip of nom-

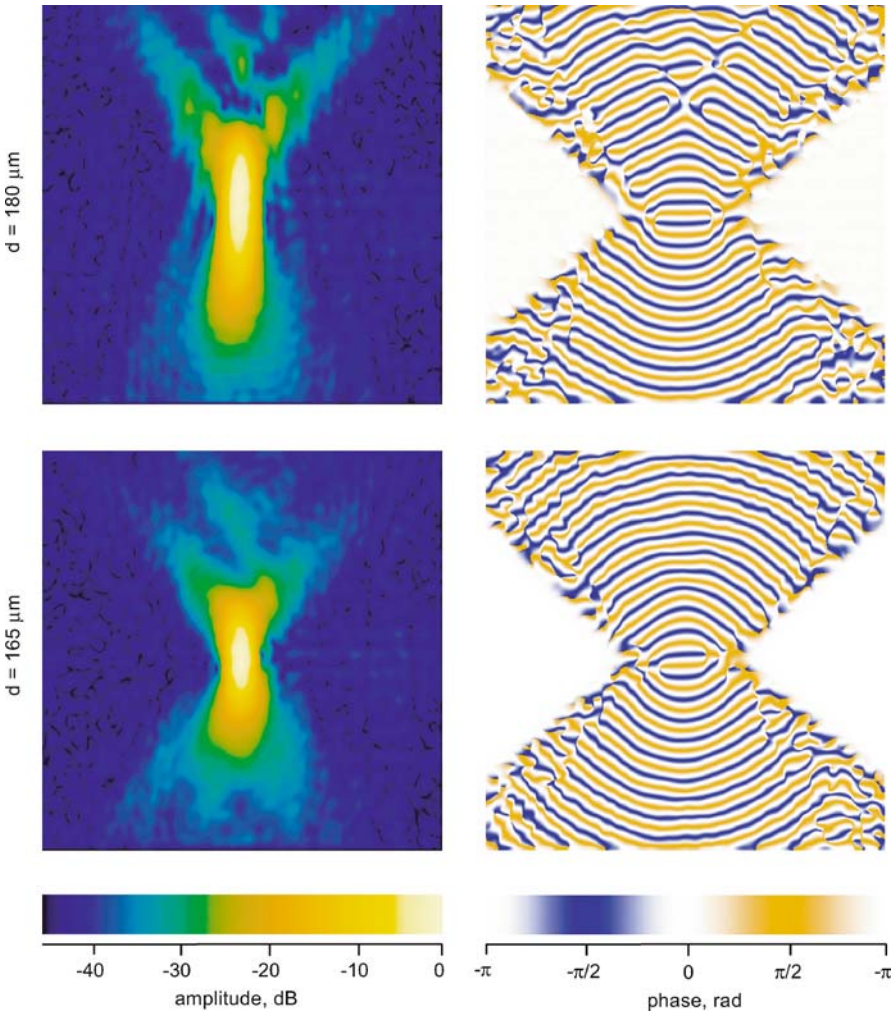


Fig. 2.3. The amplitude and phase of the effective PSF for 60× 1.2 NA water immersion lens with correction collar. Results for two different collar settings are shown. Image size in both x (horizontal) and z (vertical) are 5 µm

inal thickness 0.17 mm acted as point scatterers in this case. The coverslip was in turn mounted on a microscope slide and a gap between them was filled with immersion oil so as to eliminate reflection from the back surface of the coverslip. The size of the bead was carefully chosen experimentally in order to maximize the signal level without compromising the point-like behaviour. Indeed a control experiment using 40 nm beads yielded similar results to those presented below but with a vastly inferior signal-to-noise ratio.

In principle this apparatus is capable of producing full 3-D complex PSF data sets. It was found however that in most cases x–z cross-sections provided sufficient insight into the aberration properties of the lens without requiring too long acquisition times. Such results are shown in Fig. 2.3 for two settings of the correction collar. In order to emphasize the side lobe structure the magnitude of the PSF in displayed in dB with the peak value taken to be 0 dB. It can be seen that a collar setting of 0.165 mm gives a near-perfect form to the PSF. The axial side lobes are symmetric with respect to the focal plane and the phase fronts away from this plane quickly assume the expected spherical shape. On the other hand a small 10% deviation from the correct setting already has quite pronounced effect on the PSF in the bottom row of Fig. 2.3. The symmetry is broken, axial extent of the PSF has increased by about 30% and distinct phase singularities appeared on the phase fronts. Everything points towards a certain amount of uncompensated spherical aberration being present in the system. It is interesting to note that the phase map of the PSF seems to be more sensitive to the aberrations than the magnitude. This can be used as an early warning indicator of the trouble. It also shows the importance of measuring both the magnitude and phase of the PSF.

Although so far the measured PSF has been described in purely qualitative terms some useful quantitative information about the objective lens can also be extracted from this data. One parameter that can be readily verified is the objective's NA. Axial extent of the PSF is more sensitive to the NA than its lateral shape. Using the axial section of the PSF is therefore the preferred method to determine the NA. Besides, the interference fringes present in the z-scan provide a natural calibration scale for the distance in z. The actual measurement was obtained by finding the best fit to the curve in Fig. 2.4. A somewhat surprising result of this exercise was that the

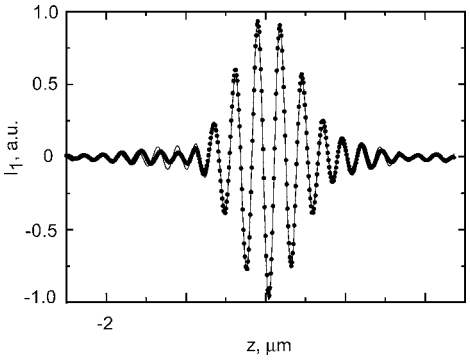


Fig. 2.4. Measured (*dots*) and calculated (*line*) amplitude axial responses for the same lens

best fit corresponded to NA of 1.15, rather than the nominal value of 1.2. This is not a coincidence: such discrepancies were found with other high NA objective lenses as well. The reason for this kind of behaviour will become clear in Sect. 2.4.4.

2.3 Chromatic Aberrations

Chromatic aberrations constitute another class of aberrations that can adversely affect the performance of any microscopy system. These aberrations are notoriously difficult to overcome in high NA objective lens design. The reason, at least in part, is relative uniformity of dispersion properties of common glasses used in objective lenses. Ingenious solutions have been found at an expense of dramatic increase of the number of lens elements – typically to more than a dozen in apochromats. Even then part of the correction may need to be carried out in the elements external to the objective.

Lateral and axial colour, as they are called by lens designers, are usually treated as separate chromatic aberrations. The former, which manifests itself as the wavelength-dependant magnification, is easy to spot in conventional microscopes as colouring of the edges of high contrast objects. Lateral chromatic aberration is also the more difficult to correct of the two. Traditionally this has been done by using the tube lens, or even the ocular, to offset the residual lateral colour of the lens. Some latest designs claim to have achieved full compensation within the objective lens itself, – claims to be treated with caution. In any case the correct testing procedure for the lateral colour should include at least a matched tube lens. The simplest test would probably be to repeat the experiments described in Sect. 2.2.2 for several wavelengths at the edge of the field of view and record the shift of the lateral position of the point image.

In confocal microscopy, where the signal is determined by the overlap of the effective excitation and detection PSFs, the loss of register between them should lead to reduction of signal towards the edge of the field of view. It has to be said, though, that in most confocal microscopes almost always only a small area around the optical axis is used for imaging, hence this apodization is hardly ever appreciable. Axial colour on the other hand is rarely an issue in conventional microscopy, but it can be of serious consequence for confocal microscopy, especially when large wavelength shifts are involved, such as in multiphoton or second and third harmonic microscopy. Mismatch in axial positions of excitation and detection PSFs can easily lead to degradation or even complete loss of signal even in the centre of the field of view. Below we describe a test setup which uses this sensitivity of confocal system to characterise axial chromatic aberration of high NA objective lenses.

2.3.1 Apparatus

Ideally one could conceive an apparatus similar to that in Fig. 2.2, whereby the laser is substituted with a broadband light source. One problem is immediately obvious: it is very difficult to couple any significant amount of power into a single-mode fibre

from a broadband light source, such as an arc lamp. Using multiple lasers provides only partial (and expensive) solution. Instead it was decided to substitute the point scatterer with a plane reflector. Scanning the reflector axially produces the confocal signal [6]:

$$I = \left[\frac{\sin u/2}{u/2}\right]^2 . \tag{2.11}$$

The maximum signal is detected when the plane reflector lies in the focal plane. This will change with the wavelength if chromatic aberration is present. Using mirror instead of a bead has another advantage: the resulting signal is one-dimensional, function of u only, and hence a dispersive element can be used to directly obtain 2-D spectral axial responses without the necessity of acquiring multiple datasets at different wavelengths.

The resulting apparatus, depicted in Fig. 2.5 and described in more detail in [7], is again based around a fibre-optic confocal microscope setup, but the interferometer part is now discarded. Instead, a monochromator prism made of SF4 glass is introduced to provide the spectral spread in the horizontal direction (i.e. in the image plane). Scanning in the vertical direction was introduced by a galvo-mirror moving in synchronism with the mirror in the focal region of the objective lens. The resulting 2-D information is captured by a cooled 16-bit slow-scan CCD (charge coupled device) camera. A small-arc Xe lamp is used as a light source providing approximately 0.2 μW of broadband visible radiation in a single-mode fibre. This is sufficient to produce a spectral snapshot of a lens in about 10 s.

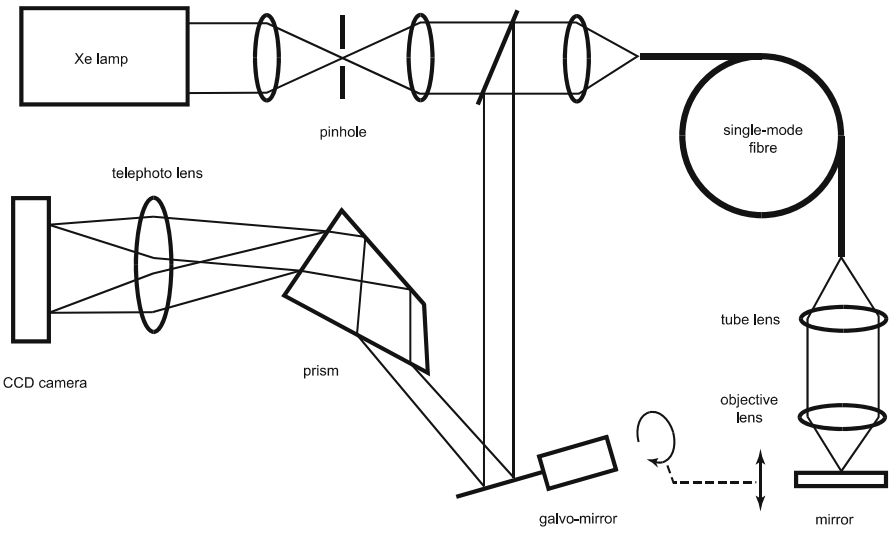


Fig. 2.5. Experimental setup for measuring axial chromatic aberration of objective lenses

2.3.2 Axial Shift

Typical results obtained by the chromatic aberration measurement apparatus are shown in Fig. 2.6. Since the raw images are not necessarily linear either in z or λ a form of calibration procedure in both coordinates is required. To achieve this the arc lamp light source was temporarily replaced by a He-Ne and a multi-line Ar^+ lasers. This gave enough laser lines to perform linearisation in λ. As a bonus coherent laser radiation also gave rise to interference fringes in the axial response with the reflection from the fibre tip acting as a reference beam, just as in the setup of Sect. 2.2.1. When a usual high NA objective lens was substituted with a low NA version these fringes covered the whole range of the z scan and could be readily used to calibrate the axial coordinate. The traces shown in Fig. 2.6 have been normalized to unity at each individual wavelength. The presence of longitudinal chromatic aberration is clearly seen in both plots. Their shapes are characteristic for achromats in which the longitudinal colour is corrected for two wavelengths only. It is interesting also to note the change of the shape of the axial response with the wavelength. This is noticeable for 32× 0.5 NA plan-achromat but it becomes particularly dramatic for 50× 0.8 NA plan-achromat, Fig. 2.6b. Clearly the latter lens suffers from severe spherical aberration at wavelengths below 550 nm, which results in multiple secondary maxima at one side of the main peak of the axial response.

An interesting problem is posed by the tube lens in Fig. 2.5. This lens may well contribute to the longitudinal chromatic aberration of the microscope as a whole and therefore it is desirable to use here a proper microscope tube lens matched to the objective. In fact it transpired that in some cases the tube lens exhibited significantly larger axial colour than the objective itself. This is hardly surprising: the tube lens is typically a simple triplet and three elements are hardly enough to achieve any sophisticated colour correction.

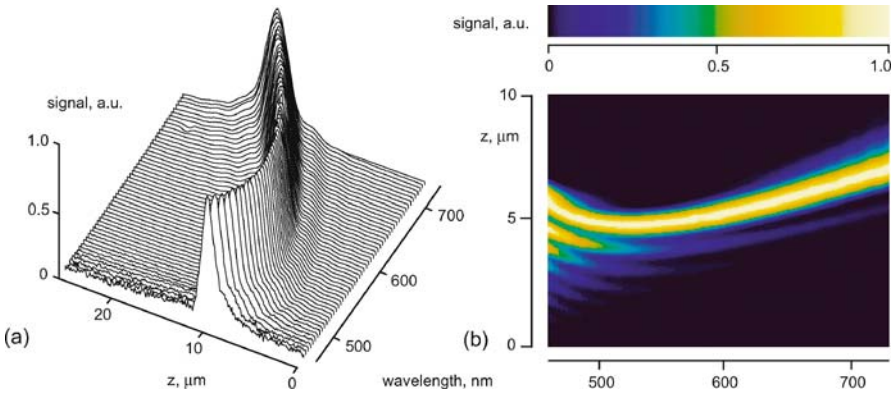


Fig. 2.6. Experimental results from (**a**) 32× 0.5 NA plan-achromat objective, displayed as 3-D plot, and (**b**) 50× 0.8 NA plan-achromat in pseudo-colour representation. Note the change in z scale

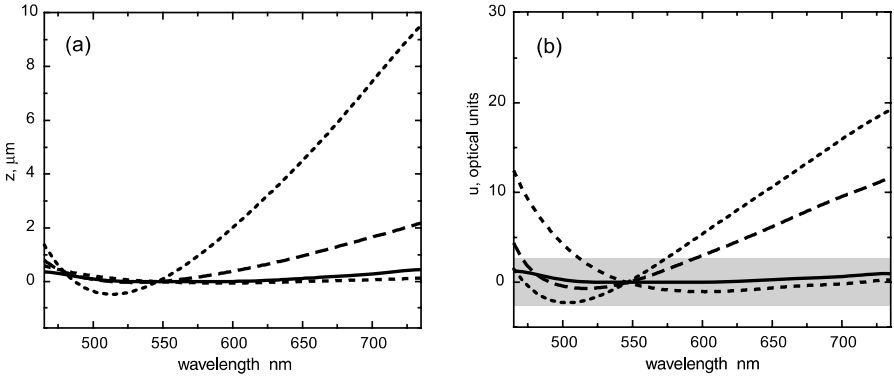


Fig. 2.7. Wavelength dependence of the axial response maxima represented in physical (**a**) and optical (**b**) axial coordinates. The four traces correspond to: 32× 0.5 NA plan-achromat (*dotted line*), 50× 0.8 NA plan-achromat (*long dashes*), 100× 1.4 NA plan-apochromat (*short dashes*), and the same lens stopped down to 0.7 NA (*solid line*)

In this experiment, however, the main task was to evaluate the properties of the objective lens itself. A different approach was therefore adopted. The same achromatic doublet (Melles Griot 01 LAO 079) collimating lens was used with all objective lenses. Because the chromatic aberrations of this lens are well documented in the company literature these effects could be easily removed from the final results presented in Fig. 2.7. Figure 2.7b presents the same data but in a form more suited to confocal microscopy whereby the chromatic shift is now expressed in optical units (2.3). The half width of the axial response to a plane mirror is given by 2.78 optical units at all wavelengths and for all NAs. This region is also shown in the figure. The zero in the axial shift is arbitrarily set to correspond to $\lambda = 546\,\text{nm}$ for all the objectives tested.

As could be expected these results show improvement in performance of apochromats over achromats. They also show that none of the tested objectives (and this includes many more not shown in Fig. 2.7 for fear of congestion) could meet the requirement of having a spectrally flat – to within the depth of field – axial behaviour over entire visible range. This was only possible to achieve by stopping down a 1.4 NA apochromat using a built-in aperture stop – the trick to be repeated several times again before this Chapter expires.

2.4 Pupil Function

Pupil function is the distribution of the phase and amplitude across the pupil plane when the lens is illuminated by a perfect spherical wave from the object side. It is related in scalar approximation to the PSF via a Fourier-like relationship (2.4). It would appear therefore that they both carry the same information and therefore the choice between them should be a simple matter of convenience. Reality is a bit

more complicated than that. Calculating the Pupil function from the PSF is an ill-posed problem and therefore very sensitive to noise. Measurements of the Pupil function provide direct and quantitative information about the aberrations of the lens – information that can only be inferred from the PSF measurements.

The trouble with mapping the Pupil function is that a source of a perfect spherical wave is required. Such thing does not exist but, fortunately, the dipole radiation approximates such wave rather well, at least as far as phase is concerned. The approach described in this Section is based on using small side-illuminated scatterers as sources of spherical waves. The actual Pupil function measurement is then performed in a phase-shifting Mach–Zender interferometer in a rather traditional fashion.

2.4.1 Phase-Shifting Interferometry

The experimental setup depicted in Fig. 2.8 comprised a frequency doubled Nd^+:YAG laser which illuminated a collection of 20 nm diameter gold beads deposited on the surface of high refractive index glass prism acting as dipole scatterers. Because the laser light suffered total internal reflection at the surface of the prism no direct illumination could enter the objective lens. The gold scatterers converted the evanescent field into the radiating spherical waves which were collected by the lens and converted into plane waves. These waves were then superimposed on a collimated reference wave. A 4-f lens system was then used to image the pupil plane of the lens onto a CCD camera. A pinhole in the middle of this projection system served to select a signal from a single scatterer. The size of this pinhole had to be carefully controlled as not to introduce artifacts and degrade resolution in the image of the Pupil function. A second CCD camera was employed to measure the PSF at the same time.

One of the mirrors in the reference arm of the interferometer was mounted on a piezo-electric drive and moved in synchronism with the CCD frame rate to produce

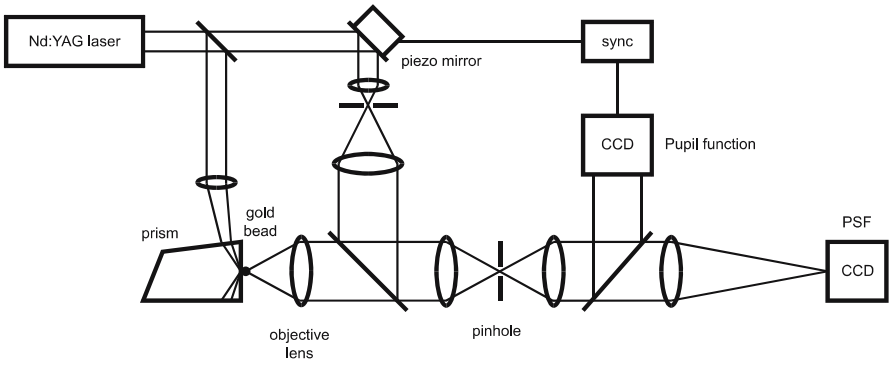


Fig. 2.8. Phase shifting Mach–Zehnder interferometer used for the Pupil function measurements

successive interferograms of the pupil plane shifted by $2\pi/3$ rad

$$I_l \sim \left| r + P(\rho,\theta)\exp\left(\mathrm{i}\frac{2\pi l}{3}\right)\right|^2, \qquad l = 0,1,2. \tag{2.12}$$

Using these three measurements the phase component of the Pupil function was then calculated as

$$\arg[P(\rho,\theta)] = \arctan\frac{\sqrt{3}(I_1 - I_2)}{I_1 + I_2 - 2I_0}. \tag{2.13}$$

The lens together with the prism were mounted on a pivoting stage which could rotate the whole assembly around the axis aligned to an approximate location of the pupil plane. Thus the off-axis as well as on-axis measurements of the Pupil function could be obtained. A set of such measurements is presented in Fig. 2.9 which clearly demonstrates how the performance of the lens degrades towards the edge of the field of view. Not only appreciable astigmatism and coma are introduced, but also vignetting becomes apparent. The presence of vignetting would be very difficult to deduce from the measurement of the PSF alone as it could be easily mistaken for astigmatism. Not that such vignetting is necessarily an indication that something is wrong with the lens: it can well be deliberate technique employed by the lens designer to block off the most aberrated part of the pupil.

2.4.2 Zernike Polynomial Fit

Traditionally the phase aberrations of the Pupil functions are described quantitatively by expanding them using a Zernike circle polynomial set:

$$\arg[P(\rho,\theta)] = \sum_{i=1}^{\infty} a_i Z_i(\rho,\theta), \tag{2.14}$$

where a_i are aberration coefficients for corresponding Zernike polynomials $Z_i(\rho,\theta)$. Significant variations between different modifications of Zernike polynomials exist. In this work a set from [9] was used. The first 22 members of this set are listed in Table 2.1 together with their common names. This list can be further extended. In practice, however, expansion the beyond second order spherical aberration is not very reliable due to experimental errors and noise in the measured image of $P(\rho,\theta)$.

The determination of the expansion coefficients a_i should, in principle, be a simple procedure, given the Zernike polynomials are orthonormal. Multiplying the measured Pupil function by a selected polynomial and integrating over the whole pupil area should directly yield the corresponding aberration coefficient. The real life is a bit more complicated, especially when processing the off-axis data, such as shown in Fig. 2.9. One obstacle is vignetting: the standard Zernike set is no longer orthonormal over a non-circular pupil. Even without vignetting the 2π phase ambiguity poses a problem. Before the expansion procedure can be applied the

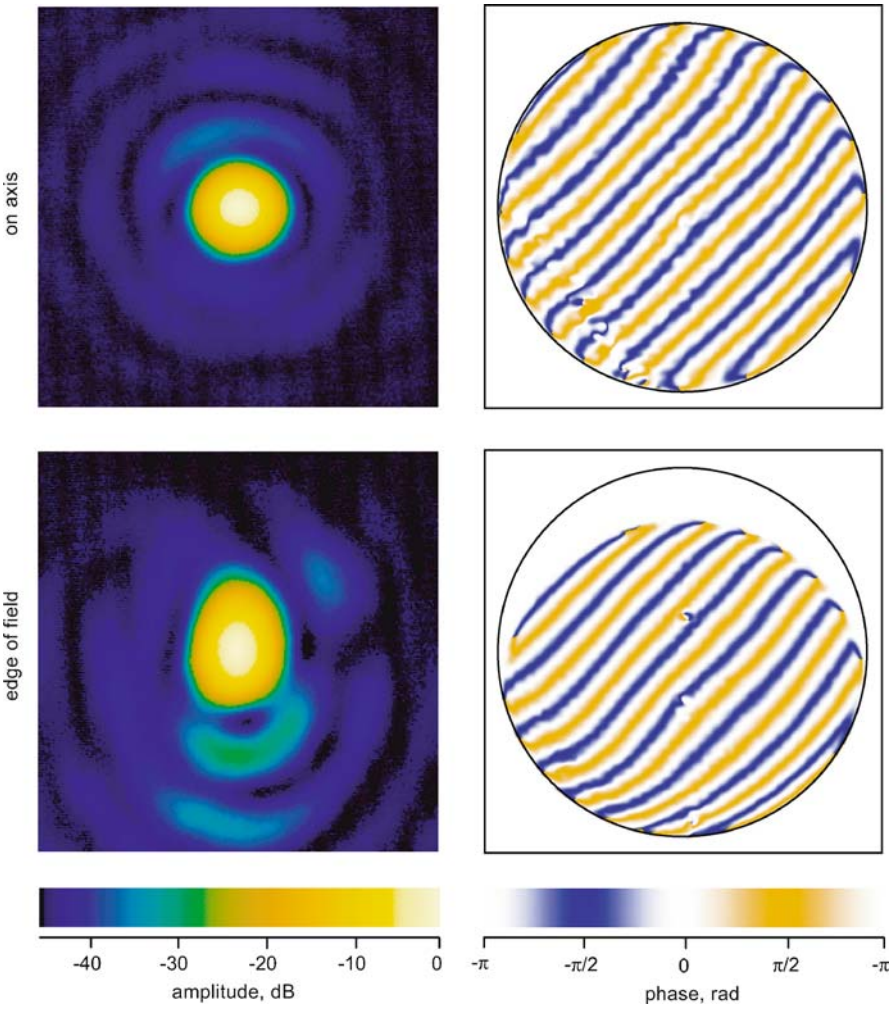


Fig. 2.9. Typical results obtained with 40× 1.2 NA Apochromat, PSF on the left, phase maps of corresponding pupil functions on the right. Performance both on axis, top, and at the edge of the field of view, bottom, is shown

phase of the Pupil function has to be unwrapped – not necessarily a trivial procedure.

An entirely different expansion technique was developed to overcome these difficulties. This technique is based on simulated iterative wavefront correction routine, originally conceived to be used in adaptive optics applications together with a diffractive optics wavefront sensor [10]. The essence of the method is that small simulated amounts of individual Zernike aberrations are applied in turns to the measured Pupil function. After each variation the in-focus PSF is calculated and

Table 2.1. Orthonormal Zernike circle polynomials

i	n	m	$Z_i(r,\theta)$	*Aberration term*
1	1	0	1	piston
2	1	1	$2r\cos\theta$	tilt
3	1	−1	$2r\sin\theta$	tilt
4	2	0	$\sqrt{3}(2r^2-1)$	defocus
5	2	2	$2\sqrt{3}r^2\cos 2\theta$	astigmatism
6	2	−2	$2\sqrt{3}r^2\sin 2\theta$	astigmatism
7	3	1	$2\sqrt{2}(3r^3-2r)\cos\theta$	coma
8	3	−1	$2\sqrt{2}(3r^3-2r)\sin\theta$	coma
9	3	3	$2\sqrt{2}r^3\cos 3\theta$	trefoil
10	3	−3	$2\sqrt{2}r^3\sin 3\theta$	trefoil
11	4	0	$\sqrt{5}(6r^4-6r^2+1)$	primary spherical
12	4	2	$\sqrt{10}(4r^4-3r^2)\cos 2\theta$	
13	4	−2	$\sqrt{10}(4r^4-3r^2)\sin 2\theta$	
14	4	4	$\sqrt{10}r^4\cos 4\theta$	
15	4	−4	$\sqrt{10}r^4\sin 4\theta$	
16	5	1	$2\sqrt{3}(10r^5-12r^3+3r)\cos\theta$	
17	5	−1	$2\sqrt{3}(10r^5-12r^3+3r)\sin\theta$	
18	5	3	$2\sqrt{3}(5r^5-4r^3)\cos 3\theta$	
19	5	−3	$2\sqrt{3}(5r^5-4r^3)\sin 3\theta$	
20	5	5	$6\sqrt{3}r^5\cos 5\theta$	
21	5	−5	$6\sqrt{3}r^5\sin 5\theta$	
22	6	0	$\sqrt{7}(20r^6-30r^4+12r^2-1)$	secondary spherical

the whole process iteratively repeated until the Strehl ratio is maximized. The final magnitudes of the Zernike terms are then taken to be (with opposite signs) the values of the Zernike expansion coefficients of the experimentally measured aberrated Pupil function. This procedure is reasonably fast and sufficiently robust, provided that the initial circular aperture can still be restored from the vignetted pupil.

The power of this technique is demonstrated in Fig. 2.10 where a 40× 1.2 NA water immersion lens was investigated at three different settings of the correction collar. As expected, adjusting the collar mainly changes the primary and secondary spherical aberration terms. Variations in other terms are negligible. The optimum compensation is achieved close to $d = 0.15$ mm setting, were small amounts of both aberrations with opposite signs cancel each other. The usefulness of the Zernike

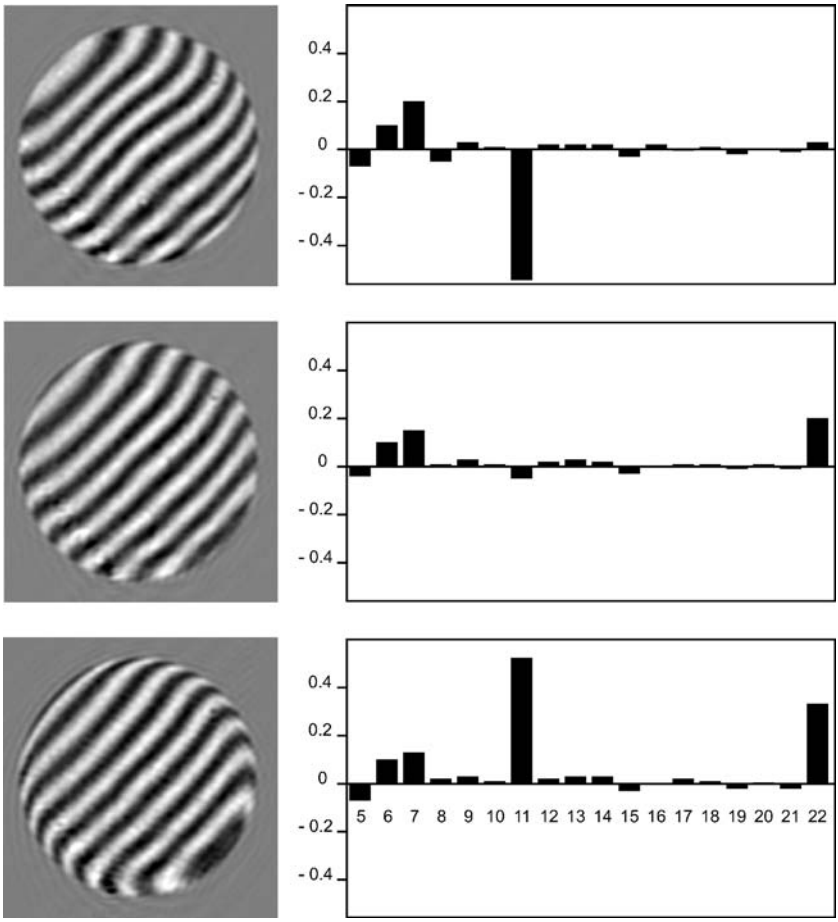


Fig. 2.10. Variations in wavefront aberration function expressed via Zernike modes when correction collar of a water immersion lens is adjusted

expansion is further demonstrated by the fact that the main residual term in this case was the defocus, which, although not an aberration itself, could be easily mistaken upon visual inspection of the interference pattern for a spherical aberration.

2.4.3 Restoration of a 3-D Point Spread Function

Nowhere the power of the Pupil function approach to the objective lens characterization is more apparent than in cases when full three-dimensional shape of the PSF needs to be determined. Such need may arise, for instance, when using deconvolution techniques to process images obtained with a confocal microscope.

As is clear from (2.4) not only an in-focus PSF can be calculated from a measured Pupil function, but the same can be done for any amount of defocus, by choos-

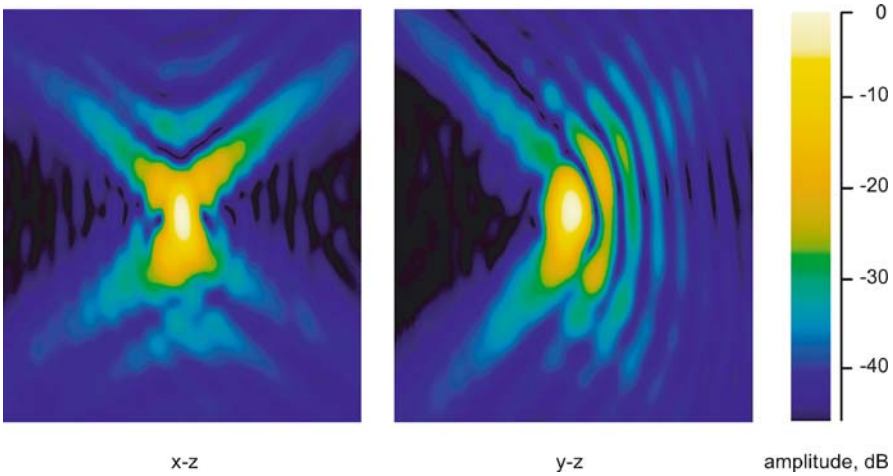


Fig. 2.11. Three-dimensional PSF restored from pupil function data, shown here via two meridional sections

ing an appropriate value for the axial coordinate u. Repeating the process at regular steps in u yields a set of through-focus slices of the PSF. These can then be used to construct a 3-D image of the PSF much in the same manner as 3-D images are obtained in a confocal microscope. Compared to the direct measurement using a point scatterer, vis Sect. 2.2.2, advantages of this approach are clear. A single measurement of the Pupil function is sufficient and no scanning of the bead in three dimensions is required. Consequently, exposures per image pixel can be much longer. As a result this method provides much improved signal-to-noise ratio in the final rendering of the PSF allowing even faintest sidelobes to be examined.

Obviously, presenting a complete 3-D image on a flat page is always going to be a problem but, as Fig. 2.11 shows, even just two meridional cross sections of a 3-D PSF provide infinitely more information than a plain 2-D in-focus section of the same PSF at the bottom of Fig. 2.9. Thus, for instance, the $y - z$ section clearly shows that the dominant aberration for this particular off-axis position is coma. Comparing the two sections is also possible to note different convergence angles for the wavefronts in two directions – a direct consequence of vignetting.

2.4.4 Empty Aperture

Testing objective lenses with highest NAs (1.4 for oil, 1.2 for water immersion) one peculiar aberration pattern is frequently encountered. As shown in Fig. 2.12, the lens is well corrected over up to 90-95% of the aperture radius, but after that we see a runaway phase variation right to the edge. Speaking in Zernike terms residual spherical aberration components of very high order are observed. Because of this high-order it appears unlikely that the aberrations are caused by improper immersion

fluid or some other trivial reasons. These would manifest themselves via low-order spherical as well. More realistically, this is a design flaw of the lens.

The portion of the lens affected by this feature varies from few to about 10 percent. For the lens in Fig. 2.12 the line delimiting the effective aperture was somewhat arbitrarily drawn at NA=1.3. What is undeniable is that the effect is not negligible. In all likelihood this form of aberration is the reason for the somewhat mysterious phenomenon when a high NA lens exhibits a PSF, which is perfect in all respects except for a somewhat reduced NA. This was the case in Sect. 2.2.2 and also described by other researchers [11].

It is quite clear that competitive pressures push the lens designers towards the boundary (and sometimes beyond the boundary) of the technical possibilities of the day. Few years ago no microscope lens manufacturer could be seen without a 1.4 NA oil immersion lens when the Joneses next door were making one. The plot is likely to be repeated with the newly emerging 1.45 NA lenses. It is also true that a hefty premium is charged by the manufacturers for the last few tenths in the NA. It is quite possible that in many cases this is a very expensive *empty* aperture, which, although physically present, does not contribute to the resolving power of the lens.

This discussion may seem to be slightly misguided. Indeed, many people buy high NA lenses not because of their ultimate resolution, but because of their light gathering efficiency in fluorescence microscopy. This property is approximately proportional to NA^2 and therefore high NA lenses produce much brighter, higher contrast images. At the first glance it may seem that the aberrated edge of the pupil will not affect this efficiency and hence buying a high NA lens, however aberrated, still makes sense. Unfortunately, this is not true. Because the phase variation at the edge is so rapid, the photons passing through it reach the image plane very far from the optical axis. They do not contribute to the main peak of the diffraction spot, instead they form distant sidelobes. In terms of real life images it means that the brightness

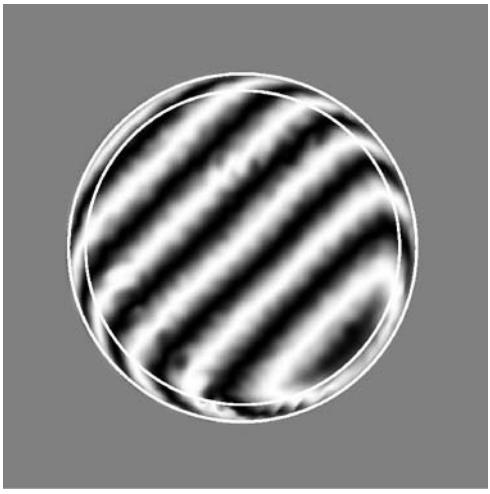


Fig. 2.12. Pupil function of 63 × 1.4 NA oil immersion lens with both nominal (outer ring) and effective working (inner ring) apertures indicated

of the background, and not the image itself, is increased. Paradoxically, blocking the outermost portion of the pupil would in this case *improve* the image contrast!

The last statement may well be generalized in the following way: in the end the only sure way of obtaining a near-perfect high NA objective lens is to acquire one with larger-than-required nominal NA and then stop it down. Incidentally, in certain situations this may be happening even without our deliberate intervention. Consider using an oil immersion lens on a water-based sample: no light at NA > 1.33 can penetrate the sample anyway and hence the outer aperture of the lens is effectively blocked.

2.5 *Esoterica*

In this Section a few more results obtained with the Pupil function evaluation apparatus are presented. These need not necessarily be of prime concern to most microscope users but might be of interest to connoisseurs and, indeed, could provide further insight into how modern objective lenses work.

2.5.1 Temperature Variations

Many of microscopists would recall seeing 23 °C on a bottle of immersion oil as a temperature at which the refractive index is specified, typically $n = 1.518$ at $\lambda = 546$ nm. But how important this "standard lab temperature" is to the performance of high NA lenses? In order to answer this question the Pupil function of a 100× 1.4 NA oil immersion lens was measured at a range of temperatures and the results were processed to obtain the variation of the primary aberration coefficients with the temperature.

The major factor in the degradation of the imaging qualities of the immersion lenses with temperature is the variation of the refractive index of the immersion oil, usually at the level of $\mathrm{d}n/\mathrm{d}T = 3\text{–}4 \times 10^{-4}$. The effect of this change is similar to that of introducing a layer of a refractive-index-mismatched material between the lens and the sample. The resulting aberrations are well understood, their exhaustive analysis can be found in [12]. In short, spherical aberrations of various orders will be generated; the relative weight of higher order terms rises dramatically with the NA of the lens. This is corroborated by the experimental data in Fig. 2.13 which show steady, almost linear variation in both primary and secondary spherical aberration terms with temperature. Less predictably the same plot also registers significant variations in the other first order aberrations: coma, astigmatism and trefoil. Because of their asymmetry these aberrations can not be explained by the oil refractive index changes. They are probably caused by small irregular movements of individual elements within the objective lens itself.

Strictly speaking aberrations caused by the refractive index changes in the immersion fluid should not be regarded as lens aberrations. In practice, however, the lenses are designed for a particular set of layers of well defined thicknesses and refractive indexes between the front of the lens and the specimen. Any change in

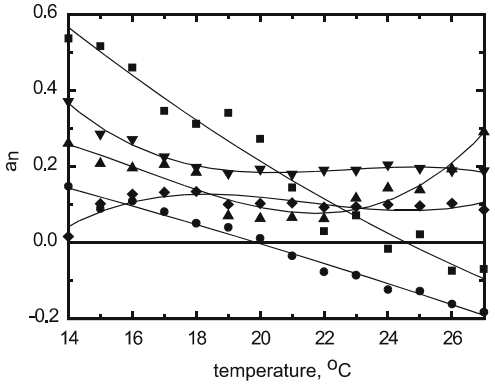


Fig. 2.13. Temperature variation of primary (■) and secondary (●) spherical aberrations as well as magnitudes of astigmatism (▲), coma (▼) and trefoil (♦). The latter three were defined as $a = \sqrt{a_s^2 + a_c^2}$, where a_s^2 and a_s^2 are the sin and cos components of the corresponding aberrations

these parameters upsets the fine optical phase balance in the system and results in aberrated PSF. This might be an argument why it makes sense to treat the immersion medium as being a part of the objective lens. The temperature effect depends dramatically of the type of the immersion fluid used. Water with $\mathrm{d}n/\mathrm{d}T \approx 8\times10^{-5}$ is far less sensitive; dry lenses, of course, are not affected at all. Large working distance lenses will be at a disadvantage too due to longer optical paths in the immersion fluid.

Closer analysis of Fig. 2.13 reveals that the aberrations are indeed minimized around 23 °C. where the combined primary and secondary spherical aberrations are close to their minimum. A small but noticeable hysteresis effect was also noted when after a few temperature cycles the aberration coefficients failed to return to their low initial values. It is tempting to draft this effect to the explanation of the fact that imaging properties of even the best of lenses always deteriorate with age – although accidental straining during experiments is still likely to remain the prevailing factor.

2.5.2 Apodization

So far the emphasis of this investigation was on measuring the phase aberrations. This is justified by the fact that the phase deviations from an ideal spherical wavefront have considerably more impact on the shape of the PSF than similar imperfections in amplitude. Nevertheless for completeness sake it might be interesting now to have a closer look at the apodization effects occurring in high NA lenses. Using dipole radiation as a probe offers unique advantages in this task. This is because the angular intensity distribution of the dipole radiation is well defined. Therefore any deviations from perfect lens behaviour should be easy to spot.

Let's assume that the polarization vector of a dipole situated in the focus is aligned in x direction. Angular intensity distributions in the x-z and y-z (i.e. meridional and equatorial) planes will be given by, respectively, $I_x \sim \cos^2\alpha$ and I_y = const. Due to purely geometric reasons these distributions will change when

light propagates to the pupil plane even if the lens is perfect. With reference to Fig. 2.1 and to the sine condition $d = nf \sin\alpha$ it is not too difficult to show that an extra factor of $\sec\alpha$ is has to be introduced when going from the object to the pupil side of the objective in order to satisfy the energy conservation law. This factor was well known since the early days of high NA lens theory [3]. Intensity distributions in the pupil plane therefore should therefore look like $I_x \sim \cos\alpha$ and $I_y \sim \sec\alpha$ or, with the help of sine condition:

$$I_x \sim \frac{nf}{\sqrt{(nf)^2 - d^2}}, \qquad I_y \sim \frac{\sqrt{(nf)^2 - d^2}}{nf}. \tag{2.15}$$

An experiment to measure these distributions was carried out on the setup of Fig. 2.8 by simply blocking the reference beam and capturing the pupil intensity image alone. To produce the results shown in Fig. 2.14 images of 8 individual scatterers were acquired and averaged in the computer. Intensity distributions in the two principal planes were then extracted. They follow the theoretical predictions rather well up to about half of the pupil radius. After that apodization is apparent which increases gradually reaching about 30–50% towards the edge of the pupil.

The cause of this apodization in all likelihood are the Fresnel losses in the elements of the objective lens. Broadband antireflection coatings applied to these elements are less effective at higher incidence angles that the high aperture rays are certain to encounter. Because of the nature of these losses they are likely to be very individual for each particular type of objective lens. It is also worth noting slight polarization dependence of the losses, which contributes to polarization effects described in the following section.

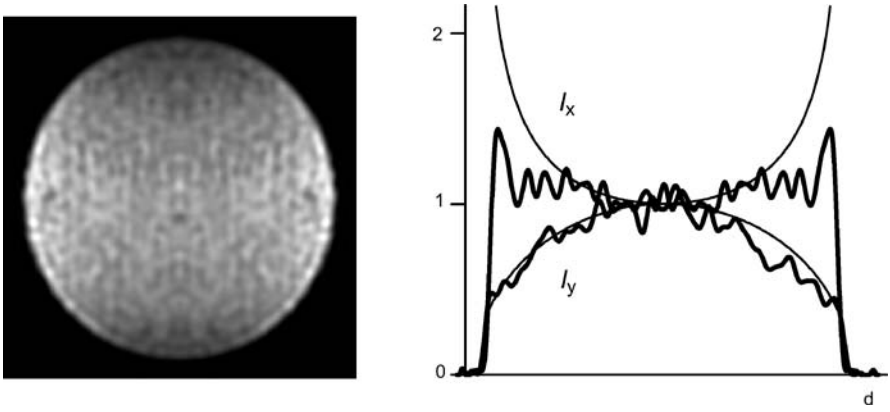


Fig. 2.14. Apodization effects on an image of a dipole viewed at the pupil plane of a 100× 1.4 NA lens. Theoretical predictions (*thin line*) according to (2.15)

2.5.3 Polarization Effects

Polarization effects encountered when imaging a dipole with a high NA lens have been covered elsewhere [13] (For a general discussion also see Chapter 1). For the purposes of this investigation they are interesting inasmuch as the imperfections of the lens contribute to these effects. When an image of a small scatter is viewed between the crossed polarizers characteristic "clover leaf" pattern emerges. An image of the pupil plane, Fig. 2.15, is particularly telling. It shows, that the only rays that travel very close to the edge of the pupil pass through the polarizer. This happens because the polarization of light in these four segments is slightly rotated from its original x direction. The reason for this rotation is twofold. First of all the radiation of a dipole is linearly polarized in the meridional plane, which can only be approximated by a uniformly polarized plane wave for small angles α. When this spherical wavefront is straightened by the lens and projected into the pupil plane only light propagating in x-z and y-z possess perfect x polarization, the remainder is rotated to some degree. The degree of rotation increases dramatically with higher aperture angles. This effect is fundamentally geometrical; its close cousin is a well know problem of how to comb a hairy ball.

The second reason for the polarization rotation is more prosaic: Fresnel losses. These tend to be higher for s than for p polarized light. For a beam travelling in a plane bisecting the x-z and y-z planes overall tendency would be to rotate the polarization towards the radial direction. Hence this effect seems to work in the opposite direction to that caused by the geometric factor, which favours the azimuthal direction.

2.6 Conclusion

The range of various experimental setups and techniques dedicated to the characterization of the high NA objective lenses could be continued. For instance the lateral chromatic aberration has not been considered so far. One has to be practical, however, and draw a line at some point.

From the tests described in this Chapter the measurement of the Pupil function provides the most detailed and insightful information about the capabilities of a particular lens. In many cases two or three Pupil functions measured across the field of view and, perhaps, tabulated in Zernike coefficient form would be more than sufficient to predict the lens performance in most practical situations. It is disheartening to think of how much wasted time, frustration and misunderstandings could be avoided if such information were to be supplied with the objective lenses by their manufacturers.

My overall conclusion is that the vast majority of currently designed objective lenses perform really well. Any imperfections visible in a microscope image are far more likely to be a result of a sloppy practice (for example tilted coverslip, incorrect immersion fluid etc.) than a fault of the lens itself. That said, cutting-edge designs are always going to be a problem and the very highest NA lenses

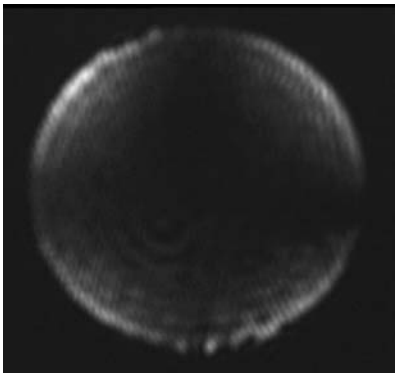


Fig. 2.15. Polarization effects on an image of a dipole viewed at the pupil plane

should be approached with caution. It is also worth pointing out that elements of the microscope other than the objective lens may also be a factor in imaging quality. The tube lens is one such element of particular concern. Having evolved little over the last years this lens is typically just a simple triplet with too few elements to achieve the aberration correction on par with that of an all-singing all-dancing objective lens. This situation is further exacerbated by the advent of a new breed of low-magnification high NA objectives with their enormous back apertures.

Acknowledgements

I would like to thank my colleague M.A.A. Neil for numerous contributions to this work, in particular the implementation of the Zernike polynomial fit routine and processing of the relevant data.

References

1. D. Malacara: *Optical shop testing* (John Wiley & Sons, New York 1992)
2. M. Born, E. Wolf: *Principles of optics* (Pergamon Press, Oxford 1998)
3. B. Richards, E. Wolf: P. Roy. Soc. Lond. A **253**, 358 (1959)
4. T. Wilson, R. Juškaitis, N.P. Rea, D.K. Hamilton: Opt. Commun. **110**, 1 (1994)
5. R. Juškaitis, T. Wilson: J. Microsc. – Oxford **189**, 8 (1997)
6. T. Wilson, C.J.R. Sheppard: *Theory and practice of scanning optical microscopy* (Academic Press, London 1984)
7. R. Juškaitis, T. Wilson: J. Microsc. – Oxford **195**, 17 (1999)
8. R. Juškaitis, M.A.A. Neil, T. Wilson: 'Characterizing high quality microscope objectives: a new approach', In: *Proceedings of SPIE, San Jose Jan. 28–Feb. 2, 1999*, ed. by D. Cabib, C.J. Cogswell, J.-A. Conchello, J.M. Lerner, T. Wilson (SPIE, Washington 1999) pp. 140–145
9. V.N. Mahajan: Appl. Optics **33**, 8121 (1994)
10. M.A.A. Neil, M.J. Booth, T. Wilson: J. Opt. Soc. Am. A **17**, 1098 (2000)
11. S.W. Hell, P.E. Hanninen, A. Kuusisto et al.: Opt. Commun. **117**, 20 (1995)
12. P. Török, P. Varga, G. Németh: J. Opt. Soc. Am. A **12**, 2660 (1995)
13. T. Wilson, R. Juškaitis, P.D. Higdon: Opt. Commun. **141**, 298 (1997)

3 Diffractive Optical Lenses in Imaging Systems – High-Resolution Microscopy and Diffractive Solid Immersion Systems

R. Brunner and H.J. Dobschal

3.1 Introduction

Mirrors and refraction based lenses are the most dominating basic design elements of imaging optical systems. This fact can be observed over a broad spectrum of optical instruments ranging from low-cost devices, e.g. optical pick-ups in data storage systems, up to high-end systems, e.g. space telescopes or optics for photolithography containing 30 lenses and more. In all these optical systems, the demands on performance, volume and costs are increasing dramatically from one product generation to the next. The search for alternative design approaches becomes more and more important. Here, the combination of diffractive and refractive components to form hybrid systems opens up an enormous potential to overcome the present limitations.

Going back in history, diffractive optical elements (DOEs) started their career as dispersive gratings and continue as the basic elements for most spectroscopic instruments. Due to the lack of suitable refractive lenses, DOEs, e.g. in the form of Fresnel Zone Plates, meanwhile have established themselves in x-ray microscopy [1]. Also in the visible spectrum of light the hybrid concept has been introduced to sophisticated systems for various applications. Especially eyepieces have to be mentioned, in which the hybrid approach allows to overcome limitations in eye relief, exit pupil, and field of view [2–6]. Closely related to them are head-mounted displays (HMDs), a micro-display application near to the eye, with challenging issues in high quality ergonomics combined with high optical performance in a minimal volume [7–9].

The dependence of optical characteristics of DOEs with respect to wavelength differs fundamentally from that of refractive elements. In particular, the dispersion of an imaging DOE is negative and much stronger compared to conventional lenses. This property allows a beneficial combination of a refractive lens and a diffractive surface to realize compact achromats. The second fundamental issue in broadband applications of DOEs is their diffraction efficiency. A single layer DOE shows a maximum efficiency only at the design wavelength and decays significantly in the adjacent blue and red spectral range. A very beneficial approach to guarantee a high diffraction efficiency over a broad spectral range was the introduction of multi-layer DOEs. This approach also allows the realization of high-quality camera lenses over the whole visible spectrum [10].

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

The following contribution starts with a condensed survey of some basic aspects of DOEs in imaging systems including a short discussion of angle-dependent diffraction efficiency and secondary spectrum. The main part of the text will focus on hybrid diffractive-refractive approaches in microscopy. Here a promising application is deep-UV defect inspection in the semiconductor industry. The impact of diverse critical specifications on the lens design, the realization of hybrid lens systems, and experimental results are presented. A different application area of the hybrid concept in the field of high-resolution imaging concerns diffraction based solid immersion lens (SIL) systems.

3.2 Basics

3.2.1 Fundamentals

In contrast to conventional optical elements that form images with either reflection or refraction, DOEs make use of diffraction and interference. Generally, these elements are based on a surface-relief structure of concentric rings with a variation of the spatial frequency. A simple approach to describe the basic principles of the focusing properties of an imaging DOE allows the analysis of the Fresnel Zone Plate (FZP) (see Fig. 3.1). The FZP is characterized by an axially symmetrical and concentric zone system of alternating opaque and transparent rings. Consider a monochromatic plane wave which is incident onto the FZP. In accordance with the Huygens-Fresnel principle, a common focal point is guaranteed, when the optical path length from two adjacent zones differs by an integer multiple of the wavelength λ. The radius, r_j, of the j-th ring and the focal length, f, are related by:

$$r_j^2 = 2j\lambda f + (j\lambda)^2 \tag{3.1}$$

This relation implies that the distance between two neighboring zones decreases with increasing radius. In contrast to a conventional refractive lens, a FZP causes multiple foci at distances of an integral fraction of the basic focal length. In the case of a transmission grating of equal sizes of bars and spaces only the odd orders occur ($f/3$, $f/5$, ...). If the alternating opaque and transparent structure is replaced by a phase-only diffractive element the light intensity obtained at the focus is increased dramatically

An additional difference between conventional spherical refractive lenses and DOE lenses is, that for the latter a zonal design following Eq. (3.1) shows no spherical aberration.

To design or analyze complex hybrid lens systems comprising several refractive components and imaging DOEs, a ray-trace based optical design software is most often used. The DOEs are treated by using the grating equation on a local basis rather than the refraction law. The action of a diffractive lens can be described by a radial phase function so that aberrations can be treated just as they are in refractive lenses.

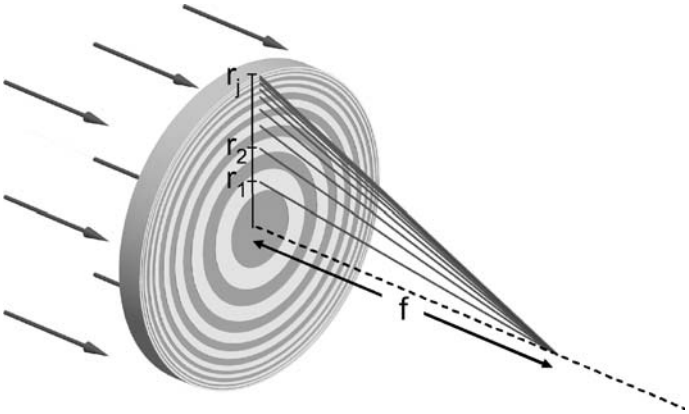


Fig. 3.1. The concentric ring system of a Fresnel zone plate lens with plane wave illumination, showing the convergent 1st order diffraction. The optical path length of the light deflected from adjacent zones towards a common focal point differ by an integral multiple of the design wavelength

3.2.2 Dispersion – Achromatization – Apochromatization

An essential difference between DOEs and refractive lenses is their chromatic behavior. The performance of refractive lenses depends on material dispersion. Most optical materials, e.g. inorganic glasses, crystals or polymers, show a normal dispersion, i.e. the refraction index $n(\lambda)$ decreases with wavelength. Quantitatively this can be expressed by the Abbe-number $\nu_{e(ref)}$, where the subscript defines the reference wavelength ($\lambda_e = 546.1$ nm, $\lambda_C = 656.3$ nm, $\lambda_F = 486.1$ nm).

$$\nu_{e(ref)} = \frac{n(\lambda_e) - 1}{n(\lambda_F) - n(\lambda_C)} \tag{3.2}$$

The Abbe-number of inorganic glasses ranges from 20 for very dense flint glasses up to 90 for fluor-crown glasses.

In contrast, the most prominent characteristic of an imaging DOE is the large negative dispersion. This can be expressed by an effective Abbe-number $\nu_{e(diff)}$ for a diffractive lens, which is the ratio of the wavelengths:

$$\nu_{e(diff)} = \frac{\lambda_e}{\lambda_F - \lambda_C} = -3.21 \tag{3.3}$$

This behavior makes DOEs well-suited for the realization of achromatic imaging systems. A conventional all-refractive achromatic doublet combines lenses of both positive and negative optical power. The opposition means that extra thickness is required to achieve a given power. In contrast, the hybrid concept allows the combination of a diffractive and a refractive element with optical power equal in sign, which results in a small and compact design.

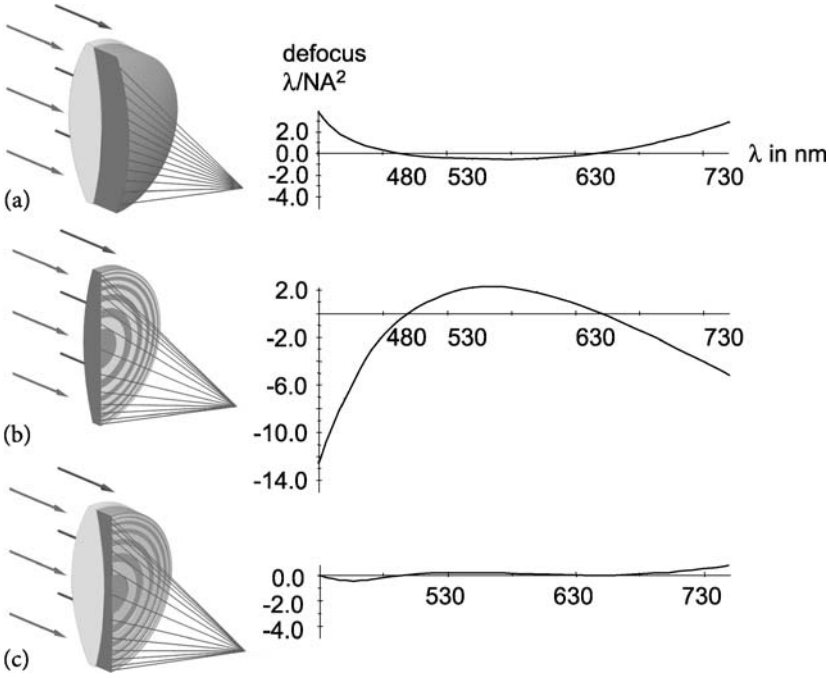


Fig. 3.2. Variation of focal distance as a function of wavelength for different lens combinations – (**a**): classical all-refractive achromat, (**b**): monolithic hybrid achromat, (**c**) diffractive-refractive apochromat

Achromatization with respect to two wavelengths does not allow a complete removal of the color error. The remaining chromatic aberration, which is expressed by the focus-shift in dependence on the wavelength, $f(\lambda)$, is known as the secondary spectrum.

Figure 3.2 shows the secondary spectrum for different design examples with a focal length of 10 mm. For the calculations we choose plastic optical components made from polymethyl methacrylate (PMMA) and polycarbonate (PC). Polymers have advantages in weight and producibility of high volume and low price systems, but also show limitations in the optical properties.

In Fig. 3.2a $f(\lambda)$ is shown for a conventional dioptric combination of a positive (PMMA) and a negative (PC) lens. The function shows a concave curvature with intersections of the wavelength axis at the reference wavelengths of 488 nm and 633 nm. One can also design a single PMMA plano-convex lens with a DOE on the plane surface (Fig. 3.2b) that is achromatic at the two reference wavelengths. The hybrid component, however, exhibits a convex curvature with a more significant deviation in both the short and the long wavelength range (see also [11]).

A further reduction of the secondary spectrum – apochromatic correction – could be achieved in the conventional approach by the introduction of a three-lens

combination, whereas in general one component has to be based on a material with special optical properties. Due to the limitations of plastic materials, designing an apochromatic system may require more than three lenses, which increases size and cost.

An alternative apochromatic design, without the need for special materials, is offered by a hybrid combination of a simple positive refractive lens (PMMA) and a hybrid, plano-concave element (PC) with the DOE-structure on the plane surface (Fig. 3.2c). The focus position varies minimally from blue to red.

3.2.3 Diffraction Efficiency

In a diffractive lens, the distribution of the local periods and the grating equation, determine the direction of the deflected light. On the other hand, the diffraction efficiency of the various orders depends on the nature of the periodic structure, e.g. geometry, material selection, and the wavelength, polarization and incidence angle of the light.

Following the scalar theory, which is valid for grating periods much larger than the wavelength, the sawtooth profile of a blazed diffraction grating possesses the most favorable profile geometry. The efficiency η_m in the desired order, m, for an ideal blazed profile is expressed by [12,13]:

$$\eta_m(\lambda) = \mathrm{sinc}^2\left\{\frac{\lambda_0}{\lambda}\frac{n(\lambda)-1}{n(\lambda_0)-1} - m\right\} \tag{3.4}$$

where λ_0 is the design wavelength for which the efficiency is 100%, and $\mathrm{sinc}(x) = \sin(\pi x)/(\pi x)$. With the change of the refractive index as a function of the wavelength, the efficiency has a maximum only at the design wavelength. At other wavelengths, the light is partly diffracted into other orders. This efficiency behavior has a dominant impact on the design and the development process of diffractive or hybrid optical systems. On the one hand, due to the varying efficiency, the images are not generated in true colors. Also, the light transmitted into the 0th order or into the higher, non-design orders may add to the background illumination, reducing the image contrast. Dependent on the individual application there are several opportunities to treat this characteristic:

- Systems acting in a limited wavelength range, especially laser-systems, don't suffer from the addressed drawback.
- Optical systems which are applied directly to the human eye, can be adapted to the visual sensitivity curves of rods and cones. The strong wavelength dependence of the visual system causes only a weak influence on the perception in the red and blue spectral range, so the aberrations are less noticeable [14].
- There is also the possibility to direct the disturbing diffraction orders to light traps or to the housing of the system, so that only a limited part of the unwanted light will reach the detection plane.

- In several applications, the wavelength selectivity of the DOEs is a necessary precondition for the mode of operation [15,16]. In such systems for example, the DOE profile is chosen so that each wavelength range is using a particular and different diffraction order [17,18].
- By the use of selected multi-layer relief configurations, it's also possible to achromatize the diffraction efficiency of a DOE and guarantee a nearly 100% efficiency over a broad spectral bandwidth. These elements take advantage of the dispersion characteristics and/or the introduction of additional interfaces into the diffractive structure [19].

Nearly index-matched (NIM) elements [20] form one DOE type which show the capability of efficiency achromatization. These NIM-DOEs consist of a diffractive structure with large profile depth, sandwiched between two materials characterized by a small difference in their refractive index. Compared to diffractive structures with an interface to air, the NIM concept has the advantage of reduced tolerances on profile depth and surface roughness. In the NIM configuration, the dispersion of the two materials must be such that the change of the refractive index, Δn, at the interface remains proportional to λ or equivalently $\lambda/\Delta n(\lambda) = \text{constant}$. This constraint limits the choice of possible material combinations. Furthermore, the material restriction becomes more significant with varying temperature, because dn/dT often differs significantly for the two materials. The need of large profile depth which is typically in the range of ten's of microns and the practicable aspect ratio in the fabrication process restricts the minimum zone width of the diffractive structure and therefore limits the diffractive optical power of the NIM-DOEs.

As already mentioned above, a second approach to realize DOEs with achromatized efficiency uses a multi-layer combination of diffractive structures and an additional air gap [21,22]. Due to the additional degrees of freedom, the above mentioned constraint on the dispersion properties of the materials are removed.

In one type of multi-layer configuration, two single-layer DOE structures are arranged to face each other and are separated by an air gap of a few microns. The two single-layer DOEs are formed to have equal grating pitches but differ in grating heights and in material parameters. For an efficiency achromatization with respect to two wavelengths, both materials can be chosen nearly independently and only the profile depths have to be calculated. On the other hand, a multi-color efficiency achromatization needs two materials with selected dispersion characteristics and adapted profile depths or the introduction of additional layers.

In an alternative multi-layer configuration the air gap is moved to the exterior of the structure [19]. The choice of best layer configuration also depends upon polarization-dependent Fresnel losses and fabrication constraints.

The multi-layer concept has recently been introduced in consumer camera lenses [23]. Compared to a classical dioptric lens design of the same specifications and performance, it was possible to reduce both the overall length by 27% and the weight by 31%.

In hybrid systems designed to image a wide field, the angle-dependent diffraction efficiency of the DOE also becomes relevant.

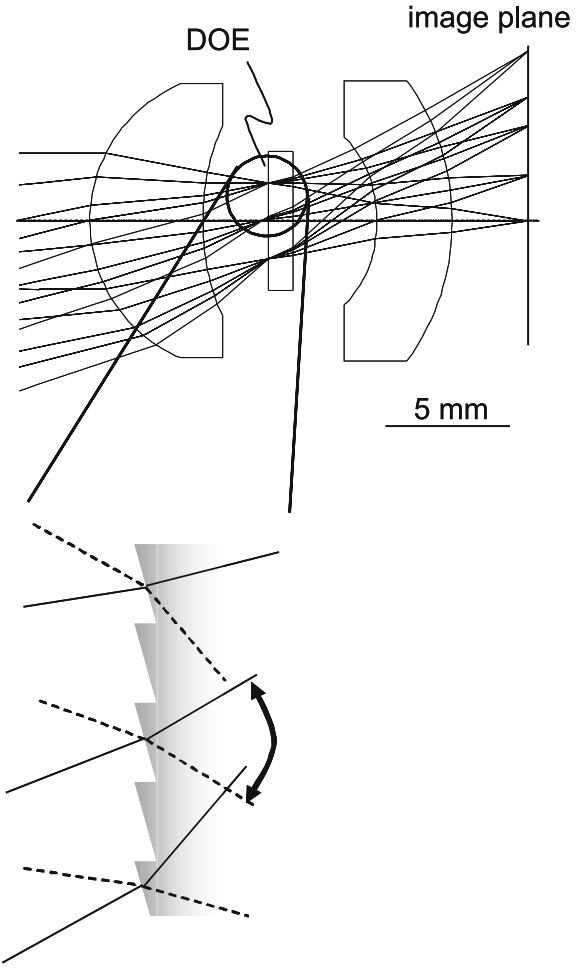


Fig. 3.3. Schematic lens drawing of a hybrid imaging system composed of two concave-convex elements and a DOE. Each field position in the image plane correlates to a different incidence angle on the DOE. The inset shows the variation of the incidence angle on the surface relief for different field positions

For example, consider an imaging system composed of two concave-convex elements opposed to each other and a DOE on a plane substrate centered between them (Fig. 3.3). Each field position in the image plane correlates to a different incidence angle on each position of the DOE.

Generally, the blaze profile of a grating is designed for a particular direction of incidence. A strong variation of the incidence angle will cause a deviation from the optimized blaze angle and profile depth which is correlated to a decrease in the diffraction efficiency. To realize a uniform field illumination it is necessary to avoid this problem.

To estimate, the impact of the angle dependency on the diffraction efficiency, consider a model system in an inorganic glass (BK7) at $\lambda = 546\,\text{nm}$. The blaze profile was optimized by a simple scalar approximation, which results in a profile depth of $d = 1.05\,\mu\text{m}$ ($d = \lambda/(n(\lambda) - 1$; $n_{\text{BK7}} = 1.519$ at 546 nm). The reference incidence

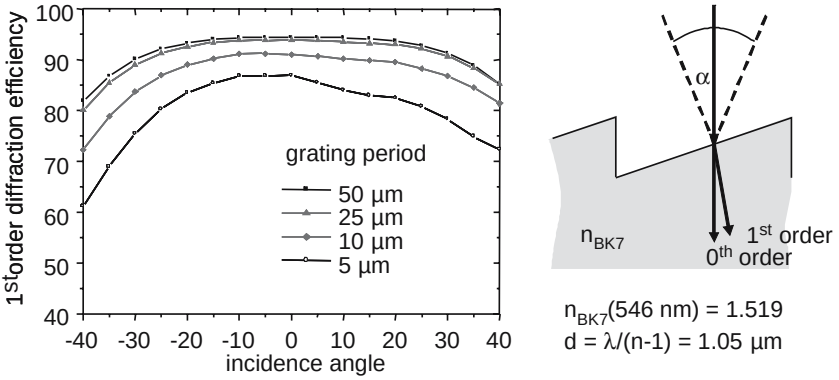


Fig. 3.4. First order diffraction efficiency as a function of the incidence angle for different blazed grating periods

angle was assumed to be perpendicular to the substrate surface, i.e. the anti-blaze facet (draft facet) is oriented parallel to the incoming ray. The calculations were carried out by a fully rigorous electromagnetic treatment in TE-polarization and an extended one-dimensional grating approximation [24,25]. The results for different gratings from 50 µm to 5 µm pitch size clearly show that the maximum efficiency increases with grating period (Fig. 3.4). This effect can be related to the loss of the validity of the scalar approximation. Essentially, it becomes obvious that the diffraction efficiency decays significantly for all grating periods when the incidence angle exceeds about ±25°. The slightly more prominent decay for negative angles is due to asymmetric shadowing effects. As a consequence, in an imaging system such angles will contribute to a nonuniform field illumination and therefore, should be avoided.

3.3 Applications

Microlithography is a key technology in the semiconductor industry. Approximately one third of the total cost of semiconductor manufacturing can be attributed to the microlithography process. Today, the heart of the microlithography process are deep-UV exposure systems designed to produce a demagnified optical image of a photomask on a resist-coated silicon wafer. In subsequent development, etch and deposition steps the latent images are transferred into a permanent structure.

The masks used in this photolithography process are one of the most crucial components, and number up to 30, for advanced semiconductor circuits. The mask contains the original image of the structures, which are transferred several thousand times to the individual chips during the exposure of the wafer. Any defects in the mask pattern will be transferred and will destroy the functionality of the printed devices. Hence defect identification and classification are indispensable tasks. Especially, the continuous progress of photomasks by the introduction of optical enhancement techniques, such as phase shift masks (PSMs) and Optical Proximity

Corrections (OPC), requires new sophisticated optical inspection systems. Usually, the instruments used for this purpose are highly specialized deep-UV microscope systems. The steady increase of the system requirements call for alternative optical concepts. Here, the hybrid approach offers new possibilities to overcome the limitations of dioptric or catadioptric systems.

3.3.1 Hybrid Lens System for High Resolution DUV Mask Inspection

The challenge in designing a complex optical system for the deep-UV regime is a consequence of the limited material selection in this spectral range, combined with the demand of cement-free optical groups. Especially for optical mask inspection where the presence of a protecting pellicle requires a very long working distance, an all-refractive solution for a high-NA objective is challenging. Mask-protecting pellicles are thin transparent membranes suspended typically 7.5 mm above the mask surface. When in the lithography process a dust particle is deposited on the pellicle, it is far out of the focal plane and thus essentially invisible to the projection optics. The pellicles consist of polymers, e.g. Teflon-like fluorocarbons, and have a typical thickness of ~1 μm. To exclude contamination during the optical inspection process, the pellicle should not be removed.

The currently proposed all-refractive designs incorporate an enormous number of single optical elements (e.g. 24 elements in [26]) and therefore involve either an extended volume or sacrifice in the working distance [27].

The hybrid concept depicted in Fig. 3.5 addresses all of these concerns [28,29]. The design combines eight refractive components and a single diffractive element on a plane substrate. The main properties of the hybrid objective are the following:

- large working distance of 7.8 mm
- numerical aperture of 0.65 and 50x magnification
- only fused silica components, calcium fluoride is not needed
- no cemented groups
- the overall length of 45 mm fits in a standard mask evaluation tool
- laser wavelength of 193 nm and a bandwidth of ±0.5 nm

3.4 Design and Realization

The light source for the microscope system, into which the hybrid lens is integrated, is a compact ArF-excimer laser ($\lambda = 193\,\text{nm}$) with a spectral bandwidth of $\Delta\lambda = \pm 0.5\,\text{nm}$. Generally, in this wavelength region the material selection is limited to fused silica (SiO_2) and CaF_2. The refractive index of fused silica changes by $\Delta n = 0.0016$, even over the narrow laser bandwidth. By comparison, the index of a typical inorganic glass (e.g. BK7) changes by a comparable amount over a bandwidth of nearly 40 nm in the visible range. Thus, despite the small bandwidth of the excimer laser, color correction is indispensable for high quality imaging.

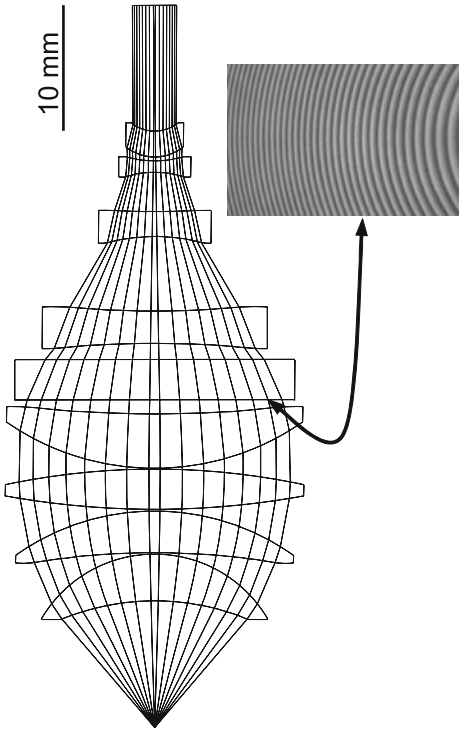


Fig. 3.5. Hybrid refractive-diffractive mask inspection lens system for 193 nm with holographic optical element

A further challenge in realizing a color-corrected high quality objective for deep-UV application is the demand of cement-free optical groups. At 193 nm the energy of the photons is comparable to the binding energy of the organic material conventionally used as an optical cement. Hence, the cemented groups are very sensitive to the exposure dose and will limit the lifetime of a lens system. To overcome these difficulties in a classical dioptric approach, a highly sophisticated air-space technology was introduced [30]. This technology requires a large number of optical elements of both fused silica and CaF_2. Furthermore, the manufacturing process of this objective shows extremely tight tolerances for the air spacing and the centering of the single elements.

As already mentioned in the previous chapter, a reduced diffraction efficiency in the working order results in additional light in the adjacent orders which may reduce the imaging contrast. Figure 3.6 shows the ray trace of both the 0th- and the 2nd-order diffraction. It is clearly observable that in both cases only a small portion is transmitted through the final optical element of the objective. Of the light captured by the NA of the objective and undergoing 0th-order diffraction, only 0.012% reaches the detection plane. For 2nd-order diffraction, the figure is 0.01%. The rest is blocked by the housing and mountings of the tube lens.

The DOE is the most determining element of the hybrid objective and the structure parameters to be realized are very challenging. The minimum feature size of

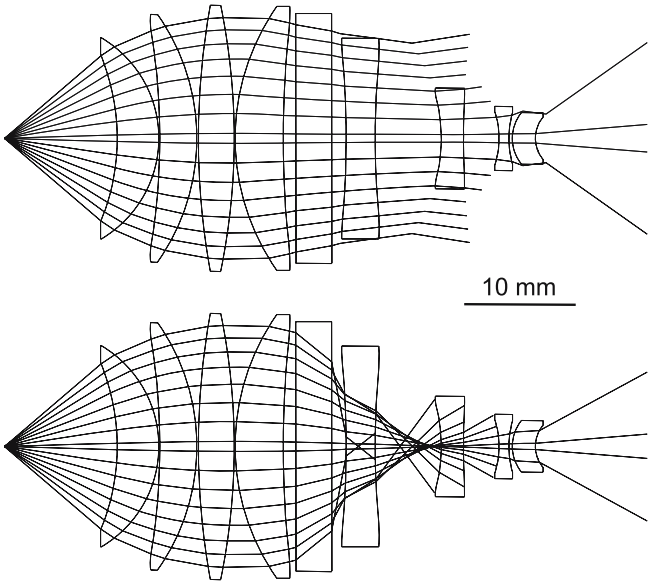


Fig. 3.6. Ray bundle of the unwanted zero order (*top*) and second order (*bottom*) diffraction. Most of the unwanted diffraction light is blocked by the housing of the objective itself and the following tube lens

the DOE is in the range of 550 nm. The necessity of a high diffraction efficiency requires a sawtooth like profile with a calculated depth of ≈300 nm in fused silica. To meet these requirements we chose a two-step process involving interference lithography followed by reactive ion etching. In the first step, two spherical wave fronts interfere to create the concentric ring pattern. The orientation of the lines of minimum intensity in the positive photoresist defines the asymmetrical groove profile required for high diffraction efficiency. The setup allows a maximum depth in the resist of ∼100 nm, i.e. a factor of three is necessary during the profile transfer with the etching process.

An advanced technique is the proportional transfer of 3D resist structures into the hard optical material by means of a reactive ion-etching (RIE) or (reactive) ion-beam-etching process (R)IBE. The main difficulty to be solved in the development of those techniques is the adjustment of the etching rates of materials with very different etching behavior namely the resist or polymer mask, on the one hand, and the hard optical substrate material, on the other. All relevant etching parameters can be used for such optimization. The problem will normally grow when the curvatures of the microstructures involve a great variety of gradients.

3.5 Application Examples

The results of a series of test measurements with the hybrid objective integrated in a mask inspection microscope working in transmission mode are shown in Fig. 3.7. As a test structure a chrome-on-glass (COG) mask with different periodic lines and spaces was used (L/S: 200 nm, 175 nm, 150 nm, and 125 nm). Lines and spaces of 125 nm half pitch on mask can be resolved clearly.

To demonstrate the further potential of the hybrid lens system we compared the high resolution mode with the mask imaging system AIMS^{TM} (Aerial Imaging Measurement System). The AIMS^{TM} microscope allows to emulate the imaging characteristics of a stepper/scanner under equivalent optical conditions [31]. Many mask defects do not print in the lithography process. To decide if a mask has to undergo a costly repair process, it is necessary to identify defects that are only emerging and which do not yet print. The emulation with the AIMS^{TM}-tool enables the assessment of the printability of defects on the mask in a fast and reliable way without the need of wafer prints. The demand of stepper/scanner equivalent settings also limits the numerical aperture of condenser and objective lens of the AIMS^{TM} microscope to values NA $\leq$ 0.23, so that the single aerial image corresponds to a low resolution mode.

The comparison of both modes, the high resolution mode with the hybrid lens system and the AIMS^{TM} mode, is of particular interest for structures on the mask having so-called OPC-features. Optical proximity correction (OPC) is a lithography enhancement technique used to compensate transcription errors caused by the finite spatial bandwidth of the optical projection lens [32]. Relevant optical proximity effects which have to be treated are linewidth differences between dense lines and isolated lines, line end shortening, where the ends of lines tend to contract, and corner rounding, which means that rectangular patterns on the mask will print as circles on the wafer. The OPC technique uses pattern modifications of the photomask to reduce effects like theses. Typical assist features added to the mask layout are so-called scattering bars, serifs or hammerheads. These features do not print in the lithography process but they modify the image to the favored shape.

The upper part of Fig. 3.8 shows the layout of a chrome on glass pattern with binary mask structures comprising main bars with attached serifs and scattering

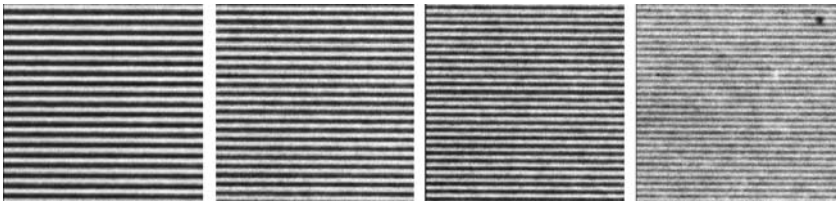


Fig. 3.7. Images of a series of test measurements with the 193 nm-hybrid lens. As a test structure a chrome on glass mask with different lines and spaces (1 : 1) was used (from *left* to *right* L/S: 200 nm, 175 nm, 150 nm, and 125 nm)

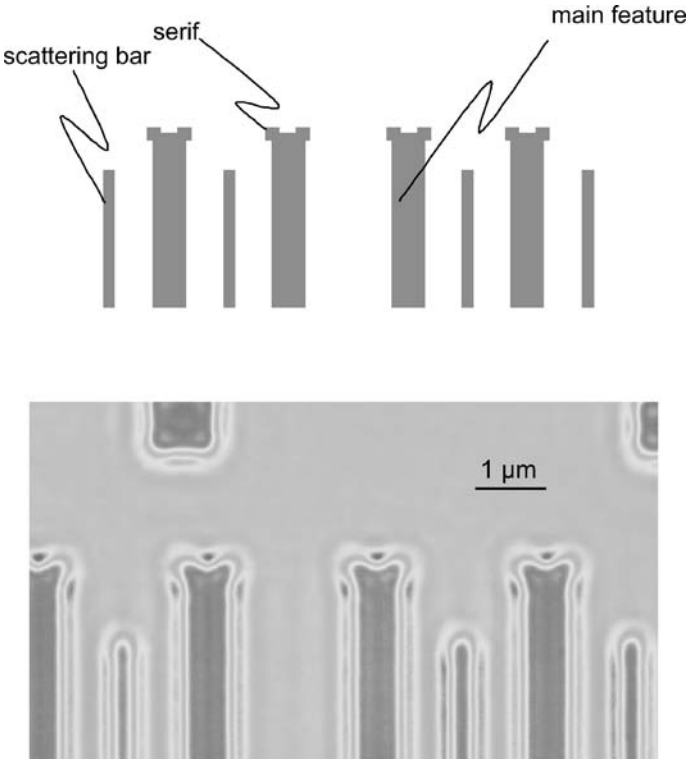


Fig. 3.8. OPC-layout (*top*) and high resolution image (*bottom*) of a chrome test pattern on a clear substrate containing serifs and assist features

bars spaced between the main features. Both positive (opaque features – surrounding transparent) and negative (transparent features – surrounding opaque) structures have been measured. The lower part of Fig. 3.8 displays a high resolution image of a positive structure taken with the hybrid lens system. The width of the main bars are 600 nm, the edge length of the serifs is 160 nm and the scattering bars are 200 nm wide. All assist features can clearly be recognized. In Fig. 3.9 the comparison of the AIMS^{TM} mode (left) and the high resolution mode (right) is shown measured on two negative structures with different features sizes in the layout:

- 600 nm main features, scattering bars 200 nm, 160 nm serifs
- 400 nm main features, scattering bars 120 nm, 100 nm serifs

For both feature sizes, 600 nm and 400 nm on mask, the aerial image shows four distinct lines which are printable in the lithography process. Due to the scattering bars, a slight intensity transmission is observable in the spacing between the main feature. The color gradient of the images displays the transmitted intensity continuously. However, since the photolithographic printing process follows a distinct

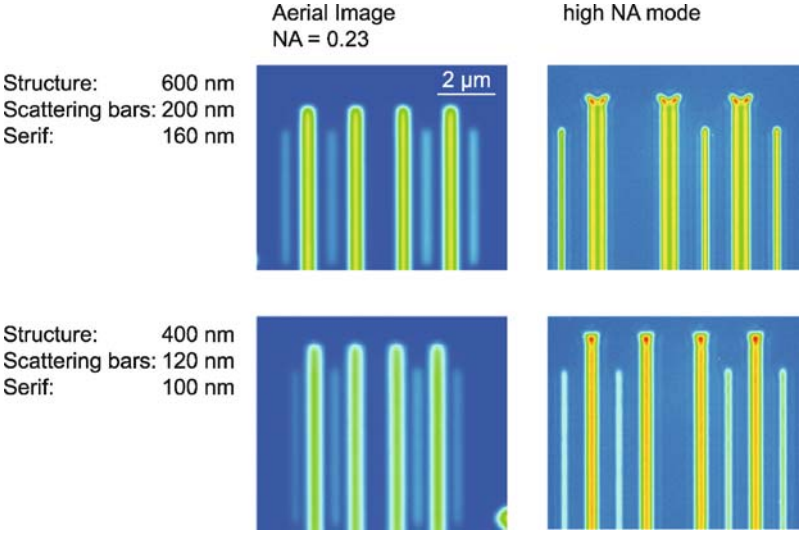


Fig. 3.9. Comparison of aerial image measurements (*left*) and high resolution mode (*right*) on different feature sizes. AIMS^{TM} investigates the printability of a lithography tool. In the high resolution mode the assist features become visible

threshold value, the transmission intensity correlated to the scattering bars will be below the particular threshold value and therefore will not print.

Applying the high resolution mode, the assist features become visible between the main structures. For the 600 nm feature, the line ends clearly show the serifs at the left and right corner. The 120nm serifs of the 400 nm structures are not fully resolved. However, we can see that there are no defects at the left and right corners, which would make the corner rounding worse in the aerial image formation.

Contact holes are also suitable objects to compare the AIMS^{TM} and the high resolution mode and to demonstrate the performance of the hybrid lens system. The structure to be investigated is an array of contact holes with contact size of 0.81 µm

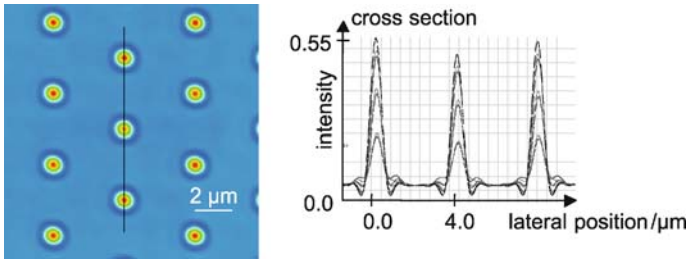


Fig. 3.10. Aerial image at best focus of an array of contact holes (*left*) and the selected through focus intensity profiles (*right*). The feature near the center shows a decrease in the intensity of approximately 10% compared to the neighboring structures

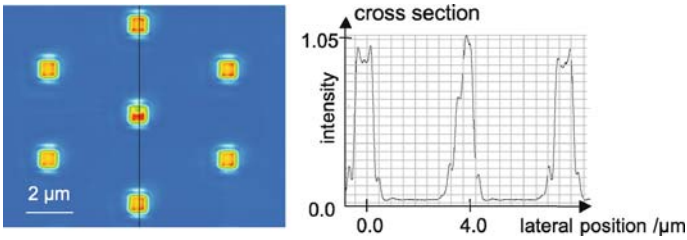


Fig. 3.11. High resolution image of the same area of contact holes as investigated before with AIMS^{TM} (*left*) and the selected intensity profile (*right*). The defect can be assigned to protrusion on edge

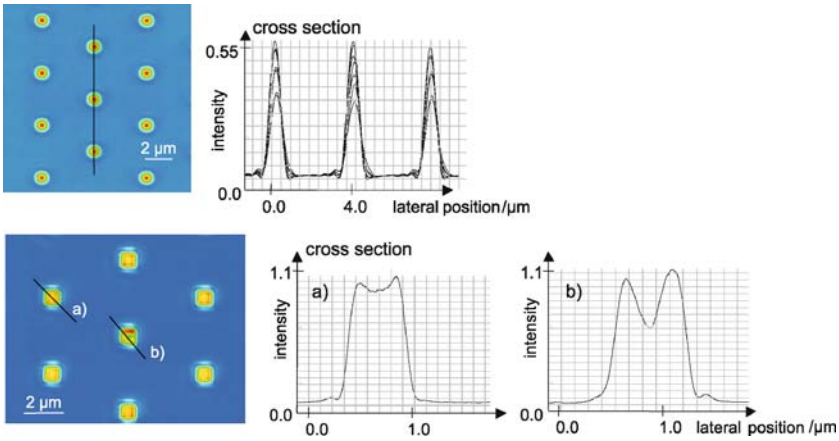


Fig. 3.12. Aerial image at best focus of an array of contact holes (*top left*) and the selected through focus intensity profiles (*top right*). The aerial image shows no differences in the intensity structure for all contact holes. The high resolution image of the same area (*bottom*) shows clearly a defect in the contact hole near the center (protrusion on corner)

on the mask. Selected contact holes are decorated with defined mask defects. Figure 3.10 (left) displays the aerial image (NA = 0.23) of some contact holes at best focus. The feature near the center of the image shows a slight decrease in the maximum intensity compared to the neighboring structures. This becomes more clear when the cross sections of a through focus series are compared (Fig. 3.10 right). The decrease in intensity of the center feature is approximately 10%. The analysis of the aerial image together with the characteristic response of the photoresist gives an answer about the printability of the special contact hole. The special kind of mask defect would not be predictable through this method with such a precision.

In Fig. 3.11 the corresponding high resolution image of a part of the same area is shown as investigated before with the AIMS^{TM} setup. The squared shape of the underlying basic structure of the contact holes can clearly be identified. The contact hole in the center of the image shows a significant intensity reduction in the upper

part of the structure. This becomes prominent when taking the cross section of three vertical aligned contact holes in the center of the image. The result can be assigned to the defined defect of protrusion on edge.

Figure 3.12 shows the results measured on a different position on the mask. In the aerial image at best focus (Fig. 3.12, upper part) no deviation in the intensity profile of the different features could be observed and the mask can be used for wafer printing of equally sized contacts. However, the high resolution image (Fig. 3.12, lower part) reveals that the center contact possesses a defective corner (protrusion on corner). There is no need for repair of this non-printing defect for use under specific lithography settings. Yet, such defects are important to know and they may matter later in the life cycle of the mask.

3.6 Resolution Enhancement with Solid Immersion Lens (SIL)

Following the classical theory of diffraction, the resolution of an optical microscope is determined by the illumination or detection wavelength λ and the numerical aperture NA of the system. The NA is the product of the sine of the half-angle θ of the cone of light accepted by the objective lens and the refractive index n of the objective space ($\mathrm{NA} = n \sin\theta$). From this relationship, the options for optical resolution enhancement techniques can be deduced. An alternative approach to wavelength reduction is the use of an immersion system, which increases the refractive index in front of the sample. In this context the use of Solid Immersion Lenses (SIL) has attracted much attention owing to the possibility of simply combining a conventional far-field optical system with a high-refractive-index lens placed in close proximity to the sample [33].

The most simple geometry of a SIL is the classical hemisphere (hSIL), in which the rays incident perpendicular to the spherical surface propagate to the center without deviation. For our experiments we combined the hybrid microscope objective with a hSIL of fused silica with a refractive index of 1.56 at 193 nm.

The introduction of the hSIL in the optical path does not influence the distance between hybrid objective and mask. Additionally to the increase in the NA of the combined system, the magnification is increased by a factor of n as well. The positioning of the hSIL with respect to the mask has to be done very carefully, because the bottom surface of the hSIL has to be in direct contact to the mask. In the present configuration with a $\sin\theta = 0.65$ and a refractive index of $n(\mathrm{SiO_2}@193\,nm) = 1.56$, only propagating fields are involved; evanescent fields are not. The capability of the hSIL combined with the hybrid objective was investigated again on the COG mask with periodic lines and spaces. Figure 3.13 depicts the clearly resolved images of a 100 nm L/S structure in high contrast (additional magnification by a factor of n). This measurement demonstrates that, in spite of the difficulties in the handling of the SIL, the advantages of the resolution improvement are significant.

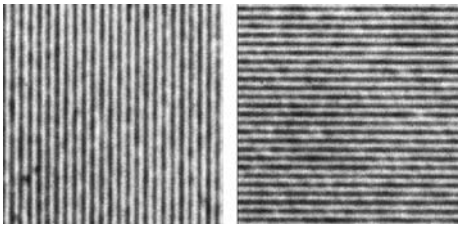


Fig. 3.13. Test measurements of a chrome on glass mask with lines and spaces. The combination of the hybrid objective and a hemispherical Solid Immersion Lens (hSIL) allows to resolve clearly features sizes down to 100 nm (L/S 1 : 1, 100 nm)

3.6.1 DUV Microscopy with NIR Autofocus: Wavelength Selective DOE Combination

High-resolution DUV-systems, developed for high-throughput defect inspection, often require an additional autofocus (AF)-system, which acts simultaneously with the imaging process. Here, the wavelength of the AF-system is usually shifted to the near infrared region (NIR) (between 680 nm and 850 nm) to exclude interfering effects with the imaging wavelength in the deep-UV. To guarantee a perfect functionality of the optical system, the foci of both wavelengths (DUV and NIR nm) must be adjusted precisely in the same position. In addition to the already addressed limitations arising in DUV regime, the two very different wavelengths cause a further complication in order to fulfill all requirements with a dioptric solution.

In the following the hybrid concept of an high-NA DUV lens system, which allows the integration of an AF-system working at 785 nm is discussed [15]. Here the application is high-throughput wafer inspection with some different requirements with respect to the pervious presented mask inspection tool. To detect and classify the defects on processed wafers, the main parameters of the lens system are an NA = 0.9, a 120x magnification, and a working distance of 200 μm. The lens system was specified for an imaging wavelength of 248 nm with a bandwidth of ±2 nm, and an autofocus wavelength of 785 ± 5 nm.

Figure 3.14 shows the scheme of the lens design and the ray trace for both the DUV wavelength (top) and the AF-wavelength (bottom). The compact solution combines 10 refractive optical elements and two DOEs, each on a plane-parallel substrate. Cemented combinations are not necessary. The light from the autofocus source is collimated and reflected into the optical path of the DUV imaging system. Inside the hybrid lens system only a part of the outer rim of the optics is used to guide the AF-wavelength to the wafer surface. After reflection on the sample, the AF-light passes through the lens system on the opposite sector and again, is reflected out of the optical path of the imaging system. Depending on the particular focus position on the wafer, the spot of the reflected and focused AF-light will illuminate a different position on a segmented diode.

Of particular importance in the lens design is the combination of the two DOEs. Here, the wavelength-dependent efficiency of diffractive elements is applied in a useful sense, so that both DOEs are acting as wavelength selective elements. In particular, the DUV-DOE is designed to guarantee high efficiency in the 1st diffraction order for the deep-UV wavelength and simultaneously a high efficiency in the

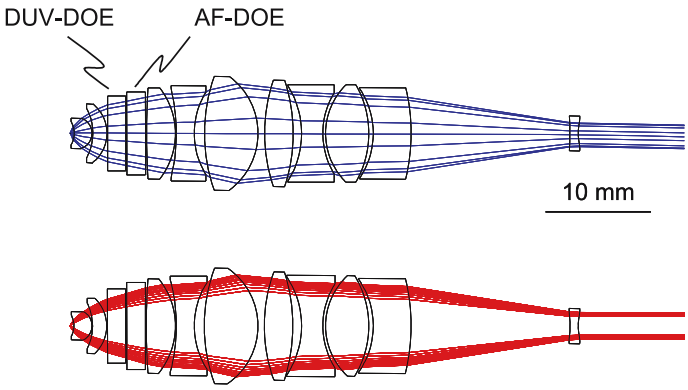
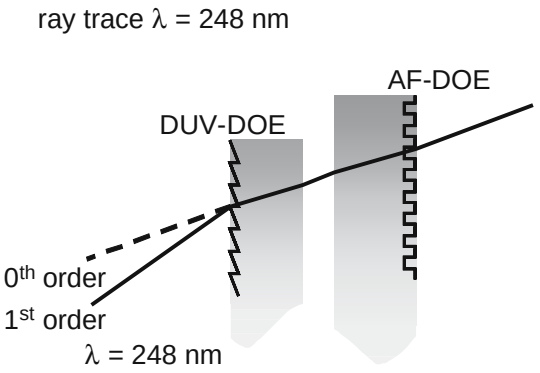


Fig. 3.14. Lens design of a hybrid high NA objective for deep-UV applications. A particular combination of two wavelength selective DOEs allows both, imaging at 248 nm (*top ray trace*) and also the integration of an AF-system working at 785 nm (*bottom ray trace*)

0th order for the autofocus wavelength. The efficiency for all other diffraction orders should be low. The diffraction behavior of the AF-DOE is reversed. This means the AF-DOE posses a high efficiency for the IR wavelength in the 1st diffraction order and a high efficiency for the DUV wavelength in the 0th diffraction order. In other words, the DUV-DOE is acting as simple plane-parallel element for the AF wavelength and for the DUV wavelength the local distribution of the spatial frequencies define the imaging properties. In the AF-DOE, the distribution of the spatial frequencies are chosen to optimize the imaging properties of the AF system, whereas the DUV wavelength is only affected by refraction. Figure 3.15 shows schematically the working principal of the proposed setup. This approach allows a partial decoupling of the design of the two different wavelengths and an almost independent correction of the lens design for both, the AF-system and the DUV imaging system.

The imaging quality of the system at 248 nm is the most important specification. Therefore, it is essential to achieve the efficiency requirements for DUV on both DOEs. An appropriate design for the DUV-DOE is a sawtooth-like, blazed profile, optimized for 248 nm, which corresponds to a depth of 400 nm in fused silica. Following the lens design, a maximum spatial frequency of 410 lines/mm is necessary for the DUV-DOE. To find an appropriate profile for the AF-DOE we start with a binary phase profile of equal size in line and space and a spatial frequency of 100 l/mm which corresponds to the maximum frequency. The efficiencies of the desired orders (0th order DUV, 1st order NIR) were calculated as a function of the profile depth. The calculations were carried out with a rigorous coupled wave analysis (RCWA) algorithm. The results are shown in Fig. 3.16. The efficiency curves of the 0th order at 248 nm show a periodic behavior with only a small deviation between TE- and TM-polarization. The maximum efficiencies are in the range of 80%. The 1st order at 785 nm also exhibits a periodic curve with a different height dependence, but with a maximum efficiency of only ~35%. This value should be sufficient for the limited



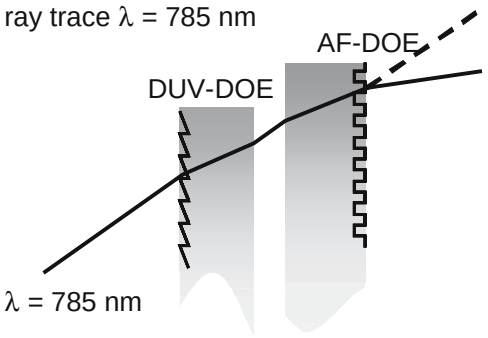


Fig. 3.15. Schematic view of the working principal of the wavelength selective DOE doublet. *Left*: ray trace of 1st diffraction order at 248 nm; *right*: ray trace of 1st diffraction order at AF wavelength. The DUV-DOE shows a high efficiency in the 1st diffraction order at 248 nm and acts as a plane parallel element at the AF wavelength. The AF-DOE shows the opposite behavior

imaging functionality of the autofocus system. A comparison of the curves shows that a height of 950 nm in binary profile should fulfill the efficiency requirements.

3.6.2 Diffraction Based Solid Immersion Lens

The use of a Solid Immersion Lens (SIL) is a promising optical method to achieve an improved optical resolution and therefore allows a higher capability in applications such as optical data storage, microscopy, or optical lithography. In the SIL technique, a conventional far-field optical system is combined with a high-index solid lens, placed in close proximity to the sample. As in liquid immersion microscopy, the numerical aperture of the optical system is enhanced by introducing high index material, which corresponds to a reduced size of the diffraction-limited spot. At the bottom surface of the SIL, at the interface between the SIL and the sample there is an air gap which has to be passed in order to interact with the sample. The small angle components of the incident light, will be refracted, but continue to propagate into the air gap outside of the SIL. When the incidence angle is larger than the critical angle θ_C ($\sin(\theta_C) = 1/n$), light experiences total internal reflection,

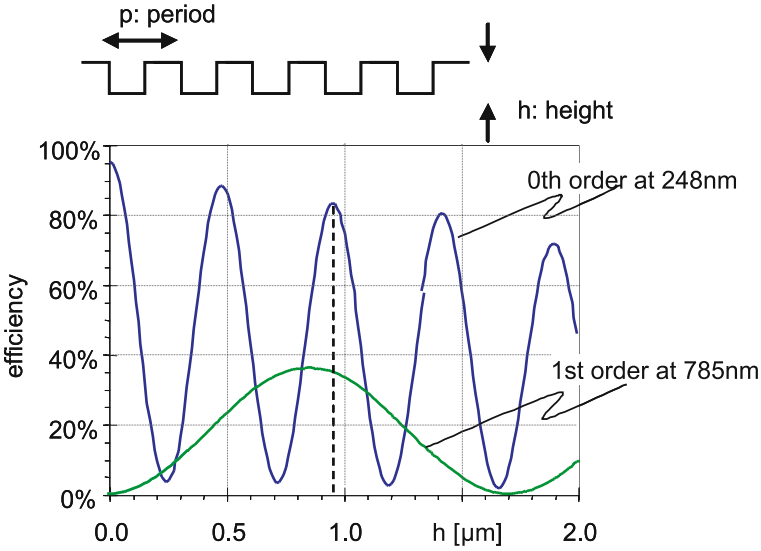


Fig. 3.16. Diffraction efficiency of a binary phase-only DOE as a function of the profile depth. Displayed is the 0th order efficiency at 248 nm and the 1st order efficiency at 785 nm. At a profile depth of ~0.950 µm both efficiency requirements are fulfilled

so that only the evanescent fields or the near-field components have the possibility to interact with the sample. These large angle components corresponds to the high spatial frequencies and to the small sample structures and therefore are essential for high resolution imaging. The associated electromagnetic near-fields decay into the air gap with a penetration depth on the order of the wavelength. Besides high spatial resolution, solid immersion microscopy also offers a high transmission efficiency, especially in fluorescence microscopy [34], and the capability of parallel imaging.

The SILs described in the literature are mainly simple hemispheres (hSIL), in which the rays incident perpendicular to the spherical surface propagate to the center without deviation, or refraction-based superspheres (sSIL) (see Fig. 3.17a,b) [33,35,36]. Superspheres, also called Weierstrass-spheres, are truncated spheres with a height equal to $(1 + 1/n)r$, where r is the radius of curvature. It is theoretically possible for both types to perform imaging near the center of their bottom surface without geometric aberration. A hemispherical SIL can improve the NA of the far-field objective by a factor of n, for supersphere the factor is n^2. In both cases the maximum achievable NA is equal to n. Due to the index-dependent dimensions of the supersphere-SIL, chromatic aberrations can arise.

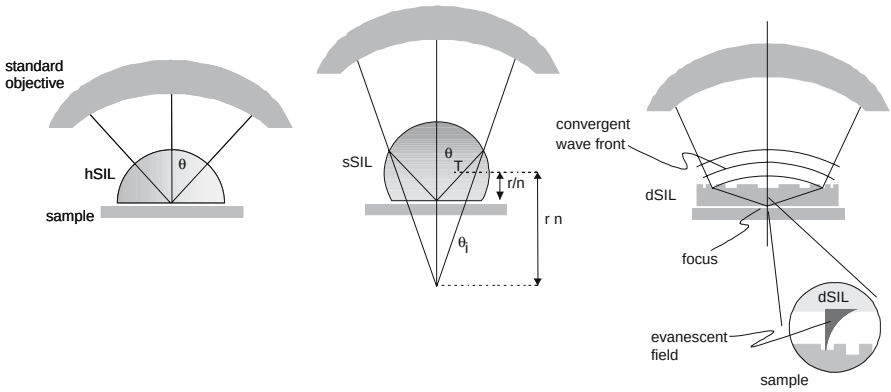


Fig. 3.17. Principle of hemispherical-SIL (*left*), Weierstrass-SIL (*center*) and diffraction based SIL (*right*)

3.7 dSIL: Concept and Phase Effects

An alternative to the spherical concepts are diffraction-based Solid Immersion Lenses (dSILs) [37]. Analogous to a Fresnel zone plate, the concentric ring system on the upper surface of the dSIL is designed so that an incoming wave front is diffracted to a focus at the lower surface of the dSIL (see Fig. 3.17c). SILs based on refraction are limited by the difficulty of fabricating anything other than a sphere, whereas the fabrication of a dSIL is amenable to "aspheric" designs. The large curvature of a refractive SIL easily introduces aberrations [38]. In comparison, diffractive structures can be etched into a plan parallel substrate, using highly accurate lithography. Diffractive SILs may also be more amenable to low-cost replication.

In the basic approach to design the annular ring system of the dSIL the Huygens-Fresnel principle has to be taken into account. The light deflected from adjacent zones of the dSIL interferes constructively at the focus when the optical path length difference (OPD) is an integer multiple m of the wavelength λ. In an extension of this approach an additional phase offset $\varphi_{\text{off};n}$ has to be considered, which is introduced by the diffractive structure itself at the upper surface of the dSIL. This phase offset varies from zone to zone and is dependent on the incidence angle, polarization, the corresponding local grating period and groove profile. The phase offset becomes crucial when the classical Rayleigh limit of $\lambda/4$ wave front aberration is exceeded. In particular, it will occur when the local period of the diffractive structure is on the order of the wavelength i.e. for large incidence angles and high numerical apertures.

An obvious example for the theoretical investigations of the phase offset is the model of a diffractive analog to a classical hSIL. Here the first order diffracted waves focus at the bottom surface of the dSIL, at the same geometric position and with the same angular spectrum as would occur without the dSIL. That means, each incidence angle is directly correlated to a spatial frequency of the dSIL.

To match the experimental data, a dielectric dSIL substrate was chosen with refractive index of $n = 2.02$ (Schott; LaSF35) at an illumination wavelength of

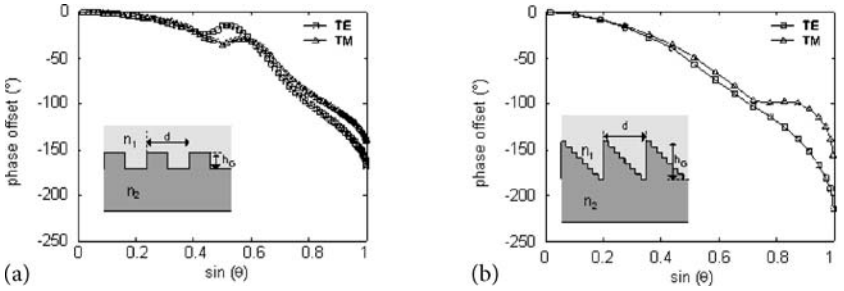


Fig. 3.18. Phase offset as a function of the incidence angle introduced by the surface relief structure of a dSIL. (**a**): Binary phase-only profile ($h = 200\,\text{nm}$); (**b**): 8-level profile ($h = 350\,\text{nm}$). Due to the application constraints, the variation of the incidence angle is correlated with a change of the grating period

632.8 nm. Both, a binary and an 8-level phase profile were investigated. Due to the difficulties to realize a variable depth in the groove profile for different spatial frequencies, a constant depth for each profile geometry was assumed (h(binary) = 200 nm, $h(8 - \text{level}) = 350\,\text{nm}$ respectively). Here the height is selected to obtain a high efficiency in the outer region of each dSIL. Figure 3.18a shows the calculated and normalized phase offset for the uncompensated binary phase profile. It is clearly observable that both polarization directions show nearly identical behavior. For decreasing spatial periods the phase offset becomes more significant, and the Rayleigh limit of $\lambda/4$ wave front aberration will be exceeded when $\sin\theta \approx 0.7$. Figure 3.18b shows the identical calculation for an 8-level phase profile. Again TE- and TM-polarization show very similar behavior. In comparison to the binary profile, here the maximum phase offset is more dominant and the Rayleigh limit is already reached for an incidence angle of $\sin\theta \approx 0.6$. To compensate the phase offset, an additional phase plate could be introduced. Alternatively, one could also tune the local period of the diffractive structure. Such a design would require the simultaneous solution of the grating and coupled-wave equations.

3.8 dSIL: Experimental

The dSILs were realized as binary phase elements by means of microstructuring techniques. They were designed and fabricated for illumination at 632.8 nm. The binary structure is realized in the high index glass LaSF35 by direct e-beam writing combined with reactive ion etching (RIE).

To evaluate the imaging properties of the dSIL, the light distribution of the focused spot at the bottom surface of the dSIL is measured directly with a near-field scanning optical microscope (NSOM) (Fig. 3.19).

Therefore the light of a HeNe–laser (632.8 nm) is focused by a microscope objective with NA = 0.6 and working distance of 3.02 mm. The wave diffracted by the dSIL comes to a focus on the flat bottom surface, parfocal to the objective operating

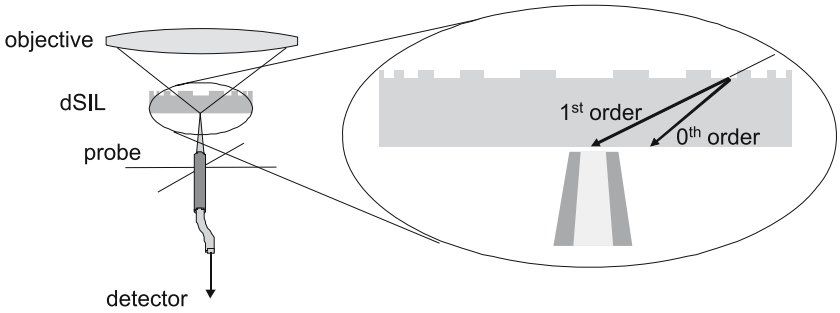


Fig. 3.19. Schematic view of the experimental setup to evaluate the resolution enhancement of the dSIL concept. Via a conventional microscope lens and the dSIL, a focused spot is realized at the bottom of the dSIL substrate. The extension of the spot size is measured by the use of a aperture probe realized by a tapered and aluminized glass fiber tip

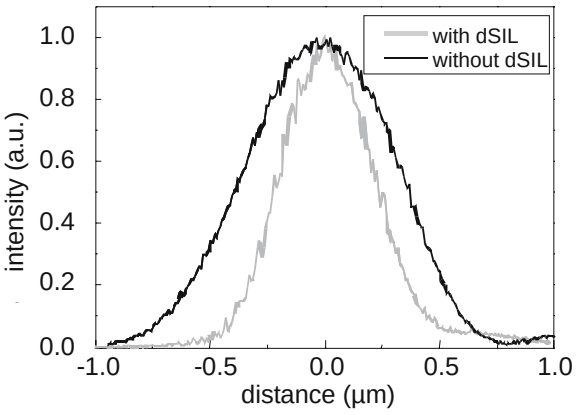


Fig. 3.20. Comparison of the measured spot size with and without dSIL. For a correct analysis of the resolution enhancement the convolution of probe aperture and optical field distribution has to be taken into account

alone. A near-field probe below the dSIL is placed in contact with the bottom surface. The probe is a tapered and aluminized glass fiber with an aperture of 100 nm. The optical field collected by the probe is detected with a photomultiplier tube. The probe scans in x and y and moves in the z direction to maintain contact. Measured in this way, the focal spot has a FWHM of 450 nm (Fig. 3.20). To characterize the focal spot of the microscope objective alone, the measurement is repeated without the dSIL, yielding a FWHM of 770 nm. The measured optical field is a convolution of the real optical field and the NSOM tip aperture. The approximate real field widths, 350 nm and 670 nm, can be obtained by subtracting the aperture diameter from the measured widths. As expected, the ratio is close to the refractive index of the dSIL and the measured spot size corresponds well to the theoretical estimation.

3.9 Final Remarks

With examples in deep-UV microscopy and the discussion of diffraction based solid immersion lenses, it was demonstrated, that high-resolution imaging profits from the utilization of the hybrid combination of diffractive and refractive lenses. In comparison to conventional setups, the hybrid concept offers an extended flexibility in the optical design, which allows the realization of lens systems with optimized performance and compactness.

To some specific optical characteristics of imaging DOEs, e.g. the wavelength selectivity of particular diffraction orders, exist no refraction based equivalent. This allows system solutions with completely new imaging properties which could not be realized with a classical dioptric or catadioptric approach.

As usual advantages are accompanied by disadvantages. This also becomes true with the hybrid concept. Especially when a hybrid system is exposed to broadband illumination, consideration should be given to the diffraction efficiency and the effects of stray light on the system performance. For a successful work on applications using DOEs in imaging systems a diversity of experiences are necessary such as knowledge of the chromatic characteristics or effects related to the diffraction efficiency of DOEs.

In this context, the interactions between the optical design and the technologies to realize DOEs are of enormous importance. The best possible knowledge of all realization steps of DOEs, involving mastering, tooling and replication, will reduce costly and time consuming iteration steps.

Acknowledgements

We would like to thank all our scientific partners, especially Jeffrey O. White (former MRL, University of Illinois, Urbana, US), Margit Ferstl (HHI, Berlin, Germany), Renate Fechner and Axel Schindler (both IOM, Leipzig, Germany) for their excellent cooperation and the Federal Ministry of Education and Research for their continued financial support. The particular respect, the acknowledgements and the thanks go to our numerous co-workers at Carl Zeiss, especially to the diffractive optics research group for their excellent work, their enthusiasm, creativity and their commitment. Their work is the foundation for the continuous progress in applications for diffractive optics.

References

1. D. Attwood, “Soft x-ray and extreme ultraviolet radiation – principles and applications”, Cambridge University Press 1999
2. M.D. Missig, and G.M. Morris, Diffractive optics applied to eyepiece design, Appl. Opt. 34, 2452–2461 (1995)

3. W. Knapp, G. Blough, K. Khajurival, R. Michaels, B. Tatian, and B. Volk, Optical design comparison of 60 eyepieces: one with a diffractive surface and one with aspherics, Appl. Opt. 34, 4756–4760 (1997)
4. Z.-Q. Wang, H.-J. Zhang, R.-L. Fu, G.-G. Mu, Z.-W. Lu, C.M. Cartwright, and W.A. Gillespie, Hybrid diffractive refractive ultra-wide-angle eyepieces, Optik 113, 159–162 (2002)
5. C.G. Blough, M.J. Hoppe, D.R. Hand, and W.J. Peck, Achromatic eyepieces using acrylic diffractive lenses, Proc. SPIE Vol. 2600 (1995) 93–99
6. Z. Yun, Y.L. Lam, Y. Zhou, X. Yuan, L. Zhao, and J. Liu, Eyepiece design with refractive-diffractive hybrid elements, Proc. SPIE Vol. 4093 (2000) 474–480
7. G. De Vos and G. Brandt, Use of holographic optical elements in HMDs, Proc. SPIE Vol. 1290 (1990) 70–80
8. J.A. Cox, T.A. Fritz, and T. Werner, Application and demonstration of diffractive optics for head-mounted displays, Proc. SPIE Vol. 2218 (1994) 32–40
9. J.P. Rolland, M.W. Krueger, and A.A. Goon, Dynamic focusing in head-mounted displays, Proc. SPIE Vol. 3639 (1999) 463–470
10. T. Nakai and H. Ogawa, "Research on multi-layer diffractive optical elements and their application to camera lenses", in "Diffractive Optics and Micro-Optics" Technical Digest, 5–7, 2002
11. J. Nowak, and J. Masajada, *Hybrid apochromatic lens*, Opt. Appl. Vol. 30 nr 2/3 (2000) 271–275
12. D.A. Buralli, Optical performance of holographic kinoforms, Appl. Opt. 28, (1989) 976–983
13. C. Londono, and P.P. Clark, Modeling diffraction efficiency effects when designing hybrid diffractive lens systems, Appl. Opt. 31, (1992) 2248–2251
14. G.M. Morris and D.A. Buralli, Diffractive optics solution for phakic intraocular lenses, in Conference Digest on Diffractive Optics and Micro-Optics, Topical Meeting, Rochester, USA (2004)
15. R. Brunner, R. Steiner, H.-J. Dobschal, D. Martin, M. Burkhardt, and M. Helgert, New Solutions to Realize Complex Optical Systems by a Combination of Diffractive and Refractive Optical Components; Proc. SPIE Vol. 5183 (2003) 47–55
16. Y. Ori, K. Takada, J. Hashimura, and N. Mushiake, Diffractive/refractive hybrids for blue LD optical storage, in Conference Digest on Diffractive Optics and Micro-Optics, Topical Meeting, Rochester, USA (2004)
17. S. Noach, A Lewis, Y. Arieli and N. Eisenberg, Integrated diffractive and refractive elements for spectrum shaping, Appl. Opt. 35 (1996) 3635–3639.
18. J.E. Ford, F. Xu, and Y. Fainman, Wavelength selective planar holograms, Opt. Lett. 21 (1996) 80–82
19. H.P. Herzig, and A. Schilling, *Optical Systems – Design Using Microoptics* in Encyclopedia of Optical Engineering Vol 2, ed. R.G. Driggers, Marcel Dekker, (2003) 1830–1842
20. S.M. Ebstein, *Nearly index-matched optics for aspherical, diffractive, and achromatic-phase diffractive elements*, Opt. Lett. 21 (1996) 1454–1456
21. Y. Arieli, S. Noach, S. Ozeri, and N. Eisenberg, Design of diffractive optical elements for multiple wavelengths, Appl. Opt.37, (1998) 6174–6177
22. Y. Arieli, S. Ozeri, T. Eisenberg, and S. Noach, Design of diffractive optical elements for wide spectral bandwidth, Opt. Lett. 23, (1998) 823–824
23. T. Nakai, and H. Ogawa, *Research on multi-layer diffractive optical elements and their applications to photographic lenses,* in Techn. Digest, 3rd International conference on optics-photonics design & fabrication, Tokyo, Japan (2002) 61–62

24. M.G. Moharam and T.K. Gaylord, "Diffraction analysis of dielectric surface-relief gratings," J. Opt. Soc. Am. **72**, 1385–1392 (1982).
25. J. Bischoff, "Beitraege zur theoretischen und experimentellen Untersuchung der Lichtbeugung an mikrostrukturierten Mehrschichtsystemen," Habilitation thesis (Ilmenau Technical University, Ilmenau, Germany, 2001).
26. Kenji Ono; United States Patent; Patent No.: US 6,392,814 B1 (2002)
27. W. Vollrath; United States Patent; Patent No.: US 5,440,422 (1995)
28. R. Brunner, R. Steiner, K. Rudolf, and H.-J. Dobschal, Diffractive-Refractive Hybrid Microscope Objective for 193 nm Inspection Systems; Proc. SPIE Vol. 5177 (2003) 9–15
29. R. Brunner, A. Menck, R. Steiner, G. Buchda, S. Weissenberg, U. Horn, and A.M. Zibold, *Immersion Mask Inspection with Hybrid-Microscopic Systems at 193 nm*; Proc. SPIE Vol. 5567 (2004) 887–893
30. T. Sure, J. Heil, and J. Wesner, *Microscope objective production: On the way from the micrometer scale to the nanometer scale*, Proc. SPIE Vol. 5180 (2003) 283–292
31. R.A. Budd, D.B. Dove, J.L. Staples, r. M. Martino, R.A. Ferguson, and J.T. Weed, *Development and application of a new tool for lithographic mask evaluation, the stepper equivalent Aerial Image Measurement System, AIMS*, IBM J. Res. Develop. 41 (1/2) (1997) 119–129
32. H.J. Levinson, "Principles of Lithography", Bellingham SPIE Press 2001
33. G.S. Kino, *Fields associated with the solid immersion lens*, Proc. SPIE Vol. 3467 (1998) 128–137
34. K. Koyama, M. Yoshita, M. Baba, T. Suemoto, and H. Akiyama; High collection efficiency in fluorescence microscopy with a solid immersion lens; Appl. Phys. Lett. **75,** 1667–1669 (1999).
35. S.M. Mansfield and G.S. Kino, Solid Immersion microscope, Appl. Phys. Lett. 57 (1990) 2615–2616
36. B.D. Terris, H.J. Mamin, and D. Rugar, Near-field optical data storage, Appl. Phys. Lett. 68 (1996) 141–143
37. R. Brunner, M. Burkhardt, A. Pesch, O. Sandfuchs, M. Ferstl, S.C. Hohng, and J.O. White, *Diffraction Based Solid Immersion Lens*; J. Opt. Soc. Am. A **21** (7) (2004) 1186–1191
38. M. Baba, T. Sasaki, M. Yoshita, and H. Akiyama, "Aberrations and allowances for errors in a hemisphere solid immersion lens for submicron-resolution photoluminescence microscopy," J. Appl. Phys. **85,** 6923–6925 (1999

4 Diffractive Read-Out of Optical Discs

Joseph J.M. Braat, Peter Dirksen, Augustus J.E.M. Janssen

4.1 Introduction

An extensive literature is available on optical data storage. Because of the multidisciplinary character of optical storage, the subjects range from optics, mechanics, control theory, electronics, signal coding and cryptography to chemistry and solid-state physics. Several books [2]–[9] have appeared that are devoted to one or several of the subjects mentioned above. This chapter will be limited to purely optical aspects of the optical storage systems. In Section 4.2 of this chapter we first present a short historic overview of the research on optical disc systems that has led to the former video disc system. In Section 4.3 we briefly review the general principles of optical storage systems that have remained more or less unchanged since the start some thirty years ago. An interesting feature of the DVD-system is its standardised radial tracking method that we will treat in some detail in Section 4.4. The appearance of new generations with higher spatial density and storage capacity has triggered solutions for the backward compatibility of the new systems with the existing ones. In Section 4.5 we pay attention to these backward compatibility problems that are mainly caused by the smaller cover layer thickness, the shorter wavelength of the laser source and the higher numerical aperture of the objective in the new generations. Finally, in Section 4.6 we indicate how an efficient modelling of the optical read-out system can be done. Important features like the radial cross-talk and inter-symbol interference can be studied with an advanced model of the optical read-out and the resulting signal jitter, an important quality factor for a digital signal, is efficiently obtained. A new approach is presented that can speed up the calculations so that the modelling can really serve as an interactive tool when optimizing the optical read-out.

4.2 Historic Overview of Video and Audio Recording on Optical Media

Optical data storage uses the principle of the point-by-point scanning of an object to extract the recorded information on the disc. In another context, such an approach was used in a somewhat primitive way in the ‘flying-spot’ film scanner. This apparatus was developed in the early fifties of the last century when a need was felt to transform the pictorial information on film into electrical information to be recorded

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

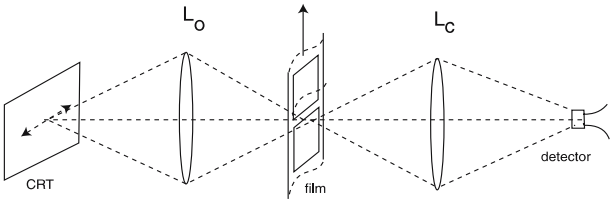


Fig. 4.1. Schematic layout of a flying-spot film scanning device

on magnetic tape. The flying-spot in the film scanner was obtained with the aid of a cathode ray tube and its resolving power was adequate for the resolution encountered on the standard cinema film format. In Fig. 4.1 the focusing lens L_o is shown that produces the scanning spot on the film. The light is partially transmitted and diffracted by the information on the film surface and the detector collects the transmitted light via the collecting lens L_c. The required resolution in the film plane was of the order of 20 to 30 μm and this was not an extreme demand neither for the electron optics nor for the 'objective' lens L_o. The flying-spot scanner was not analysed in depth for some time, it just functioned correctly for the purpose it had been developed for and produced coloured images with the required TV-resolution. In 1961 a patent was attributed to M. Minsky [10] where he describes the principle of a scanning microscope with a resolving power comparable to the standard microscope. Minsky even proposed to use several scanning spots in parallel, thus realising a large-field scanning . The practical use of his (confocal) scanning microscope was limited because of the inefficient use of the power of the (thermal) light source.

A further step towards the optical data storage principle as we know it nowadays was taken by a research group at Stanford University that focused on the recording of video information on a record with the size of a standard audio Long Play disc [11]. The group developed a disc-shaped carrier with data tracks that carried the directly modulated video signal. After a few years the research was abandoned because of the cross-talk problem between the information in neighbouring tracks and because of the bad signal to noise ratio of the detected signal. The belief in a commercial success as a consumer product dwindled away and the sponsor, the 3M-company, stopped funding.

In parallel to this early research on direct signal recording, several groups worked on the principle of storing pictures on a disc, either directly or via holographic encoding. Of course, the latter method was believed to be much more robust because of the far-field recording of the pictorial information. However, one has to conclude that the various prototypes of 'pictorial' video-players that were developed at the end of the sixties and the early 1970's were not satisfactory because of the inefficient use of the available recording surface, the difficult mechanics in the case of the 'direct picture' approach and the colour and speckle problem for the holographic devices.

4.2.1 The Early Optical Video System

At the end of the 1960's, various disciplines and techniques were present at Philips Research Laboratories in the Netherlands that enabled a breakthrough in the field of optical data storage. Important factors enabling this breakthrough were

- availability of expertise in precision optics and control mechanics,
- research on a consumer-type of HeNe-laser where the delicate problems of alignment and stability of such a coherent source were solved in such a way that relatively cheap mass-production could be envisaged,
- work on efficient modulation techniques for video recording on magnetic tape,
- a strong belief in the market for educational applications (teaching, instruction, continuous learning).

In an early stage, some decisions [12] were taken (reflective read-out, information protection via the disc substrate) that have proven to be of vital importance for the success of 'robust' optical data storage. In 1975, several companies (Philips, Thomson [13], Music Corporation of America [14]), later joined by Pioneer, have united to establish a standard for the video disc system (VLP = Video Long Play) so that worldwide disc exchange would be possible. The most important standardised features were

- HeNe laser source, $\lambda = 633\,\text{nm}$, numerical aperture (NA) of the read-out objective equal to 0.40, track pitch of information spiral is $1.6\,\mu\text{m}$, disc diameter is 30 cm, rotational velocity is 25 or 30 Hz depending on the television standard.
- frequency modulation of the analog video signal resulting in optical information pits with a (small) pit length modulation; the average pit length typically is $0.6\,\mu\text{m}$ with comparable 'land' regions in between.
- recording of two video frames per revolution (constant angular velocity, CAV) or recording with constant linear velocity (CLV). The first choice allowed for the still-picture option without using a frame memory (not yet available at that time), the second choice yielded discs with the longest playing time (up to two hours).

Further research on the video disc system (later called the Laser Disc system) has centred on an increase of the recording density, e.g. by

- halving the track pitch to $0.8\,\mu\text{m}$ and keeping the cross-talk at an acceptable level by introducing a two-level pit depth that alternates from track to track [15]
- using (high-frequency) track undulation for storing extra information, e.g. colour and/or sound channels.

Other research efforts have resulted in the introduction of the semiconductor laser ($\lambda = 820\,\text{nm}$) and the simplification of the optics, among others by introducing aspherical optics for the scanning objective [16].

Further potential changes in disc definition resulting from this research on higher density have not led to an improved video disc standard; but the semiconductor laser

source and the cheaper optics have been readily introduced at the beginning of the 1980's.

In parallel to the video application, storage of purely digital data has been pursued in the early stages of optical storage. Various systems using large-size optical discs have been put on the professional market while the recording of data was made possible by the development of high-power semiconductor lasers. In the first instance, the data were irreversibly 'burnt' in the recording layer. Later research has revealed the existence of suitable phase-change materials [17,18] that show an optical contrast between the amorphous and crystalline phase and permit rapid thermally induced switching between both phases. Other optical recording materials rely on the laser-induced thermal switching of magnetic domains [7]; the optical read-out of the domains with opposite vertical magnetisation is done using the magneto-optical Kerr effect.

4.2.2 The Origin of the CD-System

The Laser Disc video system knew a limited success due to its restricted playing time and its lack of the recording option which made direct competition with the well-established video cassette recorder rather difficult. Other opportunities for the optical storage technique were already examined for several years and the replacement of the audio Long Play disc by an optical medium was actively researched. An enormous extra capacity becomes available when switching from the mechanical LP needle to the micron-sized diffraction-limited optical 'stylus'. The approximately hundred times larger storage capacity on an equally sized optical disc has been used in two ways, first by reducing the size of the new medium and, secondly, by turning towards digital encoding of the audio signal which, at that time, asked for a very substantial increase in signal bandwidth. After a certain number of iterations where disc size, signal modulation method, digital error correction scheme and playing time were optimised, the actual Compact Disc standard was established in 1980 by the partners Philips and Sony. The disc diameter of 12 cm was established to accommodate for the integral recording on a single disc of the longest version of the ninth symphony of Beethoven (74 minutes playing time, Bayreuth 1951 recording by the Berliner Philharmoniker, conductor Wilhelm Fuertwangler).

The worldwide market introduction of the CD-system took place at the end of 1982 and the beginning of 1983. Apart from the infrared semiconductor laser source and the corresponding miniaturisation of the optics, no really fundamental changes with respect to the preceding laser disc video system were present regarding the optical read-out principles. The CD system turned out to be simpler and more robust than its video signal predecessor. This was mainly due to the digitally encoded signal on the CD-disc that is much less sensitive to inter-track cross-talk than the former analog video signal.

4.2.3 The Road Towards the DVD-System

For a long period, the application of the CD-system has been restricted to the digital audio domain. In an early stage, a read-only digital data storage system with identical capacity (650 MByte) has been defined (CD-ROM) but the market penetration of this system has been relatively late, some 8 to 10 years after the introduction of the audio CD. Once the personal computer application took off, the demand for recordable and rewritable CD-systems (CD-R and CD-RW) immediately followed. At the same time, the computer environment asked for an ever increasing data capacity to handle combination of pictorial and data content (e.g. for games). Spurred by the quickly developing market of video (using the lower quality digital MPEG1 video standard), a serious need was felt for an improved optical data storage standard. The DVD-standard (see Fig. 4.2) has been defined as of 1994 by a consortium of 10 companies and offers a seven times higher capacity than the CD-format. Because of its multi-purpose application, the acronym DVD stands for Digital Versatile Disc and both MPEG2 video information, super-audio signals and purely digital data can be recorded on this medium. Again, the optical principles used in the DVD-system have not fundamentally changed; a relatively new radial tracking method was introduced and the backward compatibility with the existing CD-family asked for various provisions. In recent years, the DVD-medium has also developed into a family with recordable and rewritable versions. The main innovation in the DVR or Blu-ray Disc system is the use of the deep blue semiconductor laser; combined with a drastically increased value of the numerical aperture of the objective, a forty times higher density with respect to the CD-system is achieved.

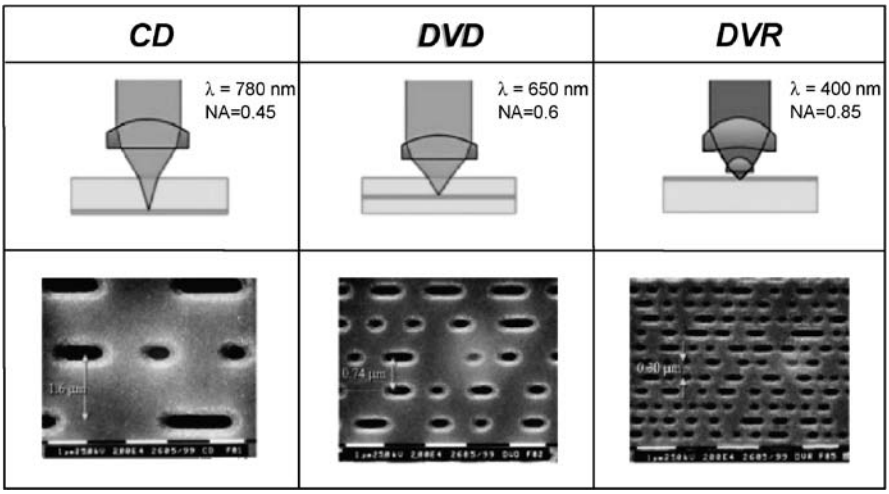


Fig. 4.2. A sketch of the light paths for the CD, DVD and DVR or Blu-ray Disc system and electron microscope photographs with identical magnification of the information pits of the three systems

Even if the basics of optical disc systems have not changed, an enormous effort in research and development has been needed to adapt the media and the playing units to the drastically increased demands with respect to spatial density, scanning speed and data rate. Another less visible development is the digitisation of the opto-mechanical and electronic control systems in the player. Without the easily programmable digitised servo and control systems, the multi-purpose disc player/recorder that easily accepts a broad range of media would be impossible.

In this chapter we will restrict ourselves to the purely optical aspects of an optical disc player/recorder. More general overviews can be found in the various textbooks mentioned in the introductory section.

4.3 Overview of the Optical Principles of the CD- and the DVD-System

In this section we briefly recall the read-out method to extract the stored information from the optical disc. The information is optically coded in terms of a relief pattern on a read-only disc while recordable and rewritable discs employ recording layers that influence both the phase and the amplitude of the incident light. The arrangement of the information is along tracks that have to be followed, both in the recorded and the 'blank' or pre-formatted situation. To this purpose, a bipolar signal is optically derived that can be connected to a servo mechanism to keep the objective head on track during reading or recording. Simultaneously, an error signal is needed to correct for any deviation of the information layer from the optimum focus position.

4.3.1 Optical Read-Out of the High-Frequency Information Signal

In Fig. 4.3, the typical data formats to be read in an optical disc player are shown. The typical track spacing on CD is 1.6 μm and the pit length varies between 0.9 and 3.3 μm with typical increments of 0.3 μm. The data disc in Fig. 4.3 (b) has a different coding scheme that initially looked more appropriate for the hole-burning process by avoiding the length modulation. In Fig. 4.3 (c) the CD-ROM format is present on a phase change disc with the low-contrast regions of micro-crystalline and amorphous structure. The disc in Fig. 4.3 (d) is not compatible with the standard (or DVD) reading principle because of the polarisation-sensitive read-out.

The reading of the information is done with the aid of the scanning microscope principle that is sketched in Fig. 4.4. As it is well-known from the literature, the scanning microscope and the classical microscope are very closely related regarding their resolution and optical bandwidth [19,20]. In terms of maximum transmitted spatial frequency, the classical microscope attains a value of $2\,NA/\lambda$ in the case of incoherent illumination via the condensor system (the numerical aperture of the condensor C then needs to be equal to or larger than the numerical aperture NA of the objective O_1). The scanning microscope achieves the same maximum frequency transfer when the detector D at least captures the same aperture NA as offered by the

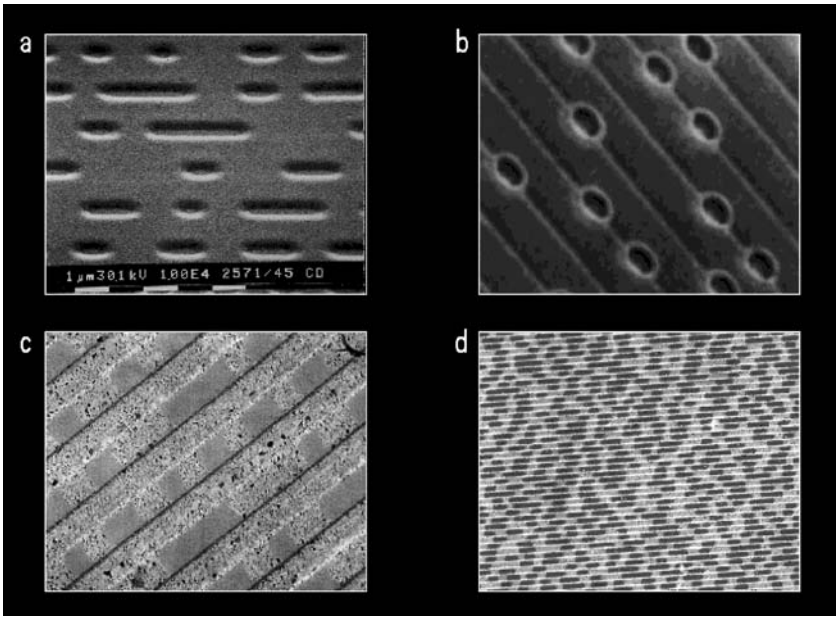


Fig. 4.3. Typical data formats present on optical storage media (**a**) CD-information recorded as a relief pattern in the surface of a disc obtained via injection molding (**b**) 'burned' data on a digital data disc (**c**) amorphous islands (gray) arranged along a pre-formatted track in a micro-crystalline 'sea' (phase-change recording) (**d**) polarisation microscope photograph of a magneto-optical layer with recorded CD-like information

scanning objective O_S (see heavy dashed lines in Fig. 4.4 (a). In the case of a reflective scanning system, the equal aperture of O_S and D is obtained in a straightforward manner because of the double role of O_S as scanning objective and collecting aperture for the detector.

The optical transfer function of a scanning microscope has been visualised in Figs. 4.5–4.6. The diffraction of a plane wave with propagation angle α by a (one-dimensional) periodic structure (frequency $\nu = 1/p$) leads to a set of distinct plane waves with propagation direction angles α_m given by

$$\sin\alpha_m = \sin\alpha + m\lambda\nu \,, \tag{4.1}$$

where m is the order number of the diffracted plane wave.

In the case of an impinging focused wave (aperture angle u), the diffracted field is a superposition of diffracted spherical waves. The incident spherical wave can be considered to be the superposition of a set of plane waves with an angular distribution ranging form $-u$ to $+u$. Note that in the case of a spherical wave, the diffracted orders can not easily be separated in the far field at a large distance from the periodic structure. This is in particular the case for low spatial frequencies and this is in line with the intuitive feeling that a finely focused wave is not capable of sharply

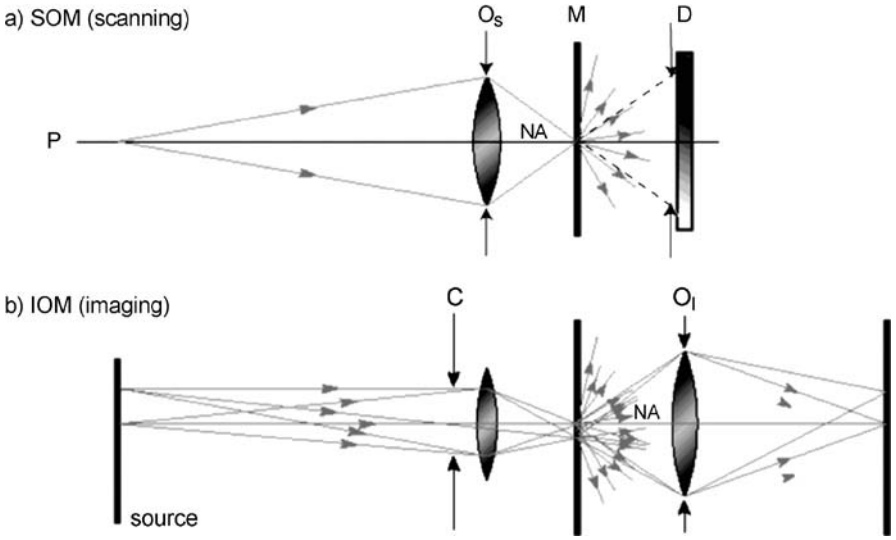


Fig. 4.4. The scanning optical microscope *SOM* (**a**) and the 'classical' imaging optical microscope *IOM* with a full field illumination (**b**). In the scanning microscope, the detector has been schematically drawn and is optically characterised by its angular extent (dashed cone) with respect to the object. In the figure, the angular extent or 'aperture' of the detector has been drawn equal to the numerical aperture *NA* of the objective O_S, a situation that corresponds to standard reflective read-out of an optical disc

discriminating a large period. As soon as the period becomes comparable with the lateral extent of a focused wave (typically of the order of $\lambda/2NA$), the diffracted spherical waves are fully separated. However, we do not need to fully separate the diffracted spherical orders to draw conclusions about the presence and the position of the periodic structure. In the regions where e.g. the spherical waves with order number $m = 0$ and $m = 1$ overlap, the two waves are brought to interference and the phase difference between them will depend on the lateral position of the periodic structure with respect to the focal point of the incident spherical wave (note that the phase of a first diffracted order shows a phase shift ϕ that equals $2\pi\nu\Delta$ if the periodic structure suffers a lateral shift of Δ).

The strength of the interfering signal depends on the size of the overlapping region and on the usefully captured signal by the detector. In the case of a centred detector, the signal depends on its lateral extent and it is normally expressed in terms of the angular beam extent ($NA = \sin u$) of the incident focused beam.

In Fig. 4.6, the normalised frequency transfer function has been depicted as a function of the commonly called coherence factor $\gamma = NA_D/NA$ (NA_D is the numerical aperture of the detector as defined by geometrical conditions or by a collecting optical system in front of the detector, see [19]). The maximum detectable frequency is $2\,NA/\lambda$. In this case ($\gamma \geq 1$), the optical system is approaching the fully incoherent limit which means that from the complex amplitude transmission

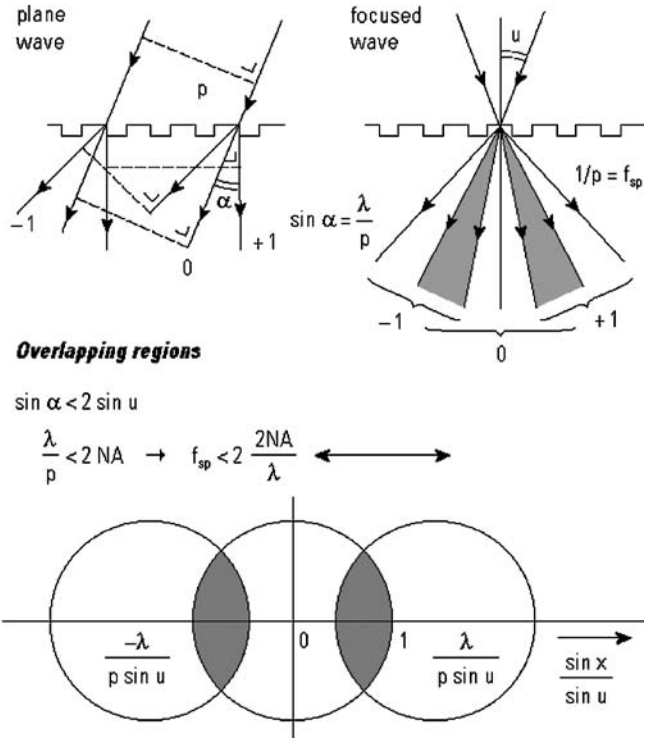


Fig. 4.5. The diffraction of a plane wave and a focused wave by a one-dimensional periodic structure. The useful region in the far-field for signal detection is given by the regions of overlap of zeroth and first orders

function of an object only the modulus part is detectable. At low values of γ, leading to the limit of 'coherent' read-out, the phase part of an object, even if it is a weakly modulated function, can also be detected in principle. Note that on an optical disc the relief pattern mainly affects the phase part of the optical transmission or reflection function. Because of the strong height modulation with respect to wavelength, the phase modulation can also be detected in the (almost) incoherent mode at $\gamma = 1$. If the phase modulation switches between 0 and π, the phase modulation even becomes a pure amplitude modulation with extreme values of ± 1. The choice of the value of γ has been fixed at unity from the very beginning of optical storage. In contrast with e.g. optical lithography where a certain minimum transfer value is required and where it can be useful to trade off the maximum transmitted frequency against a better modulation via a choice $\gamma < 1$, the transfer of an optical disc system can be improved electronically afterwards. This electronic equalisation has proved to be powerful and thus permits to virtually use the entire optical pass-band up to the optical cut-off frequency of $2\,NA/\lambda$; in practice, frequencies turn out to be detectable up to 90% or 95% of the cut-off frequency. The typical optical frequencies

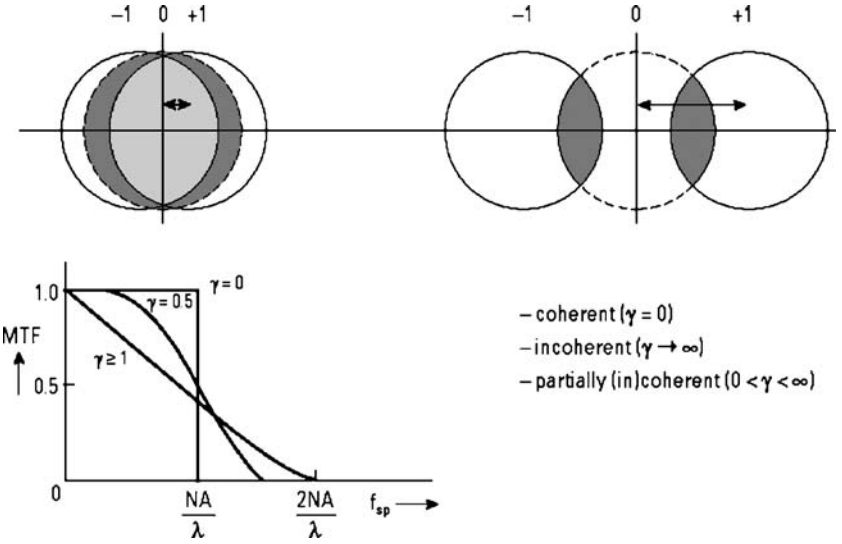


Fig. 4.6. The normalised optical transfer function as a function of the coherence factor γ, determined by the numerical aperture ratio of detector and scanning objective

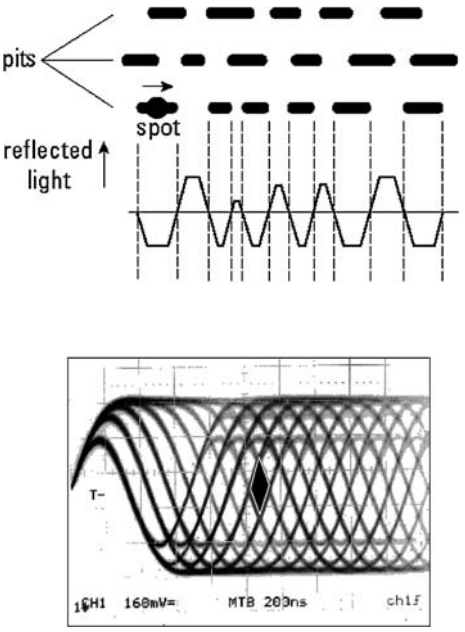


Fig. 4.7. The pit structure on an optical disc and a schematic drawing of the detected signal that shows a typical maximum slope defined by the spot size λ/NA. An oscilloscope image of a large sequence of signal sections is shown below when the scope is triggered on an arbitrary positive transient in the signal; this signal is commonly called the digital 'eye-pattern' and should preferably not contain transients that are close to the centre of a digital eye (see the drawn lozenge in the figure)

in a CD-system ($\lambda = 785\,\text{nm}$, $NA = 0.45$) are 55% of the cut-off frequency for the radial period and approximately 50% for the highest frequency in the direction of the tracks (this is the 3T–3T alternation, the shortest land and pit distances (0.9 μm)

on a CD-disc given the clock length of 0.3 μm). Of course, the binary relief pattern in the track of a CD may contain higher frequencies than this highest fundamental frequency. In the DVD-system ($\lambda = 650\,\text{nm}$, $NA = 0.60$) the optical pass band is used in a more aggressive way and the typical frequencies are shifted to 75% of the cut-off frequency $2\,NA/\lambda$.

The typical signal that is obtained from the high-frequency detector in an optical disc player is shown in Fig. 4.7. A well-opened digital 'eye' is essential for a correct read-out of the stored information.

4.3.2 Optical Error Signals for Focusing and Radial Tracking of the Information

Focus Error Signal Various methods have been proposed for deriving a focus error signal that can be fed to e.g. the coils of a miniature linear electromotor so that an active positioning of the objective with respect to the spinning disc is achieved. In Fig. 4.8 the commonly used astigmatic method [21] is shown. The light reflected

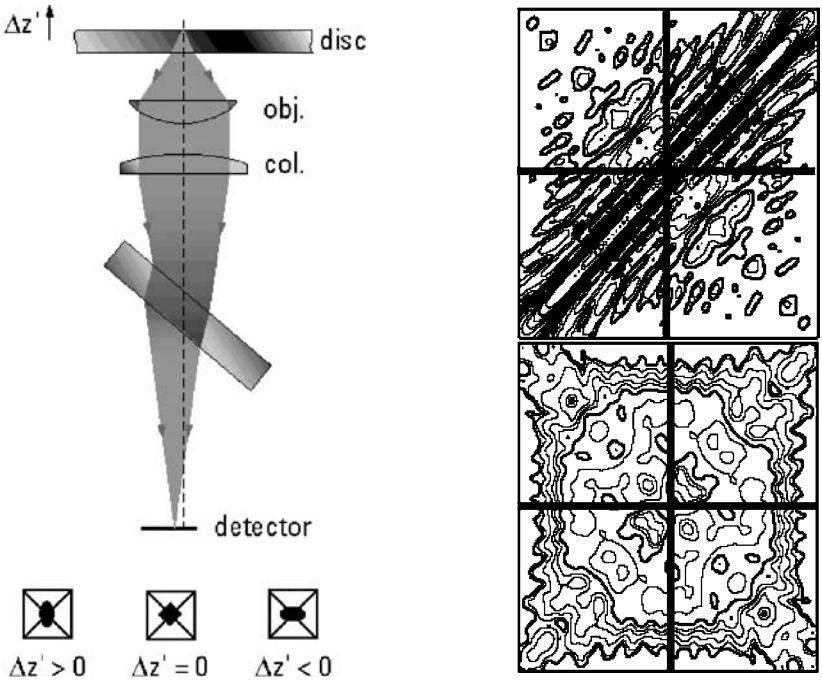


Fig. 4.8. An optical light path using the astigmatic focusing method by exploiting the astigmatism introduced by a plane parallel plate in the returning converging beam (*left*). The calculated diffraction patterns corresponding to a focal line and to the best focus position are depicted in the right-hand part of the figure, together with the separation lines of the focus quadrant detector (*horizontal* and *vertical* in this picture)

from the disc passes through a relatively thick plane parallel plate (the beam splitter on the way forth from the source to the disc). Depending on the convergence angle of the returning beam and on the thickness of the plate, a certain amount of astigmatism is introduced that generates two focal lines along the optical axis with a 'best' focal region in between them. Projected back onto the information surface and taking into account the axial magnification of the combination of collimator and objective, one aims at a distance between the astigmatic lines of some 10 to 15 μm. The best focus position corresponds to a balanced intensity distribution on the four detectors and should be made to coincide with optimum high-frequency read-out of the optical disc. In the right part of Fig. 4.8 a contour picture is shown of the diffraction image in the best focus position and in the case when one of the line images is focused on the detector.

Another candidate for the detection of a focus defect is the Foucault knife-edge test. Although its sensitivity can be higher than that of the astigmatic method, the position tolerances of the knife edge are somewhat critical. A circularly symmetric version of the knife-edge method is the 'ring-toric' lens method [22]. Another option is the 'spot-size' detection method that measures the variation of beam cross-section as a function of defocus. All these methods are capable of delivering a signal with changing polarity as a function of the sign of the focus error. Another important property of an error signal is the maximum distance through which the signal is well above the noise, the so-called effective acquisition range. For both methods, a typical value of 30 to 50 μm can be expected. More refined methods use the detected high-frequency information to monitor the absolute focus-setting and to correct for any drift in the detector balance or the mechanical settings of the light path.

Radial Error Signal The methods indicated in Fig. 4.9 have been devised for the detection of the radial position of the reading or recording light spot with respect to the centre of a track. The first method uses two auxiliary light spots generated by means of a grating (twin-spot method [23]). The light spots have a quarter track spacing off-set with respect to the central light spot. The reflected light belonging to each auxiliary spot is collected in the focal region on separate detectors and the difference in average signal from each detector provides the desired bipolar tracking signal that is zero on-track. An AC-variant of this method with the auxiliary spots uses a small lateral oscillation of the main spot itself. The detection of the phase (either 0 or π) of the oscillation in the detected AC-signal yields the required tracking error signal. The oscillation can be imposed on the scanning spot in the optical player but, preferably, a so-called track wobble is given to the track with a known phase so that no active wobble is needed in the player. Of course, the track wobble should not generate spurious information in the high-frequency signal band nor should it spoil other information (track number, address information) that is stored in an 'undulation' or lateral displacement of the information track. Both methods described above retrieve the tracking information from a refocused light distribution on one or more detectors that are optically conjugate with the information layers and

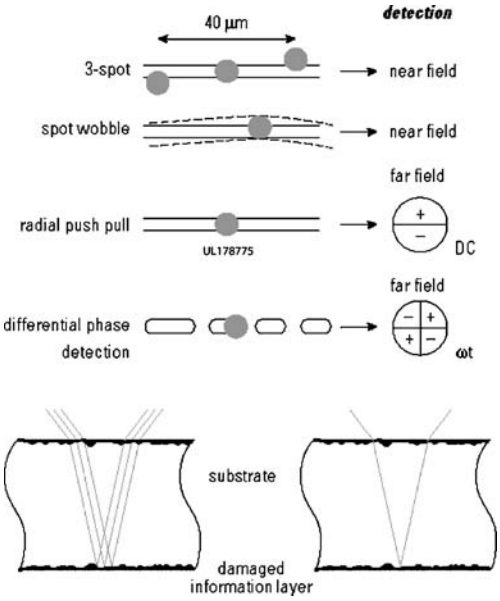


Fig. 4.9. The various options for deriving a bipolar optical tracking signal. The label 'near field' here means that the detector arrangement is positioned in the image plane of the information layer. In the case of 'far field' detection, the detector is effectively located in the limiting aperture of the objective (or in an image of this diaphragm)

for this reason they are called near-field methods. A well-known far-field method is the so-called Push-Pull method [21] that uses a duo-detector in subtractive mode for radial tracking (see Fig. 4.10). The method is based on the fact that in the presence of a tracking error (the centre of the scanning spot hits e.g. the edges of the information pits) the average propagation direction is slightly bent away from the perpendicular direction. For relatively small values of the phase shift of the light reflected by the pits, the difference signal of the duo-detector shows a change of sign when the scanning spot crosses the information track. If we define the phase depth $\Delta\phi$ of the information pits as the phase shift suffered by the light on reflection from the information pits with respect to the light reflected from non-modulated land regions, we observe that the difference signal of the duo-detector reduces to zero if $\Delta\phi$ equals π. The obvious reason is that a phase shift of π between pits and land regions has reduced the information structure to effectively an amplitude structure with no directional phase information available. For this reason, the Push-Pull method is critical with respect to the phase shift induced by the information pits and it also behaves in a rather complicated way once the optical 'effects' have a combined amplitude and phase character (phase-change recording material). A far-field method that, in a first instance, does not suffer from the restrictions on the phase depth $\Delta\phi$ is the differential phase (or time) detection where the phase relations between signal components from a far-field quadrant detector at the instantaneous signal frequency ω are used to derive a tracking signal [24]. A tracking method based on this principle has been standardised for the DVD-system and will be discussed in more detail in a following section.

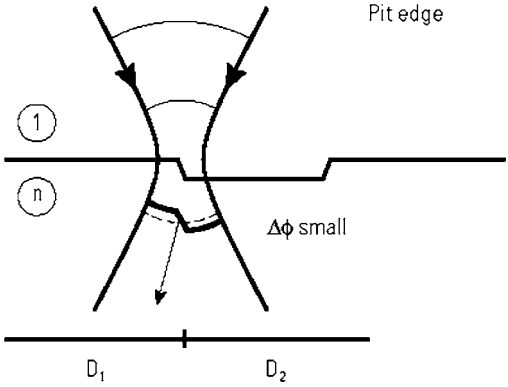


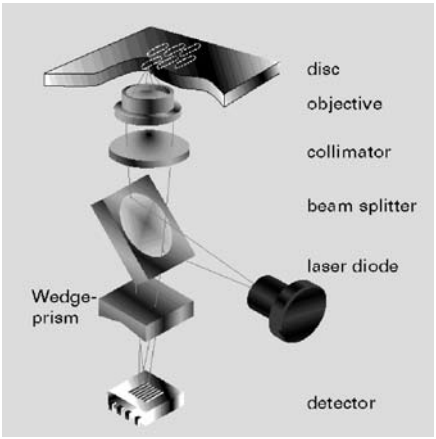
Fig. 4.10. The phase change introduced by a pit edge to the focused wave (here shown in transmission). The average propagation direction of the transmitted beam is slightly tilted causing unbalance between the detectors D_1 and D_2

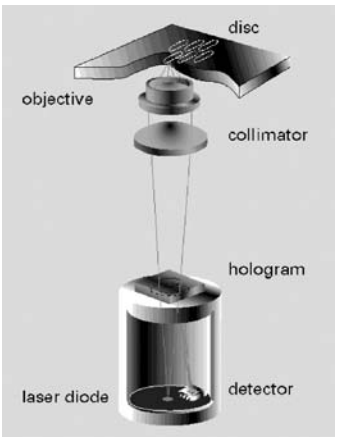
The possible perturbations of the near-field or far-field methods is correlated with the conjugate position of the respective detectors. In Fig. 4.9, lower part, the near-field methods should be sensitive to damage of the information layer while the far-field based methods should be more sensitive to damage or scratching of the upper surface (non-info) of the disc. Because of the limited bandwidth of the tracking error signals, their (average) information is collected from a relatively large disc area (a stripe on the info-layer, an elongated circle on the non-info surface). Because of the spatial averaging effect, no substantial difference of preference is found for the near- or far-field methods.

4.3.3 Examples of Light Paths

The various methods for deriving the high-frequency signal and the optical error signals have to be incorporated in a light path that has wide tolerances both during manufacturing and product life (optical damage, thermal and mechanical drifts). In Fig. 4.11, left-hand side, a classical light path for CD is shown with the light from the laser diode source coupled in via a semi-transparent plate towards the collimator-objective combination. The reflected light is transmitted through the beam splitter towards a double-wedge prism that separates the far field region in order to accommodate the detection of a Push-Pull radial tracking signal. The double wedge also serves as 'knife-edge' for each beam half thus producing two complementary knife-edge images for focus detection. This light path is made from simple, easily available components. The tightest tolerance regarding temperature and mechanical drift is found in the relative position of source and detector with respect to the beam splitter.

In the right-hand figure, the beam splitting action is obtained with the aid of a holographic element. The relatively thick substrate carries a grating on the lower surface for generating the three spots needed for radial tracking. The upper surface contains a more complicated structure, the holographic or diffractive element. At the way forth to the disc, the three beams generated by the grating continue as the zeroth order of the holographic structure and they are captured by the collimator and objective. Consequently, three light spots are focused on the disc. On the way back,





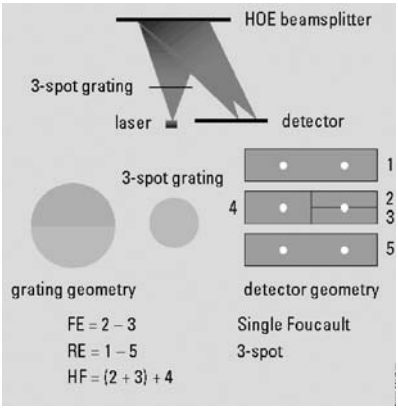


Fig. 4.11. An optical light path using separate optical components for achieving the beam separation and the generation of the focus and radial error signals (left). In the right-hand picture, the splitting function and the generation of the error signals is obtained by means of holographic or diffractive optical element. In the lower picture, the various beams are shown that propagate towards the detector after traversal of the holographic beam splitter *HOE*

the light diffracted by the hologram propagates towards the detector region and a far-field splitting is obtained by sending one half circle of the beam cross-section to the detectors 2 and 3 and the other half to detector 4. The auxiliary spots produced by the grating are directed towards the detectors 1 and 5. A focus error signal is obtained by taking the difference between detectors 2 and 3 while the Push-Pull radial tracking error signal can be obtained by the signal combination (2 + 3) – 4. Moreover, an extra twin-spot radial error signal is obtained from the detector difference signal 1–5 (the radial offset of each spot is one quarter of a pitch). The high frequency signal is derived from the sum of the detectors 2, 3 and 4. The attractiveness of the hologram solution is its mechanical and environmental stability. The directions of the diffracted orders are rather insensitive to hologram tilt and displacement and the

source and detector are mounted on one single substrate. A wavelength change will shift the diffracted spots along the dividing line between detectors 2 and 3 but does not cause an offset.

4.4 Radial Tracking for DVD

Some disadvantages of the radial tracking methods had become manifest while using them for the CD-system. The twin-spot method requires an extra grating with a rather delicate angular orientation and, in the recording mode, precious optical power (some 20%) is spread out to the auxiliary twin spots. The Push-Pull method is sensitive to the phase depth $\Delta\phi$ of the information and does not work in the presence of pure amplitude information patterns. The spot or track wobble method, due to its very nature, consumes precious space in the radial direction and is likely to increase the cross-talk between tracks when compared to the other methods at equal density. For this reason, the Differential Phase Detection (*DPD*) method [24], originally devised for the analog video disc system, has been revisited to adapt it to the digital signals on the high-density DVD disc. It is now commonly called the Differential Time Detection (*DTD*) method.

4.4.1 A Diffraction Model for the DPD and DTD Tracking Signal

An easy picture for the understanding of the differential phase (or time) detection method is obtained when the information in the tangential direction is replaced by a single frequency, so that, together with the radial periodicity, a purely two-dimensional grating is present on the disc. In Fig. 4.12 the various quasi-DC and AC signals are depicted that are present in the far-field on a quadrant detector (quadrants A, B, C and D). Each modulation term depends on the diffraction direction. Radially diffracted orders show a (varying) intensity level in an overlapping region with the zeroth order that is proportional to $\cos(\psi_{10} \pm \phi_\mathrm{r})$. The reference phase ψ_{10} between the first and zeroth orders is determined by the disc structure; it varies from $\pi/2$ for very shallow structures to π if the phase depth $\Delta\phi$ of the optical effects attains the value of π. The quantity $\phi_\mathrm{r} = 2\pi r/q$ is proportional to the tracking error r (q is the track pitch). The sign of ϕ_r depending on the order number ± 1 of the diffracted order. In the same way, the overlapping regions in the tangential direction show an intensity variation according to $\cos(\psi_{10} \pm \omega t)$, caused by the scanning with uniform speed of the tangentially stored periodic pattern. Obliquely diffracted orders show a mixed phase contribution and are carrying the interesting information for deriving the tracking error.

Collecting the interference terms on the various quadrants, we obtain

$$
\begin{aligned}
S_\mathrm{A}(t, \phi_\mathrm{r}) &= \cos(\omega t + \psi) + \alpha \cos(\omega t - \phi_\mathrm{r} + \psi) \\
S_\mathrm{B}(t, \phi_\mathrm{r}) &= \cos(\omega t - \psi) + \alpha \cos(\omega t + \phi_\mathrm{r} - \psi) \\
S_\mathrm{C}(t, \phi_\mathrm{r}) &= \cos(\omega t - \psi) + \alpha \cos(\omega t - \phi_\mathrm{r} - \psi) \\
S_\mathrm{D}(t, \phi_\mathrm{r}) &= \cos(\omega t + \psi) + \alpha \cos(\omega t + \phi_\mathrm{r} + \psi) \,, \qquad (4.2)
\end{aligned}
$$

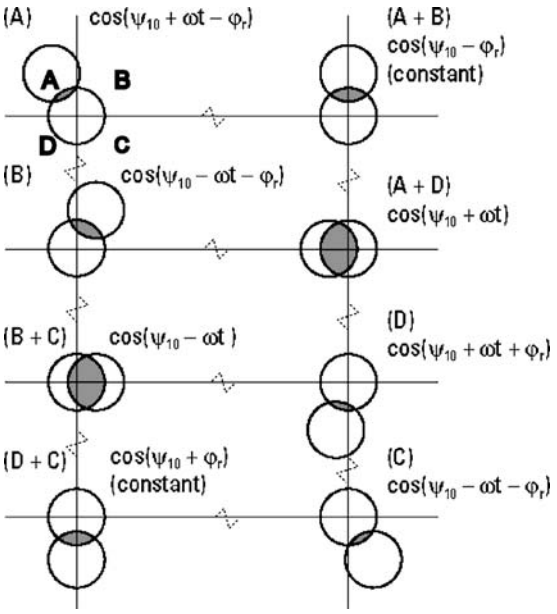


Fig. 4.12. Overlapping regions in the far-field quadrants A, B, C and D with typical modulation terms in the DC-domain and in the high-frequency domain (frequency ω). In each of the eight examples shown, the letters are printed (between brackets) of the quadrants in the far field where a certain harmonic modulation term of the signal is detected. The dependence on the disc translation is given by the term ωt and a radial displacement by the phase ϕ_r. If no translation term ωt is present, the signal only shows variations of average intensity (denoted by 'constant') due to radial position. The tangential direction is chosen horizontally in the figure

where all possible reference phase angles between zeroth and first diffracted orders, for reasons of simplicity, have been replaced by a single phase angle ψ and where α is a factor less than unity that accounts for the relatively small contribution of the diagonal orders with respect to the tangential orders.

With this approximation, the various high-frequency signals are now given by:

$$
\begin{aligned}
S_{\mathrm{CA}}(t,\phi_\mathrm{r}) &\propto +(1+\alpha\cos\phi_\mathrm{r})\cos\psi\cos(\omega t)\\
S_{\mathrm{tPP}}(t,\phi_\mathrm{r}) &\propto +(1+\alpha\cos\phi_\mathrm{r})\sin\psi\sin(\omega t)\\
S_{\mathrm{rPP}}(t,\phi_\mathrm{r}) &\propto +\alpha\sin\phi_\mathrm{r}\sin\psi\cos(\omega t)\\
S_{\mathrm{dPP}}(t,\phi_\mathrm{r}) &\propto -\alpha\sin\phi_\mathrm{r}\cos\psi\sin(\omega t)\quad .
\end{aligned}
\tag{4.3}
$$

The subscripts here denote the signals according to the detector schemes Central Aperture (CA $= A + B + C + D$), tangential Push-Pull (tPP $= A - B - C + D$), radial Push-Pull (rPP $= A + B - C - D$) and diagonal Push-Pull (dPP $= A - B + C - D$).

For the derivation of a tracking signal, a reference signal with a well-defined time-dependent phase is needed for comparison with a signal whose phase depends on the tracking error. As a reference signal we can choose the Central Aperture (CA)

signal or the tangential Push-Pull signal. If we take the CA-signal and multiply it with the dPP-signal after a phase shift of the latter over $\pi/2$ we obtain after low-pass filtering:

$$S_1(\phi_r) \propto -\alpha \cos^2\psi \left(\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r \right). \tag{4.4}$$

The multiplication of the tangential PP-signal and the diagonal PP signal directly yields after low-pass filtering:

$$S_2(\phi_r) \propto -\alpha \sin\psi \cos\psi \left(\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r \right). \tag{4.5}$$

Inspection of the signals $S_1(\phi_r)$ and $S_2(\phi_r)$ shows that they provide us with the required bipolar signal of the tracking error ϕ_r. The factor in front, containing ψ, depends on the phase depth $\Delta\phi$ and makes $S_2(\phi_r)$ less apt for discs with relatively deep optical structures while $S_1(\phi_r)$ performs optimum in that case.

Other possible combinations are obtained when combining the high-frequency radial Push-Pull signal with either the CA- or the tPP-signal. One easily deduces that the following signals result after multiplication, phase-shifting and low-pass filtering of, respectively, rPP and CA $S_3(\phi_r)$ and rPP and tPP Radial $S_4(\phi_r)$:

$$S_3(\phi_r) \propto \alpha \sin\psi \cos\psi \left(\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r \right). \tag{4.6}$$

$$S_4(\phi_r) \propto \alpha \sin^2\psi \left(\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r \right). \tag{4.7}$$

The ψ-dependence of especially $S_4(\phi_r)$ makes it unsuitable for standard optical discs but especially favourable for extra shallow disc structures.

4.4.2 The Influence of Detector Misalignment on the Tracking Signal

The single-carrier model can be equally used for studying the effects of misalignment of the quadrant detector with respect to the projected far-field pattern, the so-called beam-landing effect. This effect is produced during quick access of a remote information track. The sledge (coarse movement) and the scanning objective (precise positioning) are actuated together and this action produces a lateral displacement of the objective with respect to the fixed collimator and detector that can amount to some 10% of the beam foot-print on the detector. As a consequence, the whole diffraction pattern is shifted on the detector in the radial direction (see the arrow 'spot offset' in Fig. 4.13). Taking into account such a mechanically induced offset with a size of ϵ (normalised with respect to the half-diameter of the beam foot-print), we write the signals corresponding to the interference terms in the over-

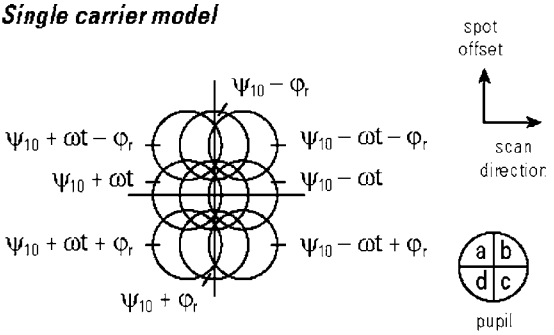


Fig. 4.13. Various methods for the extraction of a Differential Time radial error signal using the signals present in the far-field quadrants

lapping regions of the far field as

$$
\begin{aligned}
S_{\mathrm{A}}(t,\phi_{\mathrm{r}}) &= (1+\epsilon)\cos(\omega t+\psi)+\alpha\cos(\omega t-\phi_{\mathrm{r}}+\psi)\\
S_{\mathrm{B}}(t,\phi_{\mathrm{r}}) &= (1+\epsilon)\cos(\omega t-\psi)+\alpha\cos(\omega t+\phi_{\mathrm{r}}-\psi)\\
S_{\mathrm{C}}(t,\phi_{\mathrm{r}}) &= (1-\epsilon)\cos(\omega t-\psi)+\alpha\cos(\omega t-\phi_{\mathrm{r}}-\psi)\\
S_{\mathrm{D}}(t,\phi_{\mathrm{r}}) &= (1-\epsilon)\cos(\omega t+\psi)+\alpha\cos(\omega t+\phi_{\mathrm{r}}+\psi)\quad .
\end{aligned}
\tag{4.8}
$$

The beam-landing effect has only been applied to the purely tangential diffraction orders; the diagonal orders are rather far away from the horizontal dividing line of the quadrants.

Two of the four basic signals derived from the quadrants A, B, C and D depend on the radial misalignment and they become:

$$
\begin{aligned}
S_{\mathrm{rPP}}(t,\phi_{\mathrm{r}}) &\propto \alpha\sin\phi_{\mathrm{r}}\sin\psi\cos(\omega t)-\epsilon\cos\psi\cos(\omega t)\\
S_{\mathrm{dPP}}(t,\phi_{\mathrm{r}}) &\propto -\alpha\sin\phi_{\mathrm{r}}\cos\psi\sin(\omega t)-\epsilon\sin\psi\sin(\omega t)\quad .
\end{aligned}
\tag{4.9}
$$

The four possible DPD-signals now become

$$
S_1(\phi_{\mathrm{r}}) \propto -\alpha\cos^2\psi\left(\sin\phi_{\mathrm{r}}+\frac{1}{2}\alpha\sin 2\phi_{\mathrm{r}}\right)-\epsilon\sin\psi\cos\psi(1+\alpha\cos\phi_{\mathrm{r}})
$$

$$
S_2(\phi_{\mathrm{r}}) \propto -\alpha\sin\psi\cos\psi\left(\sin\phi_{\mathrm{r}}+\frac{1}{2}\alpha\sin 2\phi_{\mathrm{r}}\right)-\epsilon\sin^2\psi(1+\alpha\cos\phi_{\mathrm{r}})
$$

$$S_3(\phi_r) \propto \alpha \sin\psi \cos\psi \left(\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r\right) - \epsilon \cos^2\psi(1 + \alpha\cos\phi_r)$$

$$S_4(\phi_r) \propto \alpha \sin^2\psi \left(\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r\right) - \epsilon \sin\psi\cos\psi(1 + \alpha\cos\phi_r) \quad . \tag{4.10}$$

The signals S_1 and S_4 show a beam landing sensitivity that is maximum for the $\lambda/4$ pit depth ($\psi \approx 3\pi/4$); for deep pits (optical depth is $\lambda/2$) or for amplitude structures ($\psi = \pi$) the beam landing sensitivity is zero.

An interesting combination of the signals S_1 and S_4 is given by

$$S_5(\phi_r) = S_1(\phi_r) - S_4(\phi_r) \propto \left(\sin\phi_r + \frac{1}{2}\alpha \sin 2\phi_r\right) \quad . \tag{4.11}$$

The detection scheme can be written as follows:

$$S_5(\phi_r) = [B - C][A + D]^{90^\circ} - [A - D][B + C]^{90^\circ} \quad , \tag{4.12}$$

where the index 90° implies that the 90 degrees phase shifted version of the signal has to be taken. The signal S_5 is fully independent of pit depth and does not suffer at all from beam landing off-set.

Another possible linear combination is given by

$$\begin{aligned} S_6(\phi_r) &= S_2(\phi_r) - (1 + s)S_3(\phi_r) \propto \\ &\propto -(1 + s/2)\sin\psi\cos\psi\left(\sin\phi_r + \frac{1}{2}\alpha\sin 2\phi_r\right) \\ &\quad +\epsilon/2\left[\sin^2\psi - (1 + s)\cos^2\psi\right](1 + \alpha\cos\phi_r) \quad . \end{aligned} \tag{4.13}$$

The beam landing influence on this signal is reduced to zero by the condition

$$s = \tan^2\psi - 1 \quad . \tag{4.14}$$

The detection scheme for S_6 is

$$S_6(\phi_r) = [CD - AB] + (s/4)\left[(C + D)^2 - (A + B)^2\right] \quad . \tag{4.15}$$

In the particular case that $\psi = 3\pi/4$, the signal is free from beam landing effects when using its simplest form $[CD - AB]$; the corresponding tracking signal is obtained after low-pass filtering of this high-frequency signal product.

4.4.3 The DTD Tracking Signal for the DVD-System

Shifting from analog to digital signals with their associated clock reference signal, it is evident that the measured phase shifts will now be replaced by time differences between the clock signals associated with specific digital signals derived from the quadrant detector.

The principle of differential time detection can be explained with the aid of Fig. 4.14. The focused light spot (centre in M) is scanning an information pit slightly

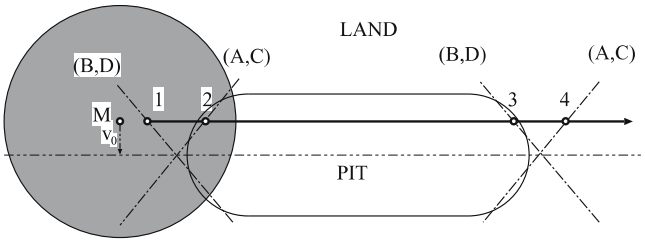


Fig. 4.14. A schematically drawn focused spot (gray area) is scanning an optical pit with a tracking error v_0. The positions 1 and 3 mark the front and end of the pit as they are detected by the detector pair (B, D); the positions 2 and 4 correspond with the beginning and end of the pit when detected by the diagonal detector pair (A, C)

off-track (distance v_0) in the direction of the arrow. The leading and the trailing edge of the pit induce diffraction of the light perpendicularly to the hatched lines (A, C) and (B, D). When the centre M of the scanning spot is in the position 1, the detector pair (B, D) perceives the leading edge of the pit and the intensity on this detector pair

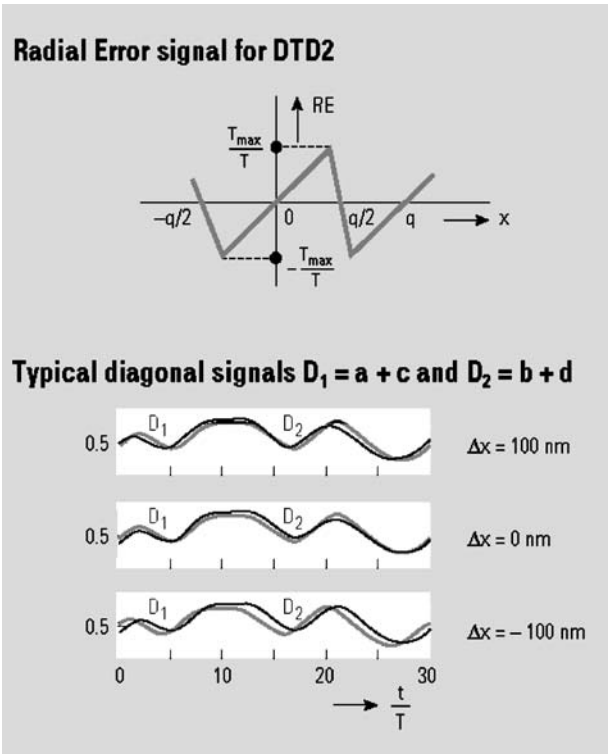


Fig. 4.15. The mutual shift in time of the diagonal quadrant signals D_1 and D_2 as a function of the tracking error and the resulting radial error signal, schematically drawn in the upper part of the figure as a function of the radial off-track position with periodicity q

will go down. A short time later, the detector pair (A, C) will measure the passage of the leading edge because the intensity goes down at the position 2. Once the scanning spot has advanced towards the pit end, the trailing edge will be detected at the positions 3 and 4 by a rise in intensity on respectively the detector pairs (B, D) and (A, C). By measuring the time difference between the information signal on the diagonal detector pairs, a bipolar function of the tracking error v_0 can be obtained. Further away from the edges, in the longer pit sections and land regions, the summed intensities on the two quadrant pairs are virtually equal and there is no contribution to the DTD-signal.

The standard DTD-signal is derived by first summing the intensities on a diagonal pair of detector and then detecting the time shift by electronic means. It is also possible to compare the phase difference between individual quadrants and we conclude that the following set of detection schemes is possible:

- $\tau_{(A+C)} - \tau_{(B+D)}$
  The time difference between the sum signal from quadrants $A + C$ and the sum signal from quadrants $B + D$ is taken. This is the standard DTD-method based on a diagonal difference signal and the denomination is *DTD2*-signal.
- $(\tau_B - \tau_C) + (\tau_D - \tau_A)$
  This difference signal requires four independent high-frequency detectors; on track, each difference signal $(D - A)$ and $(B - C)$ is identical zero. The common denomination is *DTD4*-signal
- $(\tau_B - \tau_A) + (\tau_D - \tau_C)$
  This signal is comparable to the preceding one; the time differences $B - A$ and $D - C$ are not zero on track but depend, among others, on the optical pit depth.

4.4.4 The DTD2 and the DTD4 Signal in the Presence of Defocus

In the case of the standard analog *DPD*-signal, we have seen that the phase reference value ψ between the zeroth order and the first order diffracted light played an important role via a $\cos^2 \psi$-dependence. It is a well-known fact that the effective phase difference between zeroth and first order light is affected by a defocusing. Zernike's discovery of phase contrast was triggered by his observation that a phase grating image showed beautiful contrast in a defocused setting. This focus-dependence of ψ means that the the amplitude of the standard *DPD*-signal but also the *DTD2*-signal will vary when defocus is present. If $\psi = \pi$, the signal amplitude is symmetric around focus and the maximum is found in the optimum focus setting. If $\psi < \pi$, the tracking signal will show a maximum value in a defocused situation and can even become zero in focus. Special *DPD*-signals ($S_5(\phi_r)$ and $S_6(\phi_r)$) remained unaffected by the value of ψ and they also show a behaviour that, in a first instance, is independent of defocus. It turns out that for the *DTD4*-signal, where the time difference is evaluated independently for each quadrant detector, a comparable robustness holds with respect to defocus. In Fig. 4.16 the *DTD2* and *DTD4* signals have been depicted for a rather shallow disk pattern with DVD-density. It is clear from the figure that there is a pronounced asymmetry around focus for the *DTD2*-signal while the

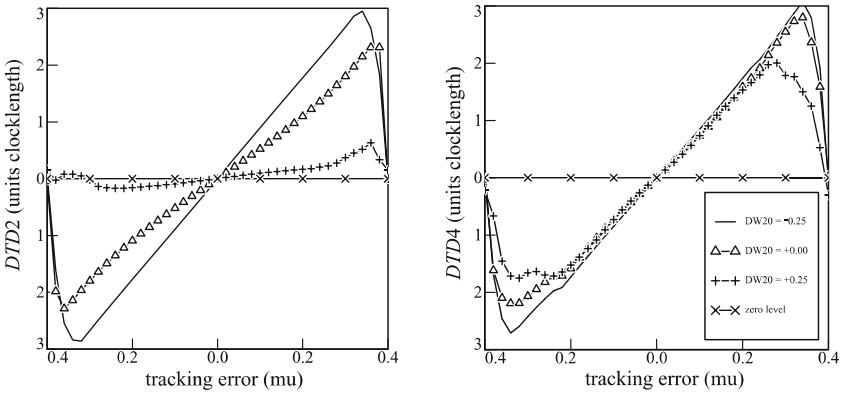


Fig. 4.16. The radial tracking signals *DTD2* and *DTD4* with different amounts of defocus. The defocus parameter *DW20* is expressed in wavefront aberration at the pupil rim in units of the wavelength λ of the light. A value of 0.25 corresponds to a defocusing Δz_s of one focal depth (typically $0.8\,\mu m$ for a DVD-system). The tracking error signal has been numerically computed using a long sequence of pits obeying the standard EFM (*E*ight to *F*ourteen *M*odulation) scheme [25]. The disk layout and the reading conditions are those encountered in the DVD-system but the depth of the information pits has been reduced to 60 nm ($\psi \approx 3\pi/4$) instead of the common 100 nm. *drawn line*: Δz_s=-0.8 μm, *triangles*: Δz_s=0, *crosses*: Δz_s=+0.8 μm

DTD4-signal virtually remains unaffected by defocus; in particular the slope at the central zero set-point of the servo system remains unaltered on defocusing.

We conclude by reminding that the introduction of the DTD-tracking method in the CD-system was hampered by a signal instability that appeared when neighbouring track portions showed comparable information with a fixed mutual position over a distance that was within the bandwidth of the radial servo system. An example of such a problematic situation was found in the fixed digital pattern used to represent absolute silence on a CD audio disc. In the DVD-system, this fixed pattern problem has been avoided by an appropriate extra coding step that suppresses the prolonged appearance of fixed patterns.

4.5 Compatibility Issues for the DVD-and the CD-System

The compatibility issues between the two standards *CD* and *DVD* are caused by the change in substrate thickness, the change in wavelength and, to a lesser degree, the change in numerical aperture. The combination of smaller wavelength and higher *NA* would lead to a density increase by a factor of 2.5. Further gain in density has been obtained by the higher frequency choice in the DVD-system with respect to the available optical pass-band; this has led to a total factor of five. The final factor of 7.2 has been achieved by more efficient modulation schemes and error correction techniques.

In Table 4.1 the specifications of both systems have been tabulated and we observe that the substrate thickness has been halved for *DVD*. The reason for this reduction in thickness was the specification on disc tilt during playback. The higher *NA* and shorter wavelength of the DVD-system causes a two times larger comatic aberration of the read-out spot at equal tilt angle. The amount of coma being linearly dependent on the substrate thickness, a reduction of thickness was required. The 600 μm compromise between tolerances and mechanical stability was solved by bonding two substrates together, leading to a doubled capacity. As it appears from Fig. 4.17, the doubled capacity can be further increased by introducing in both

Table 4.1. Table with the most relevant specifications of the *CD*- and the *DVD*-system

CD and DVD Specifications

	CD	**DVD**
Disc diameter	120 mm	120 mm
Disc thickness	1.2 mm	1.2 mm
Disc structure	Single substrate	Two bonded 0.6 mm substrates
Laser wavelength	780 nm (infrared)	650 and 635 nm (red)
Numerical aperture	0.45	0.60
Track pitch	1.6 μm	0.74 μm
Shortest pit/land length	0.83 μm	0.4 μm
Reference speed	1.2 m/sec CLV	4.0 m/sec CLV
Data layers	1	1 or 2
Data capacity	Approx. 680 Mbytes	Single layer: 4.7 Gbytes Dual layer: 8.5 Gbytes
Reference user rate	Mode 1: 153.6 kbytes/sec Mode 2: 176.4 kbytes/sec	1.108 kbytes/sec 1.1 Mbytes/sec

Video format

	Video CD	**DVD-video**
Video data rate	1.44 Mbits/sec (video, audio)	1 to 10 Mbits/sec variable (video, audio, subtitles)
Video compression	MPEG1	MPEG2
Sound tracks	2 Channel-MPEG	Mandatory (NTSC) 2-channel linear PCM; 2-channel/5.1-channel AC-3 Optional: up to 8 streams of data available
Subtitles	Open caption only	Up to 32 languages

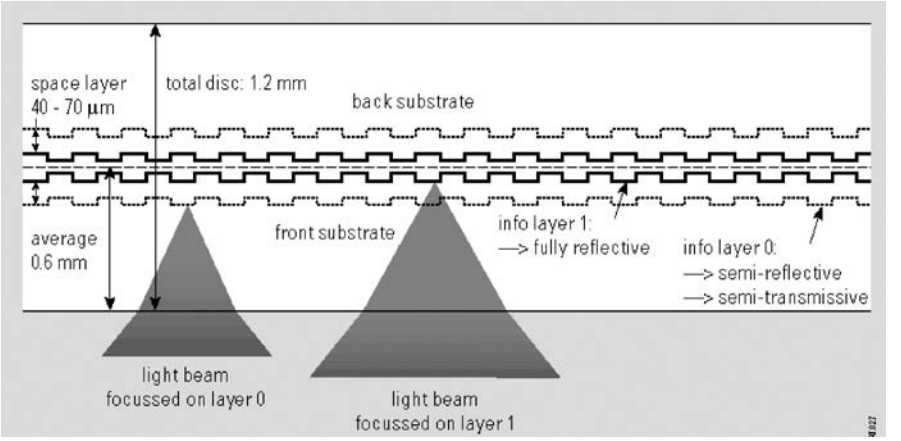


Fig. 4.17. Cross-section of a double-sided DVD-disc with an extra semi-transparent information layer (dotted) on each side

substrates a buried information layer that is approximately semi-transparent. The distance between the buried layer and the fully reflecting information layer is specified in a margin of 40 to 70 µm to avoid a too large value of spherical aberration when reading the layers at an *NA* of 0.60 in red light (λ=650 nm).

4.5.1 The Substrate-Induced Spherical Aberration

In a DVD player, the most critical task is the reading of the high-density DVD disc with an *NA* value of 0.60 and with light of a wavelength of typically 650 nm. Although a CD or CD-ROM disc could also be read with the same wavelength, this does not necessarily apply to a CD-R or a CD-RW disc. The correct optical contrast between recorded optical effects and the land regions is only guaranteed at the original CD-wavelength of 780 nm. The reading of a CD-type disc thus first requires an extra source and, secondly, a solution needs to be found for the reading through a substrate that is 600 µm thicker than prescribed in the specification for the DVD-objective. The thicker substrate causes spherical aberration in the CD-reading spot due to the fact that the marginal rays are more refracted than the paraxial rays, thus enlarging the focal distribution in the axial direction away from the objective. Simultaneously, a serious lateral smearing out of the light intensity distribution is observed and the correct reading of the CD-information is prohibited by this aberration phenomenon. In Fig. 4.18, images of intensity distributions at various focus settings are shown for a focused wave that suffers from spherical aberration. The picture at the height of the chain-dotted line is found at the axial position corresponding to the paraxial focus; approximately two pictures up, the marginal focus position is found where the aberrated marginal rays cut the optical axis. The amount of wavefront spherical aberration W_S, written as a power series of the numerical aperture ($NA = \sin\alpha$), and normalised with respect to the wavelength of the light is

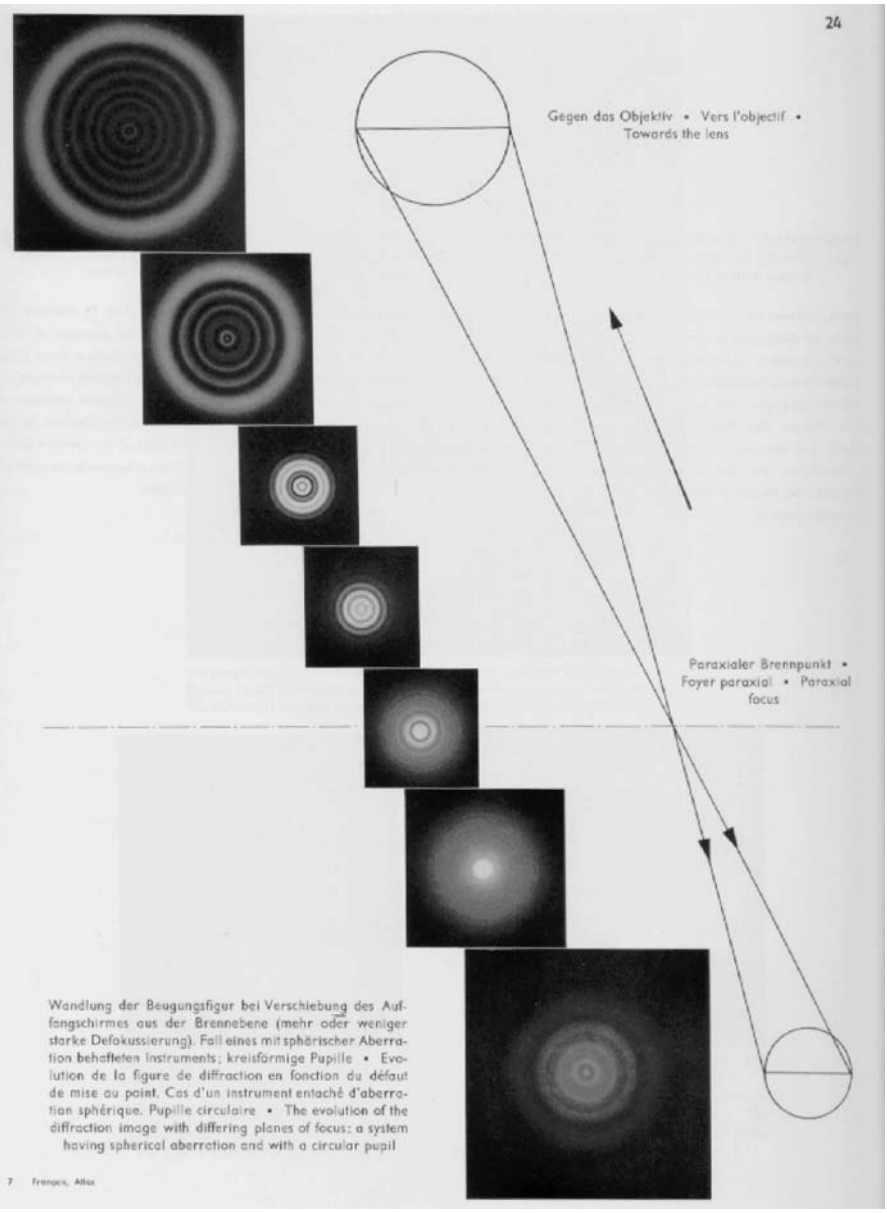


Fig. 4.18. Intensity profiles at different positions along the optical axis for a pencil suffering from spherical aberration (from: M. Cagnet, M. Françon, J.-C. Thrierr: *Atlas of optical phenomena* (Springer-Verlag, Berlin 1962)). In the case of a DVD-player reading a CD-disc, the extra substrate thickness introduces spherical aberration with a sign such that the lower part of the figure is closest to the objective and the light travels in the direction given by the large arrow in the upper part of the figure

given by

$$W_S = \frac{d}{\lambda}\left\{\left(\frac{n^2-1}{8n^3}\right)\sin^4\alpha + \left(\frac{n^4-1}{16n^5}\right)\sin^6\alpha + \left(\frac{5(n^6-1)}{128n^7}\right)\sin^8\alpha + \cdots\right\}, \quad (4.16)$$

where n is the refractive index of the disc substrate material and d the thickness deviation.

If we use the numbers $\lambda = 650\,\text{nm}$, $d = 600\,\mu\text{m}$, $n = 1.586$ (refractive index of polycarbonate) and $\sin\alpha = 0.60$, the numerical values of the three factors in the power series of $\sin\alpha$ in (4.16) amount to, respectively, 5.68, 1.43 and 0.36 in units of wavelengths. At a first sight, knowing that optical disc read-out has to be done in diffraction-limited conditions, these values are far too large. In Fig. 4.19 we have depicted the behaviour of the central intensity along the axis and in Fig. 4.20 we present calculated intensity distributions at axial positions between the paraxial and the 'best' focus (halfway the paraxial and the marginal focus). From Fig. 4.20, it is obvious that the reading of a CD-disc is impossible close to the 'best' focus setting (e.g. using the light distribution from the second row, last picture) . The reading spot is very much degraded and almost all energy has drifted towards strong diffraction rings; at read-out, the diffraction rings will yield an unacceptable intersymbol interference and the digital eye pattern fully disappears. However, the asymmetry with respect to the 'best' focus position shows a peculiar effect: towards the paraxial region, the strong diffraction rings disappear and a central lobe is observed superimposed on a reduced intensity background. The central lobe is much broader than in the best focus but could have an acceptable size for reading a low-density CD-disc when the background intensity is somehow made inoffensive by the read-out method. The distribution that most resembles a standard reading spot (appropriate half width, low level of side lobes) is found around the third position of Fig. 4.20. The focus setting corresponds to an offset from the paraxial focus of some 6 to 8 focal depths (one focal depth $\approx 0.8\,\mu\text{m}$) and yields an intensity distribution that

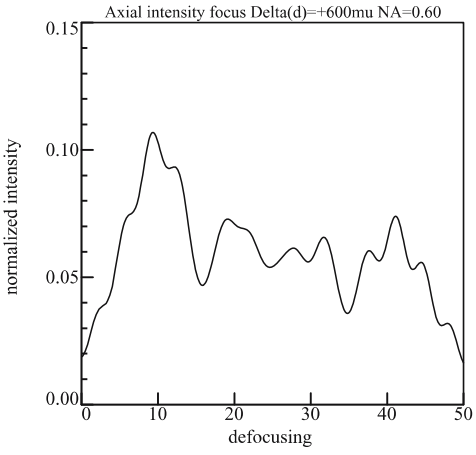


Fig. 4.19. A plot of the axial intensity of a pencil suffering from spherical aberration due to a substrate thickness increase of $600\,\mu\text{m}$ ($NA = 0.60$, $\lambda = 650\,\text{nm}$). The paraxial focus is at $z = 0$, the best focus at $z = 24\,\mu\text{m}$ and the marginal focus at $z = 48\,\mu\text{m}$

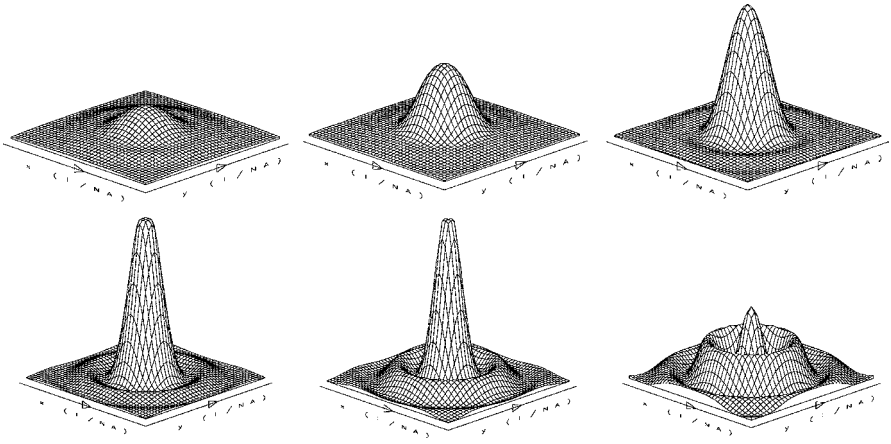


Fig. 4.20. 3D plots of the intensity distribution (from left to right) going away from the paraxial focus towards the best focus in axial steps of 3 μm (one focal depth equals 0.81 μm ($NA = 0.60$, $\lambda = 650$ nm). The maximum intensity in the figures is about 0.10 with respect to 1.0 for the diffraction-limited case. The length and width of the square plot area are $4\lambda/NA$ (4.33 μm). The upper right figure represents the optimum focus setting for reading a CD-disc

should be adequate for CD read-out. In Fig. 4.19, this interesting axial position is approximately halfway up-hill the first maximum.

Given the low peak intensity, an important amount of low-intensity aberrated stray light can be expected around the central peak and this background should be prevented from reaching the detector. This can be done by introducing an aperture in the light path towards the detector, or, alternatively, by using a detector of finite size; the strongly aberrated rays, after double passage through the mismatched substrate, will have a large transverse aberration in the image plane and miss the quadrant detector. The effect on the calculated digital eye pattern of the effective reduction of the detector aperture NA_D is illustrated in Fig. 4.21. These graphs show the resulting digital eye pattern when the detection diameter in the far field is reduced from e.g. 80% to 55% at a focus setting with $z = +6\,\mu$m from the paraxial focus P. In the case of a digital disc system, the quality of the detector signal is generally assessed by means of the so-called root mean square digital jitter. This quantity is obtained by binarising the analog signal read from a disc and comparing it with the desired binary wave form. The root mean square value of the shifts of the transients in the detected signal with respect to the corresponding positions in the originally recorded signal is called the digital jitter of the signal. The digital jitter Δ_d is expressed in the time domain or in the spatial domain and counted in units of the clock length T of the digital signal. In the right-hand figure, the residual jitter amounts to 6.7%; this value is obtained in the case of a maximum decorrelation between neighbouring tracks. When the information between the tracks is fully correlated, the jitter goes down to 3% and in practice one finds values of 5.5 to 6% in optimum CD read-out conditions.

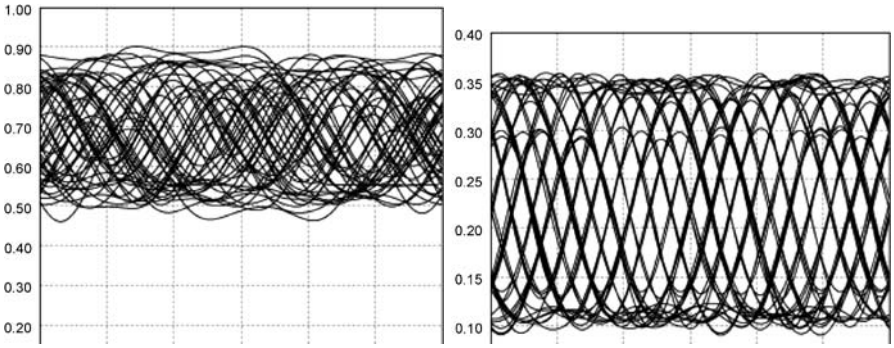


Fig. 4.21. A plot of the digital eye that is detected when focusing at a position 6 μm away from the paraxial focus (left-hand figure: far-field detection region is 80% of full aperture; right-hand figure: 55% of the aperture is effectively used for detection. Note the loss in signal strength in the second picture

A further reduction in jitter is possible when the residual spherical aberration present over the central area is corrected in the objective by tailoring the surface profile of the first aspheric surface of the single element objective lens. In this case, a wavefront correction is applied to the central 55% region while the outer region is left unaltered and this leads to a further reduction of the jitter in the detected signal down to 5% for a CD-disc.

4.5.2 The Effective Optical Transfer Function

The changes at read-out by a reduction of the far-field detection area to e.g. 55% of the total beam diameter can be expressed in terms of a modified MTF (modulation transfer function) of the optical read-out system. A reduction of the detection aperture is equivalent to a more coherent or partially coherent detection of the information on the disc and this increases the response at the mid-spatial frequencies although, simultaneously, the cut-off frequency of the optics is reduced. This latter effect is not very important when reading a CD-disc at $NA = 0.60$ and $\lambda = 650\,\text{nm}$ because the cut-off frequency is far too high ($1850\,\text{mm}^{-1}$ instead of $1150\,\text{mm}^{-1}$ for a standard CD-player).

In Fig. 4.22 we observe that at the optimum focus setting the resulting MTF (curves with triangles) is not much different from the standard MTF encountered in a classical CD-player with $NA = 0.45$ and $\lambda = 785\,\text{nm}$ (the cut-off frequency would be found at an axial value of 1.15).

In the figure we also show the slight improvement that is possible when the residual spherical aberration present over the central area of the reading beam is corrected in the objective (drawn line). Some 5 to 10% improvement in MTF-value is possible at the mid CD-frequencies and this becomes visible also in the reduction of the bottom jitter. Moreover, the absence of residual spherical aberration over the

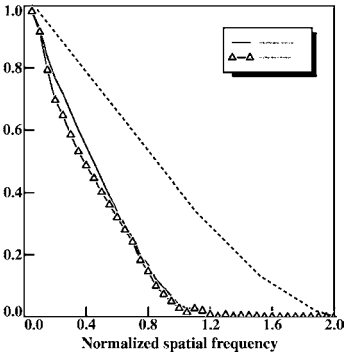


Fig. 4.22. MTF-curves of the read-out with reduced *NA* value at the detection side. The horizontal axis is expressed in units NA/λ. In the figure the optimum MTF obtained by selecting a relative aperture of 55% in the reflected beam (*triangles*) is further improved by introducing a correction for the residual spherical aberration over the central section of the objective (*drawn line*). For comparison, the optimum transfer function at full detection aperture has been sketched (*dashed line*)

effective detection area leads to a better symmetry of the read-out signal around the optimum focus setting.

4.5.3 The Two-Wavelength Light Path

As we remarked before, recordable or rewritable CD-discs require a read-out at the appropriate wavelength of $\lambda = 780 - 840\,\text{nm}$ and the corresponding numerical aperture amounts to 0.45. The partially coherent read-out at $NA_\text{D} = 55\%$ that was possible at the red wavelength with a slight objective modification at the most, does not yield a sufficiently large optical bandwidth at $\lambda = 780\,\text{nm}$. A solution is found by a further tailoring of a ring section of the surface profile of the first aspheric surface of the objective, thereby reducing the spherical aberration at CD read-out. The objective then can be considered to consist of three regions. The central paraxial region contributes to both DVD and CD read-out. A relatively narrow annular region corrects the CD read-out but is effectively lost for the DVD read-out. The outer region is only useful for the DVD read-out; this part of the read-out bundle becomes heavily aberrated at CD read-out and is then lost for detection [26]. The rather low light efficiency at CD read-out that was observed in Fig. 4.21 does not occur once a second independent source is used for CD read-out. The far-field of the second source can be adjusted to the maximum aperture ($0.45 \leq NA \leq 0.50$) needed for CD read-out.

Other compatibility solutions have been proposed based on a diffractive structure applied to the first aspheric surface of the objective and capable of correcting the spherical aberration in the first diffraction order [27]. More advanced solutions have recently been described in [28].

4.6 Efficient Calculation Scheme for the Detector Signal

In this section we will treat the problem of the efficient calculation of the detector signal in an optical disc player. The reliable and numerically efficient modelling of detected signals is a very welcome addition to the experimental verification of new

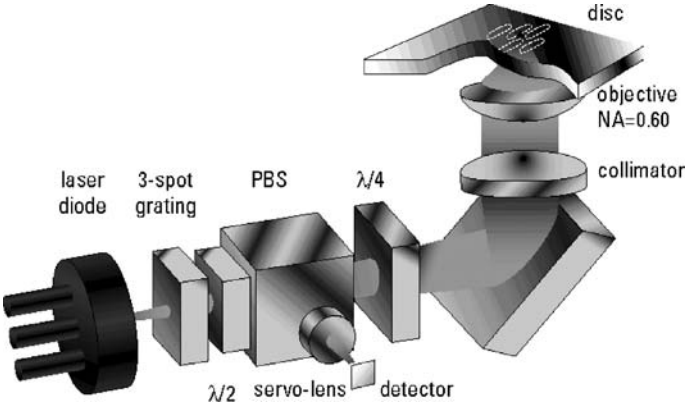


Fig. 4.23. The layout of an optical light path for DVD and CD with the possibility to inject an extra read/write beam of different colour via a second beam splitter (not shown in the figure), either before the beam splitter or in the parallel beam section after the collimator. The polarisation-sensitive light path with quarter-wave plate and polarising beam splitter (*PBS*) is used to maximise writing power

ideas. With a reliable modelling tool, one obtains a quick estimate of tolerances and robustness of new options for recording schemes and read-out methods. In this section we present a numerical procedure that carries the mathematical analysis of the read-out process as far as possible and thus gains precious time as compared to the purely numerical approach. Especially in the case of optics with general pupil shape and aberrations, the analytical approach can bring an important reduction in calculation time.

4.6.1 Optical Configuration and the FFT-Approach

The light path of an optical disc player is well represented by a scanning microscope of Type I according to the terminology given in [20]. In the schematic drawing of a light path according to Fig. 4.24, the detector could also be positioned at the location of the aperture of O' and in many modelling approaches, this is effectively done to minimize the numerical burden. But the way in which the high-frequency and the optical error signals are derived is more complicated and the diffraction step from the aperture of O' to the composite detectors in plane D is essential if a detailed knowledge of the intensity distribution on the detectors is needed. As an example we take the hologram beamsplitter of Fig. 4.11 where each beam propagating to the detector plane effectively originates from one half of the collecting aperture at O'. Another example is the light path introducing astigmatism in Fig. 4.8. The intensity distribution on the detector in best focus is some 20 times larger in linear measure as compared to the standard diffraction image. It is important to know the distribution of the high-frequency signal in the various quadrants in both the in-focus situation and the defocused case.

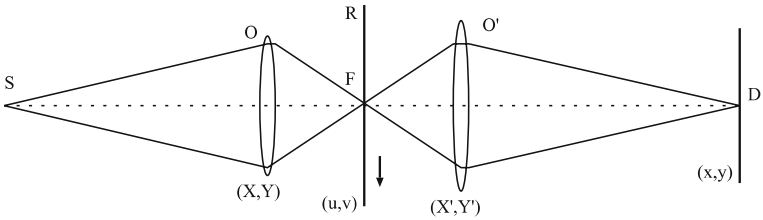


Fig. 4.24. Schematic drawing illustrating the progation of the light from the source S via the objective O towards the disc surface in the focal plane F of O. The disc presents a (spatially modulated) reflection function $R(u, v)$. After reflection the diffracted light is captured by the same objective (here denoted by O') and an image of the disc is produced at the location of the detector plane D. The pupil coordinates (X, Y) and (X', Y') are commonly normalized with respect to the half-diameter of the objective aperture. The disc coordinates (u, v) and the detector coordinates (x, y) are normalized with respect to the diffraction units λ/NA in both planes

The standard way to calculate a detector signal is the repeated application of a propagation step from an object or image plane to an aperture and vice versa, in optical terms by propagating from the near field to the far field and back. The degree of sophistication with which the propagation is carried out is not crucial in the sense that each propagation step can be reduced to a Fourier transform. Especially when a scalar diffraction approach is sufficient, a single Fourier transform operation per step is adequate. For an efficient numerical implementation, the Fourier integral is replaced by a Fast Fourier Transform (FFT) that is faster with respect to execution time.

The assessment of the digital signal is carried out by means of the digital jitter Δ_d. As a rule of thumb, one can say that the *rms* jitter preferably should not exceed 10% of the clock length. At a value of 15%, prolonged over large signal portions, the digital signal fully breaks down and cannot be reconstructed without errors. The fact that some statistics are needed to calculate a reliable jitter value means that an extended track portion with a digitally modulated signal has to be scanned. For a reliable reconstruction of the optical signal, the sample points where the detector signal is calculated should be rather densely spaced, e.g. with a spatial increment of 0.05 to 0.1 of the diffraction unit λ/NA. For each sample point at the disc, the two FFT-steps from the disc to the aperture O' and to the detector plane D have to be executed, provided that the incident amplitude distribution at the focal point F is stored in memory. Several thousands of FFT-steps can be needed for obtaining one single value of Δ_d.

4.6.2 The Analytic Approach

The analysis given in this subsection uses the material presented in references [29] and [2]. In order to accommodate for more general signal sequences without loosing the advantage of dealing with periodic structures, we define a larger unit cell

(length p) that contains several optical effects in the tangential direction and, if needed, several neighbouring tracks in the radial direction with a formal radial period of q (see Fig. 4.25).

In the presence of the double periodic disc structure, the complex amplitude in the entrance aperture of the collecting objective O' (see Fig. 4.24) is given by the sum of a certain number of diffracted waves according to

$$A'(X', Y') = \sum_{m,n} \rho_{m,\hat{n}} \exp\left\{2\pi \mathrm{i}\left[\left(\frac{m}{p}\right)u + \left(\frac{\hat{n}}{q}\right)v\right]\right\} f\left(X' - \frac{m}{p}, Y' - \frac{\hat{n}}{q}\right) \quad , \qquad (4.17)$$

where m is the tangential order number and the quantity $\hat{n}$ equals the radial order number $(n - ms/p)$ of a disc with a track-to-track shift s of the information. The argument of the exponential function is proportional to the mutual displacement (u, v) of the scanning spot and the disc structure in, respectively, the tangential and radial direction. The factor $\rho_{m,\hat{n}}$ is equal to the complex amplitude of the diffracted

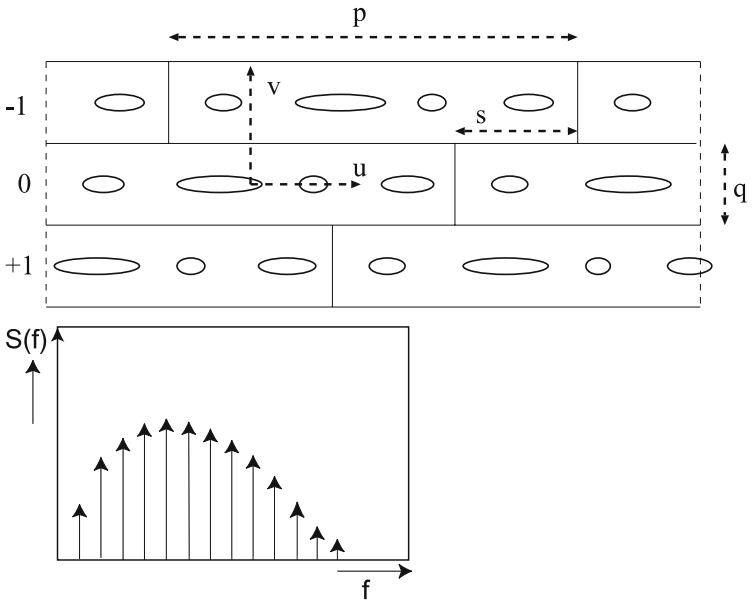


Fig. 4.25. The disc structure possesses a double periodicity. The fundamental frequency in the radial direction is given by $1/q$ with q in principle being equal to the pitch of the information track. The periodicity in the tangential direction is a 'synthetic' one. Within one period of length p several optical effects are present that obey the modulation rule of the optical structure with respect to the lengths of the optical effects and the land sections in between. The tangential information is repeated in the neighbouring tracks but a shift (distance s) has been provided to avoid directly correlated cross-talk. The tangential repetition with period p leads to a typical sampled power spectrum $S(f)$ of the digital information on the disc (lower graph); the distance between the frequency sample points along the normalised frequency axis is given by $\Delta f = \lambda/p(NA)$

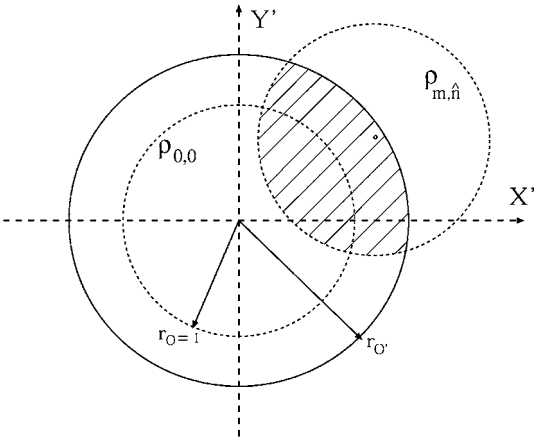


Fig. 4.26. Position of the diffracted orders of the disc structure with double periodicity in the exit pupil of the collecting objective O' with half diameter $r_{O'}$. The zeroth order with complex amplitude $\rho_{0,0}$ has been shown and a general diffraction order $(m, \hat{n})$. The hatched area indicates which part of the order $(m, \hat{n})$ is transmitted to the detector plane

wave with the corresponding order number $(m, \hat{n})$. The complex function $f(X', Y')$ stands for the lens function and carries the information about the lens transmission function (e.g. pupil shape, Gaussian filling factor of the incoming beam) and the lens aberration. Figure 4.26 shows the zeroth diffracted order and a general diffracted order (dotted circles) as they are located in the pupil of the collecting objective O'. The coordinates (X', Y') in the exit pupil are linearly related to the sines of the angles in the far-field of the disc structure if the collecting objective O' satisfies the sine condition of geometrical optics [30].

The imaging step from the exit pupil of O' to the detector plane is another Fourier transform from the coordinates (X', Y') to the detector plane coordinates (x, y). The lens function of O' is denoted by $g(X', Y')$; in the special case of a reflective system, we put $g(X', Y') = f(-X, -Y)$. The complex amplitude $A''(X', Y')$ in the exit pupil of the collecting objective becomes the product of $A'(X', Y')$ and the lens function $g(X', Y')$. If the disc is uniformly rotated, resulting in a local linear speed of s_0, the period $1/p$ is transformed into a temporal frequency $f_t = s_0/p$. We finally find for the complex amplitude in the detector plane

$$
\begin{aligned}
B(x, y; t, v) &= \iint_{O'} A''(X', Y') \, \exp\{2\pi \mathrm{i}\,(X'x + Y'y)\} \, \mathrm{d}X' \mathrm{d}Y' \\
&= \sum_{m,n} \rho_{m,\hat{n}} \, \exp\left\{2\pi \mathrm{i}\left[m\left(f_t t - \frac{sv}{pq}\right) + \frac{nv}{q}\right]\right\} \\
&\times \iint_{O'} g(X', Y') f\left(X' - \frac{m}{p}, Y' - \frac{n - \frac{ms}{p}}{q}\right) \\
&\times \exp\{2\pi \mathrm{i}\,(X'x + Y'y)\} \, \mathrm{d}X' \mathrm{d}Y' \quad , \qquad (4.18)
\end{aligned}
$$

where the integral is formally carried out over the full area of the exit pupil of O'.

In (4.18) we separate the integral over the exit pupil area O', that only depends on the imaging properties of the lenses, from the part that is scanning-dependent and carries information about the structure of the disc via the factors $\rho_{m,\hat{n}}$. The expression

for $B(x, y; t, v)$ now reads

$$B(x,y;t,v) = \sum_{m,n} \rho_{m,\hat{n}} \exp\left\{2\pi \mathrm{i}\left[\left(m\left(f_{\mathrm{t}}t - \frac{sv}{pq}\right) + \frac{nv}{q}\right)\right]\right\} F_{m,\hat{n}}(x,y) \ , \tag{4.19}$$

where the factor $F_{m,\hat{n}}(x, y)$ equals the integral over X' and Y' in (4.18).

The detected quantity is the intensity, obtained by multiplying $B(x, y; t, v)$ with its complex conjugate and this yields

$$\begin{aligned} I(x,y;t,v) &= |B(x,y;t,v)|^2 = \\ &= \sum_{m,n} \sum_{m',n'} \rho_{m,\hat{n}}\, \rho^*{}_{m',\hat{n}'}\ F_{m,\hat{n}}(x,y)\ F^*{}_{m',\hat{n}'}(x,y) \\ &\times \exp\left\{2\pi \mathrm{i}\left[(m-m')\left(f_{\mathrm{t}}t - \frac{sv}{pq}\right) + \frac{(n-n')v}{q}\right]\right\} \ . \end{aligned} \tag{4.20}$$

In an optical disc system we use special detector geometries, e.g. a quadrant detector. The detector signal is obtained by integrating the intensity over the detector area, taking into account a locally varying detection sensitivity function $S(x, y)$ and this yields a k^{th} detector signal according to

$$\begin{aligned} S_{\mathrm{D}_k}(t,v) &= \sum_{m,n} \sum_{m',n'} \rho_{m,\hat{n}}\, \rho^*{}_{m',\hat{n}'} \\ &\times \exp\left\{2\pi i\left[(m-m')\left(f_{\mathrm{t}}t - \frac{sv}{pq}\right) + \frac{(n-n')v}{q}\right]\right\} \\ &\times \iint_{\mathrm{D}_k} S(x,y)\ F_{m,\hat{n}}(x,y)\ F^*{}_{m',\hat{n}'}(x,y)\mathrm{d}x\mathrm{d}y \ , \end{aligned} \tag{4.21}$$

where D_k denotes the area of the k^{th} detector. In the expression for the detector signal we have separated the quantities related to the disc structure from the integral that only contains information on the imaging process by the objective and collecting lens. This means that one single effort in evaluating the integral in (4.21) can serve for the many times repeated calculation of detector signals corresponding to different structures on the disc that fit into the basic period of length p. In this way we can gather statistical data on the digital jitter Δ_{d} with a seriously reduced numerical effort; in practice, a factor of the order of hundred is observed.

4.6.3 The Harmonic Components of the Detector Signal

With the double periodicity present on the disc, the detector signal can be written as a double Fourier series where the uniform scanning speed s_0 transforms the periodic components in the tangential direction into time harmonic components with a fundamental frequency $f_{\mathrm{t}} = s_0/p$.

Using (4.21), we formally obtain the temporal harmonic components $A_{\mu,k}$ and $B_{\mu,k}$ from detector D_k according to

$$
\begin{aligned}
A_{0,k}(v) &= P_{1,k}(0,0) \\
&\quad +2\sum_{\kappa=1}^{\kappa_{\max}} \left\{ P_{1,k}(0,\kappa)\cos\left(2\pi\frac{\kappa}{q}v\right) + P_{2,k}(0,\kappa)\sin\left(2\pi\frac{\kappa}{q}v\right)\right\} \\
A_{\mu,k}(v) &= 2\left\{ P_{1,k}(\mu,0)\cos\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right) + P_{2,k}(\mu,0)\sin\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right)\right\} \\
&\quad +2\sum_{\kappa=1}^{\kappa_{\max}} \left\{ \left[\{P_{1,k}(\mu,\kappa)+Q_{1,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right.\right. \\
&\qquad \left. +\{P_{2,k}(\mu,\kappa)+Q_{2,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right]\cos\left(2\pi\frac{\kappa}{q}v\right) \\
&\qquad + \left[\{P_{2,k}(\mu,\kappa)-Q_{2,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right. \\
&\qquad \left.\left. -\{P_{1,k}(\mu,\kappa)-Q_{1,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right]\sin\left(2\pi\frac{\kappa}{q}v\right)\right\} \\
B_{\mu,k}(v) &= 2\left\{ -P_{2,k}(\mu,0)\cos\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right) + P_{1,k}(\mu,0)\sin\left(2\pi\frac{s}{p}\frac{\kappa}{q}\mu\right)\right\} \\
&\quad +2\sum_{\kappa=1}^{\kappa_{\max}} \left\{ \left[\{-P_{2,k}(\mu,\kappa)-Q_{2,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right.\right. \\
&\qquad \left. +\{P_{1,k}(\mu,\kappa)+Q_{1,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right]\cos\left(2\pi\frac{\kappa}{q}v\right) \\
&\qquad + \left[\{P_{1,k}(\mu,\kappa)-Q_{1,k}(\mu,\kappa)\}\cos\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right. \\
&\qquad \left.\left. +\{P_{2,k}(\mu,\kappa)-Q_{2,k}(\mu,\kappa)\}\sin\left(2\pi\frac{s}{p}\frac{v}{q}\mu\right)\right]\sin\left(2\pi\frac{\kappa}{q}v\right)\right\} \quad , \qquad (4.22)
\end{aligned}
$$

where the harmonic coefficients A and B, that depend on the off-track distance v of the scanning spot, generate the detector signal with the aid of the expression

$$
S_{\mathrm{D}_k}(t,v) = A_{0,k}(v) + \sum_{\mu=1}^{\mu_{\max}} \left\{ A_{\mu,k}(v)\cos(\mu f_{\mathrm{t}} t) + B_{\mu,k}(v)\sin(\mu f_{\mathrm{t}} t)\right\} \quad , \qquad (4.23)
$$

where $\mu_{\max} f_{\mathrm{t}}$ is the maximum frequency transmitted by the optical read-out system.

The calculation of the detector signal $S_{\mathrm{D}_k}(t,v)$ requires the evaluation of the coefficients $P_{j,k}(\mu,\kappa)$ and $Q_{j,k}(\mu,\kappa)$, $j=1,2$, that can be written as

$$
\begin{aligned}
P_{1,k}(\mu,\kappa) &= \mathrm{Re}\left\{ \sum_m \sum_n \rho_{\mu+m,\hat{\kappa}+\hat{n}}\, \rho^*{}_{m,\hat{n}}\, Z_k(\mu+m,\hat{\kappa}+\hat{n};m,\hat{n})\right\} \\
P_{2,k}(\mu,\kappa) &= \mathrm{Im}\left\{ \sum_m \sum_n \rho_{\mu+m,\hat{\kappa}+\hat{n}}\, \rho^*{}_{m,\hat{n}}\, Z_k(\mu+m,\hat{\kappa}+\hat{n};m,\hat{n})\right\}
\end{aligned}
$$

$$Q_{1,k}(\mu,\kappa) = \mathrm{Re}\left\{\sum_m \sum_n \rho_{\mu+m,\hat{n}}\ \rho^*{}_{m,\hat{\kappa}+\hat{n}}\ Z_k(\mu+m,\hat{n};m,\hat{\kappa}+\hat{n})\right\}$$

$$Q_{2,k}(\mu,\kappa) = \mathrm{Im}\left\{\sum_m \sum_n \rho_{\mu+m,\hat{n}}\ \rho^*{}_{m,\hat{\kappa}+\hat{n}}\ Z_k(\mu+m,\hat{n};m,\hat{\kappa}+\hat{n})\right\} \quad . \tag{4.24}$$

The factor $Z_k(m,\hat{n};m',\hat{n}')$ is the integral that was present in (4.21) and is written

$$Z_k(m,\hat{n};m',\hat{n}') = \iint_{\mathrm{D}_k} S(x,y)\ F_{m,\hat{n}}(x,y)\ F^*{}_{m',\hat{n}'}(x,y)\mathrm{d}x\mathrm{d}y \quad . \tag{4.25}$$

The efficient calculation of the Z_k-functions is the subject of the remainder of this chapter. Especially the numerical evaluation of the basic function $F_{m,n}(x,y)$ can be time consuming and we will show how a further analysis of this function can improve the efficiency of the final evaluation.

4.6.4 The Representation of the Function $F_{m,n}(x,y)$

The function $F_{m,n}(x,y)$ is the Fourier Transform of the product of the displaced objective lens function $f(X'-m/p, Y'-n/q)$ and the collecting lens function $g(X',Y')$ where the integration area is defined by the half diameter $r_{O'}$ of the collecting objective ($r_{O'}$ is commonly normalised to unity). The standard way to calculate the Fourier transform is using a numerical two-dimensional fast Fourier Transform (FFT). In this paragraph we will show that the use of orthogonal functions in the diffracting exit pupil and in the image plane where the detectors are situated can lead to an explicit expression for the detector intensity. This expression is then readily integrated over the detector area defined by D_k.

The function to be Fourier transformed generally is a complex function that is nonzero on a typical domain within the exit pupil of O' like the hatched area in Fig. 4.26. Orthogonal functions with uniform weighting on the circular exit pupil area of O' (pupil radius normalised to unity) are the well-known Zernike polynomials [31]. These polynomials are normally used to represent the argument of the complex pupil function, in particular the aberration or the focus defect of an imaging optical system. Here we propose to extend the use of the Zernike polynomials and to represent the complete complex function $A(\rho,\theta)\exp[\mathrm{i}\Phi(\rho,\theta)]$ in the exit pupil by a set of Zernike polynomials yielding complex expansion coefficients a_{kl} according to

$$A(\rho,\theta)\exp[\mathrm{i}\Phi(\rho,\theta)] = \sum_{k,l} \alpha_{kl}\ R_k^l(\rho)\cos l\theta \quad , \tag{4.26}$$

where (ρ,θ) are the polar coordinates in the exit pupil, $A(\rho,\theta)$ the amplitude and $\Phi(\rho,\theta)$ the phase of the pupil function. $R_k^l(\rho)$ is the radial part of the Zernike polynomial with radial degree k and azimuthal degree l ($k \geq l \geq 0$, $k-l$ even). We have limited ourselves to the cosine-polynomials, thus restricting the functions to those that are symmetric with respect to $\theta = 0$. The expansion of a general function would require an extra set of coefficients b_{kl} associated with the sine-polynomials.

The expansion in (4.26) of $A\exp(\mathrm{i}\Phi)$ can be obtained, for instance, by a least squares fit with a finite series of polynomials, a procedure which is common practice for expanding Φ itself. It can be shown that, with the transition from cartesian coordinates (x, y) to normalised polar coordinates (r, ϕ), the complex amplitude U in the detector plane is given by

$$U(r,\phi) = 2\sum_{k,l} a_{kl}\,\mathrm{i}^l\,V_{kl}\cos l\phi \quad , \tag{4.27}$$

where

$$V_{kl} = \int_0^1 \rho\exp(\mathrm{i}f_z\rho^2)\,R_k^l(\rho)\,J_l(2\pi\rho r)\mathrm{d}\rho \tag{4.28}$$

for integers $k, l \geq 0$ with $k-l \geq 0$ and even (J_l denotes the Bessel function of the first kind of order l). A quadratic phase factor $\exp(\mathrm{i}f_z\rho^2)$ has been added that is needed in the case of a defocusing of the image plane with respect to the optimum (paraxial) focus.

The Bessel series representation of V_{kl} is given in [32,33] and reads

$$V_{kl} = \exp(\mathrm{i}f_z)\sum_{s=1}^{\infty}(-2\mathrm{i}f_z)^{s-1}\sum_{j=0}^{p_z} v_{sj}\frac{J_{l+s+2j}(t)}{st^s} \tag{4.29}$$

with v_{sj} given by

$$v_{sj} = (-1)^{p_z}(l+s+2j)\binom{l+j+s-1}{s-1}\binom{j+s-1}{s-1}\binom{s-1}{p_z-j}\Big/\binom{q_z+s+j}{s} \quad , \tag{4.30}$$

for $s = 1, 2, \ldots;\ j = 0, \ldots, p_z$. In (4.29) we have set

$$t = 2\pi r \quad , \qquad p_z = \frac{k-l}{2} \quad , \qquad q_z = \frac{k+l}{2} \quad . \tag{4.31}$$

For the number S of terms to be included in the infinite series over s we have the following rule. It can be shown that, when S=25, the absolute truncation error is of the order 10^{-6} for all f_z, t, k, l specified by

$$|f_z| \leq 2\pi, \qquad t \leq 20, \qquad 0 \leq p_z \leq q_z \leq 6 \quad . \tag{4.32}$$

In the absence of defocusing ($f_z = 0$), the expansion of the complex pupil function in terms of Zernike polynomials leads to the relationship

$$U(r,\phi) = 2\sum_{k,l} a_{kl}\,\mathrm{i}^l\,(-1)^{\frac{k-l}{2}}\,\frac{J_{k+1}(2\pi r)}{2\pi r}\,\cos l\phi \quad . \tag{4.33}$$

This analytic result for the in-focus amplitude can be obtained using formula (39), p. 772 in [31]

$$\int_0^1 \rho R_k^l(\rho)J_l(2\pi\rho r)\mathrm{d}\rho = (-1)^{\frac{k-l}{2}}\,\frac{J_{k+1}(2\pi r)}{2\pi r} \quad , \tag{4.34}$$

but it also follows as the limiting case for $f_z \to 0$ of (4.28)–(4.29).

4.6.5 Orthogonality in Pupil and Image Plane

Having seen the one-to-one relationship between the orthogonal Zernike polynomials in the pupil and the corresponding Bessel functions of the first kind in the image plane (in-focus situation), we are left with the evaluation of the inner products of the Bessel functions to check their orthogonality and establish a normalisation factor.

We define the inner product of two functions from the series in (4.33) as

$$\begin{aligned} I_{k,l;k',l'} &= C \int_0^{2\pi}\int_0^{\infty} U_{kl}U_{k'l'}\, r\mathrm{d}r\mathrm{d}\phi \\ &= 4C\ \mathrm{i}^{k+k'} \int_0^{2\pi}\int_0^{\infty} \frac{J_{k+1}(2\pi r)J_{k'+1}(2\pi r)}{4\pi^2 r^2} \cos l\phi \cos l'\phi\ r\mathrm{d}r\mathrm{d}\phi \\ &= \frac{C}{\pi}\ \mathrm{i}^{k+k'} \int_0^{\infty} \frac{J_{k+1}(2\pi r)J_{k'+1}(2\pi r)}{r}\ \mathrm{d}r\ \ \delta(l,l')\ , \end{aligned} \tag{4.35}$$

with $\delta(l,l')$ the Kronecker symbol (= 2 if $l = l'$=0, = 1 if $l = l' \neq 0$ and equals 0 if $l \neq l'$). The common conditions for the indices of Zernike polynomials apply too (both $k - l$ and $k' - l'$ are even).

For the calculation of the definite integral over r we use the special result for the integral of the product of two Bessel functions in paragraph 11.4.6 of [34]

$$\begin{aligned} \int_0^{\infty} t^{-1} J_{\nu+2n+1}(t)J_{\nu+2m+1}(t)\mathrm{d}t &= 0 \qquad\qquad (m \neq n) \\ &= \frac{1}{2(2n+\nu+1)} \qquad (m = n), \end{aligned} \tag{4.36}$$

with the restriction $\nu + n + m > -1$.

Applied to (4.35) with $\nu + 2n = k$, $\nu + 2m = k'$ and $t = 2\pi r$, we find the result

$$\begin{aligned} I_{k,l;k',l'} &= \frac{C}{\pi} \qquad\qquad (k = k' = l = l' = 0) \\ &= \frac{(-1)^k}{2(k+1)}\frac{C}{\pi} \qquad (k = k' \neq 0;\ l = l' \neq 0) \end{aligned} \tag{4.37}$$

For normalisation purposes, we multiply the basic functions with $\sqrt{\pi}$ so that the value of $I_{00;00}$ equals unity.

The Evaluation of the Function $Z_k(m, \hat{n}; m', \hat{n}')$ The quantity $Z_k(m, \hat{n}; m', \hat{n}')$ was the key quantity to be calculated for the evaluation of the harmonic signal components (see (4.25)).

Using the orthogonal expansion for the functions $F_{m,n}(x, y)$ in the exit pupil of the objective O' according to (4.33) we obtain

$$\begin{aligned} Z_j(m, \hat{n}; m', \hat{n}') = &\sum_{kl}\sum_{k'l'} a_{kl}a'^{*}_{k'l'}\ i^{l-l'}\ (-1)^{\frac{k+k'-l-l'}{2}} \\ &\times \iint_{\mathrm{D}_j} S_{\mathrm{r}}(r,\phi)\frac{J_{k+1}(2\pi r)J_{k'+1}(2\pi r)}{4\pi^2 r^2} \cos l\phi \cos l'\phi\ r\mathrm{d}r\mathrm{d}\phi\ , \end{aligned} \tag{4.38}$$

where the asterisk denotes the complex conjugate function and the detector sensitivity function S_r has been written in polar coordinates.

In the general situation with an arbitrary detector sensitivity function, the integrals above have to be evaluated numerically. In the whole analytic process described until now, the calculation of the coefficients a_{kl} of the Zernike expansion of the complex function in the exit pupil of O' is the crucial step with respect to speed and convergence and this aspect is actively researched at this moment.

4.7 Conclusion

In this chapter, we have described some important aspects of the optical disc systems that are actually widely used, the CD- and the DVD-system. The subject of optical storage being vast, we have treated some restricted topics that are of interest for the understanding of the CD- and DVD-system themselves and the exchange of media between these two standardized platforms of optical recording. A detailed analysis has been given of the signal detection in an optical disc system. This is of particular interest when simulations are used as a predictive tool, in parallel to experimental research. The efficiency of the calculations is greatly enhanced when the data belonging to the information structure on the disc are separated from the imaging process in the optical recording unit; a procedure to achieve this separation has been described in this chapter and an important reduction in calculation time is obtained. We have also indicated that, by using Zernike polynomials to represent complex amplitude functions, the analytic calculation process can be pursued further and the speed in evaluating the detector signal is again improved.

Because of space limitations, we have not discussed recent developments in optical storage like the Blu-ray Disc system and still other research activities like near-field optical recording that aim at super high-density discs. With respect to these new developments, the analysis given in this chapter has to be adapted to the vectorial diffraction aspects of high-aperture image formation.

References

1. G. Bouwhuis, J. Braat: 'Recording and reading of information on optical disks', In: *Applied Optics and Optical Engineering*, ed. by R.R. Shannon, J.C. Wyant (Academic, New York 1983) pp. 22–45
2. G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, K. Schouhamer Immink: *Principles of optical disc systems* (Adam Hilger, Bristol 1984)
3. J. Isailovic: *Videodisc and optical memory systems* (Prentice Hall, Englewood Cliffs 1985)
4. A. Marchant: *Optical recording; a technical overview* (Addison Wesley, Boston 1990)
5. G. Sincerbox, J. Zavislan: *Selected Papers on Optical Storage* (SPIE Optical Engineering Press, Bellingham 1992)
6. K. Schwartz: *The physics of optical recording* (Springer, Berlin 1993)

7. M. Mansuripur: *The physical principles of magneto optical data recording* (Cambridge University Press, Cambridge 1995)
8. E.W. Williams: *Textbooks in electrical and electronic engineering Vol 2, The CD-ROM and optical recording systems* (Oxford University Press, Oxford 1996)
9. S. Stan: *The CD-ROM drive, a brief system description* (Kluwer Academic Publishers, Dordrecht 1998)
10. M. Minsky: Microscopy Apparatus , U.S. Patent 3,013,467 (1961)
11. Ph. Rice et al.: J. Soc. Motion Pict. T. **79**, 997 (1970)
12. K. Compaan, P. Kramer: Philips Tech. Rev. **33**, 178 (1973)
13. G. Broussaud, E. Spitz, C. Tinet, F. LeCarvenec: IEEE T. Broadc. Telev. **BTR–20**, 332 (1974)
14. G.W. Hrbek: J. Soc. Motion Pict. T. **83**, 580 (1974)
15. J. Braat: Appl. Optics **22**, 2196 (1983)
16. J. Haisma, E. Hugues, C. Babolat: Opt. Lett. **4**, 70 (1979)
17. M. Terao, T. Nishida, Y. Miyauchi, S. Horigome, T. Kaku, N. Ohta: P. Soc. Photo–Opt. Inst. **695**, 105 (1986)
18. T. Ohta, et al.: Jap. J. Appl. Phys. **28**, 123 (1989)
19. W.T. Welford: J. Microsc. – Oxford **96**, 105 (1972)
20. T. Wilson, C.J.R. Sheppard: *Theory and practice of the scanning optical microscope* (Academic Press, London 1984)
21. C. Bricot, J. Lehureau, C. Puech, F. Le Carvennec: IEEE T. Consum. Electr. **CE–22**, 304 (1976)
22. M. Mansuripur, C. Pons: P. Soc. Photo–Opt. Inst. **899**, 56 (1988)
23. G. Bouwhuis, P. Burgstede: Philips Tech. Rev. **33**, 186 (1973)
24. J. Braat, G. Bouwhuis: Appl. Optics **17**, 2013 (1978)
25. J. Heemskerk, K. Schouhamer Immink: Philips Tech. Rev. **40**, 157 (1982)
26. N. Arai, H. Yamazaki, S. Saito: Method for recording/reproducing optical information recording medium, optical pickup apparatus, objective lens and design method of objective lens, U.S. Patent 6,243,349 (2001)
27. Y. Komma, K. Kasazumi, S. Nishino, S. Mizumo: P. Soc. Photo–Opt. Inst. **2338**, 282 (1994)
28. B.H.W. Hendriks, J.E. de Vries, H.P. Urbach: Appl. Optics **40**, 6548 (2001)
29. H.H. Hopkins: J. Opt. Soc. Am. **69**, 4 (1979)
30. W.T. Welford: *The aberrations of optical systems* (Adam Hilger, Bristol 1986)
31. M. Born, E. Wolf: *Principles of Optics* (Pergamon Press, Oxford 1970)
32. A.J.E.M. Janssen: J. Opt. Soc. Am. A **19**, 849 (2002)
33. J. Braat, P. Dirksen, A.J.E.M. Janssen: J. Opt. Soc. Am. A **19**, 858 (2002)
34. M. Abramowitz, I. Stegun: *Handbook of Mathematical Functions* (Dover, New York 1970)

5 Superresolution in Scanning Optical Systems

Roy Pike, Deeph Chana, Pelagia Neocleous, and Shi-hong Jiang

5.1 Introduction

In this contribution we describe some approaches which attempt to increase the resolution of optical imaging systems. In particular, we are interested in scanning systems, such as the confocal microscope, using both coherent and fluorescent light, and the optical disc. The basic optical imaging principles of both compact discs and current DVDs and those of a confocal scanning microscope using coherent illumination are very similar and thus some of our considerations can be applied to both. In so far as one might consider a disc system using fluorescent media, even more of the ground would be common.

We will first briefly consider several "direct" approaches to superresolution, namely, the Pendry "perfect lens" idea, Kino's solid immersion lens and Toraldo di Francia's apodising masks. We will then discuss our own approach by summarising work over recent years with the group of Professor Bertero at the University of Genoa and other colleagues which uses the methodolology of the burgeoning field of "inverse problems"[1]. However, this work comprises some fifty papers, including some very recent papers and reviews, [1–4], and we will not labour through the mathematics in any details here, but confine our remarks to general principles and results. We will, nevertheless, include some introductory and didactic analytical and numerical material and also present some general conclusions regarding potential gains in optical disc systems which we have reached in recent joint studies with the optical storage group at Philips laboratories Eindhoven, [5].

"Inverse problems" is now a wide interdisciplinary field in its own right. It provides a generalised form of Shannon's information theory and attempts to quantify the information content of experimental data. For example, apart from its application to optical imaging, we have used similar mathematical approaches to make progress in other fields of modern theoretical physics such as high-temperature superconductivity, early universe problems (quark-gluon plasma, baryogenisis), the design of radar and sonar antennas and speech recognition.

We will show how the application of inverse-problem theory to optics allows us to obtain improvements beyond the classical Rayleigh resolution limit for so-called type I systems and also beyond the performance of ideal so-called type II or confo-

[1] see, for example, the well-established Journal *Inverse Problems* published by the Institute of Physics in the UK

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

cal systems, both with coherent and incoherent light. We have applied these results to achieve robust increases in resolution in high-aperture confocal fluorescence microscopy, using specially calculated image-plane masks, which can reach nearly twice the conventional Rayleigh limit. By the use of such masks in optical disc systems we predict "point spread functions" to be twice as sharp as in the present CD/DVD systems when compared in the square-law detector electrical outputs. Due to the non-linearity of the detection process the resulting potential storage capacity gains are not clear as yet, as we shall discuss below.

5.2 Direct Methods

In proposing new methods for important practical applications it is clearly prudent to note advantages and disadvantages with other techniques aimed at the same end. Here we consider three such alternative schemes.

5.2.1 Pendry Lens

It has been suggested recently by Pendry, [6], that a thin metal slab can make a perfect lens. The refractive index of a metal can be negative. If it becomes -1 the metal supports surface plasmons. Then, according to Pendry, if the dimensions in the propagation direction of the imaging system are small compared to the wavelength of light we can use Maxwell's equations in the electrostatic limit and all diffraction blurring effects disappear. Indeed the lens can amplify the evanescent components of the electromagneic field back to their boundary values at the object. This is independent of the magnetic permeability, μ. The arrangement and the results for a two-slit object are shown in Fig. 5.1.

We have repeated Pendry's calculations for a silver slab of thickness 40 nm with the geometry of Fig. 5.1, taken from Pendry's paper. The refractive index ϵ is $5.7 - 9.0^2/\omega^2 - 0.4\mathrm{i}$ which gives $\epsilon = -1 - 0.4\mathrm{i}$ at 3.48 eV. The result is shown in Fig. 5.2 and, in fact, does not depend on where the slab is placed betweeen the object and image.

Although difficult to achieve in practice, this might look interesting and the use of an optical system with these dimensions would not be impossible to contemplate, for example, for an optical disc system. The question is, therefore, whether 80 nm is sufficiently small compared with an optical wavelength to approximate the electrostatic limit. Fortunately, the geometry of this system is sufficiently simple for an exact calculation to be carried out. We used the longer wavelength of 520 nm than that for silver to help the situation, with the same imaginary part of the refractive index. The results for the TE wave are

$$E_0(x,z) = \sum_{\sin\theta_0} u(\sin\theta_0)\, a_0\, e^{-k_0 z\sqrt{\sin^2\theta_0 - 1}}\; e^{-\mathrm{i}k_0 x\sin\theta_0} \tag{5.1}$$

$$E_2(x,z) = \sum_{\sin\theta_0} u(\sin\theta_0)\, T_\perp\, e^{k_0(h-z)\sqrt{\sin^2\theta_0 - 1}}\; e^{-\mathrm{i}k_0 x\sin\theta_0}\ , \tag{5.2}$$

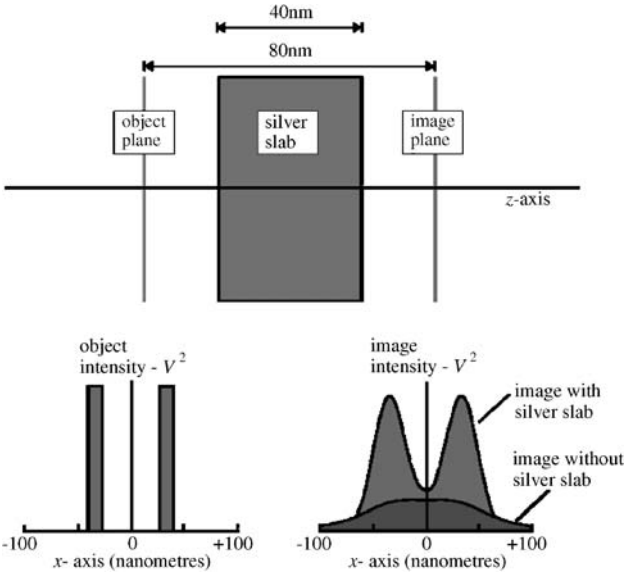


Fig. 5.1. Pendry perfect lens, silver slab, electrostatic limit

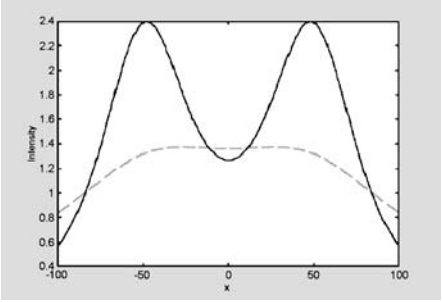


Fig. 5.2. Analytic calculations: silver slab, electrostatic limit. Diffracted image at 80 nm (*dashed line*), image with 40 nm silver-slab lens (*solid line*)

where $E_0(x, z)$ and $E_2(x, z)$ are the electric fields without and with the slab, respectively, x is the transverse coordinate, z is the propagation direction, θ_0 is the incident angle, k_0 is the incident wavevector, $u(\sin\theta_0)$ is the spatial-frequency amplitude of the incident field at the angle θ_0, $T_\perp$ is the transmission coefficient of the slab, a_0 is the amplitude of the incident wave and h is the slab thickness. The transmission coefficient of the slab is given by

$$T_\perp(\theta_0) = \frac{4n_0n_1}{(n_1+n_0)(n_2+n_1)} \frac{e^{-u_1h}}{1-\frac{(n_0-n_1)(n_2-n_1)}{(n_0+n_1)(n_2+n_1)}e^{-2u_1h}} , \tag{5.3}$$

where $n_0 = n_2 = \cos(\theta)/\eta_0, n_1 = \sqrt{\epsilon_r/\mu_r}\cos(\theta_1)/\eta_0, \eta = \sqrt{i\mu\omega/(\sigma+i\epsilon\omega)}$, the characteristic frequency of the medium for plane-wave propagation, $u_1 = ik_0\sqrt{\mu_r\epsilon_r}\cos(\theta_1)$, $\sin(\theta_1) = \sin(\theta)/\sqrt{\mu_r\epsilon_r}$ and μ_r, ϵ_r and σ are the complex relative permeability, the complex relative susceptibility and the conductivity of the medium, respectively.

For a double slit with slit-width a and separation d

$$u(\sin(\theta_0)) = \frac{\sin(ak_0 \sin(\theta_0)/2)}{ak_0 \sin(\theta_0)/2} \cos(dk_0 \sin(\theta_0)/2) \; . \tag{5.4}$$

Unfortunately for the perfect-lens idea, our results show that the answer to the question above is negative and, given that the magnetic permeability of such a lens would be 1 or greater it is not until we get down to a lens thickness of 20 nm that the image is satisfactory, by which time the diffraction losses themselves are not large. This is shown in Figs. 5.3, 5.4 and 5.5.

We have also used a simple finite-element numerical simulation to confirm these analytical results. We conclude that the use of a Pendry lens in optical-disc storage

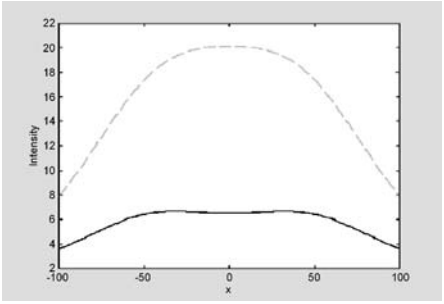


Fig. 5.3. Exact analytic calculations at 520 nm. Diffracted image at 80 nm (*dashed line*), image with 40 nm Pendry lens (*solid line*)

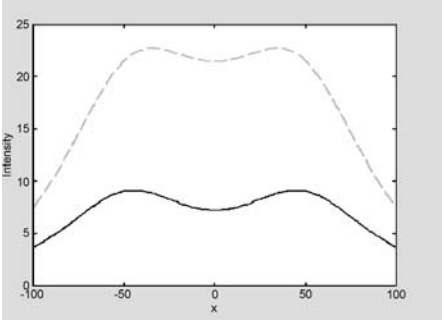


Fig. 5.4. Exact analytic calculations at 520 nm. Diffracted image at 60 nm (*dashed line*), image with 30 nm Pendry lens (*solid line*)

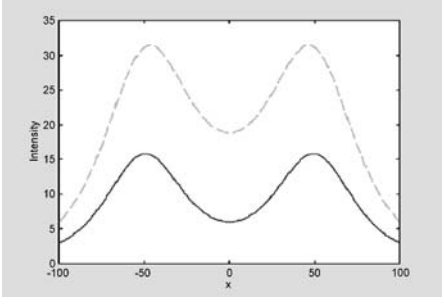


Fig. 5.5. Exact analytic calculations at 520 nm. Diffracted image at 40 nm (*dashed line*), image with 20 nm Pendry lens (*solid line*)

systems is not a promising way forward, unless materials with both surface plasmons and negative magnetic permeability in the visible region could be developed; this seems unlikely in the near future although it is not impossible in principle. See also the work by Shamonina et al. [7], in which Pendry's electrostatic approximation is not made, with similar results.

5.2.2 Kino's Solid Immersion Lens

The so-called solid immersion lens achieves effective apertures greater than are possible with conventional oil-immersion lenses by being operated in the near field. A schematic is shown in Fig. 5.6, [8]. The point spread function depends on the air gap as shown in Fig. 5.7

We illustrate the performance of such a lens for completeness in this chapter since it is a possible way forward for optical storage systems but since the air gaps required are so small, a sealed system would almost certainly be necessary to exclude dust and one then has the question of why use an optical disc rather than a hard-disc magnetic system which would have a much higher speed and capacity?

5.2.3 Toraldo di Francia's Apodising Masks

Making an analogy with superdirective antennas, it was suggested first by Toraldo di Francia, [9], that superresolution could be achieved by use of a pupil-plane mask. An annular ring system of amplitude- and/or phase-modulated pupil functions can

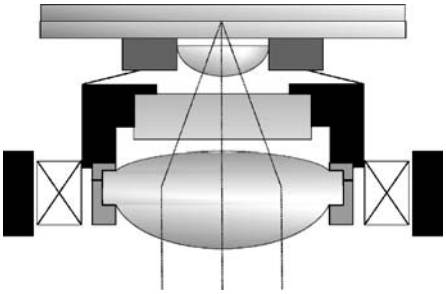


Fig. 5.6. Schematic of solid immersion lens

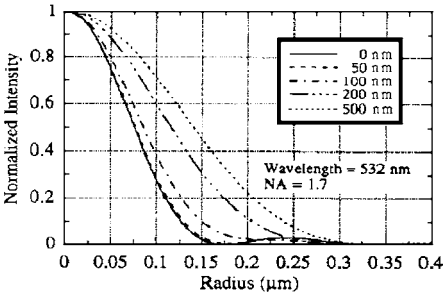


Fig. 5.7. Point spread functions with air gaps, stigmatic SIL, $N.A._{\text{eff}} = 1.7$

be used to modify the point spread function. With such a mask one could trade a resolution gain against a lower Strehl ratio and higher sidelobes.

Di Francia pointed out that there was no theoretical limit to the narrowness of the central peak and no limit either to the diameter of the surrounding dark region but, in practice, the size of the sidelobes increases prohibitively quickly and the Strehl ratio goes down to extremely small values as soon as one departs significantly from the uniform pupil and the normal Airy pattern.

A number of workers have investigated the performance of Toraldo masks using scalar theory (low numerical aperture) and numerical non-linear optimisation schemes. We show a recent example of a pair of conjugate phase-only pupil-plane filters, [10], in Fig. 5.8 which are designed to be used in the confocal transmission system of Fig. 5.9 to achieve axial superresolution.

The point spread functions are given in Fig. 5.10 which shows a 30% gain in axial resolution but with an increased sidelobe structure and an effective Strehl ratio for the combined illumination and collection optics of 0.0625 [11].

A similar pair of three-level, phase-only filters showed a small transverse resolution gain ($\approx 12\%$) with a Strehl ratio of 0.31 in one dimension, x, of the image plane at the expense of a similar loss of resolution in the orthogonal, y, direction.

A further attempt at tailoring axial resolution, [12], in the non-paraxial but still scalar Debye approximation shows similar results but the mask must now be used at a fixed aperture. Inversion of the imaging integral equation, which is discussed in

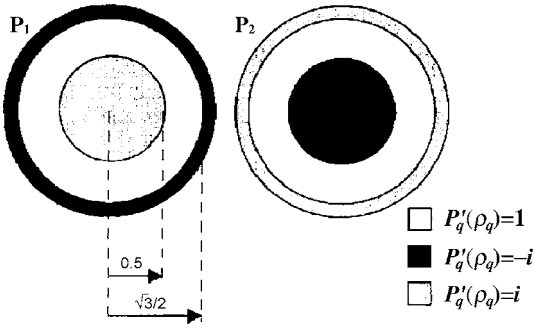


Fig. 5.8. Pair of conjugate phase-only pupil-plane filters

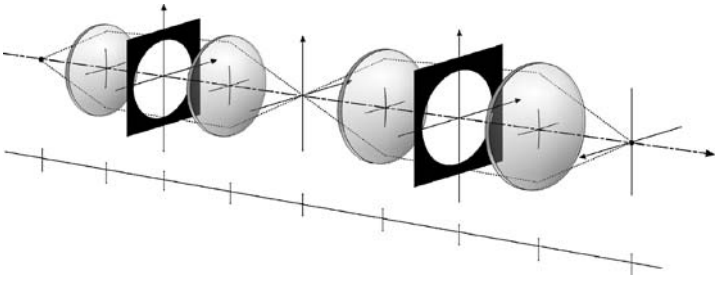


Fig. 5.9. Confocal transmission system with conjugate pupil-plane filters

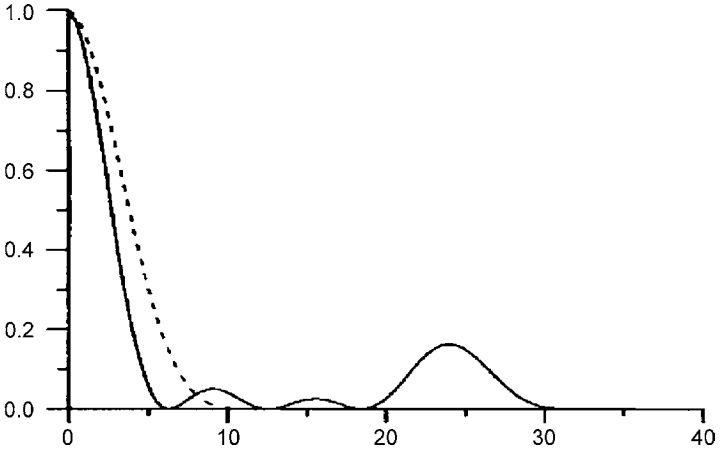


Fig. 5.10. Axial response funtion (normalised) of the conjugate filters of Kowalczyk et al.

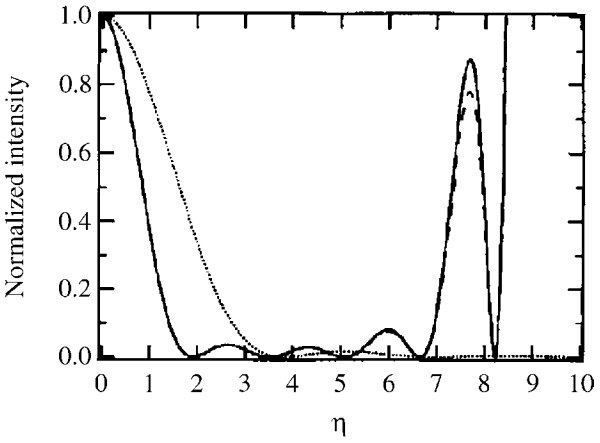


Fig. 5.11. Transverse response funtion (normalised) of the five-zone, pupil-plane, phase mask of Sales and Morris. The Airy pattern is also shown (*dotted line*)

detail below, is actually mentioned as a possibility in this paper but not considered further.

As another example we show in Fig. 5.11 the point spread function for the transverse response of a five-zone, pupil-plane, phase mask, [13]. One may see that the Toraldo principle is demonstrated with a 50% gain in transverse resolution but at the cost of enormous sidelobes (off the top of the graph) and a Strehl ratio of 2.10^{-7}.

Again in this subsection we have considered the Toraldo mask as a possible component for high-density storage systems but the rapid loss of energy in the central peak seems to preclude any realistic application. We should also say that the trend towards the use of higher numerical apertures for optical storage systems is not consistent with the limitations to the scalar case and circular symmetry of ex-

isting calculations of the Toraldo mask. We have also considered souce arrays but similar conclusions seem to apply.

5.3 Inverse Methods and Image-Plane Masks

In the application of the methods of the field of inverse problems to confocal scanning imaging systems one considers the image plane to contain "data" which must be used to determine as much information as possible about the object which generated it. It is also made clear that this information, in general, cannot be used to reconstruct the input completely since, as we shall see in a moment, the mathematical operator which transforms the input into the data is linear and compact and is thus a "smoothing" operator. Such an operator, in the presence of noise, irretrievably loses any components of the object which vary too rapidly. This gives rise to a defined "resolution" limit.

The theory shows that there is a "natural" basis for each such operator which provides a complete set of orthonormal components in which to expand the input and output (not necessarily the same for both) which are ordered in such a way that the components which are most robust to noise interference are those with the lower indices. A generalised, noise-dependent "Shannon number" can be defined, which is the index which limits the so-called "signal subspace". The noise prevents reliable recovery of the components with indices higher than this value, unless present in inordinate strength, which will live in the "noise subspace".

The process of reconstruction of the object in a scanning optical system using these concepts can be effected by placing a specially calculated, generally complex, optical mask in the image plane and integrating the transmitted complex amplitudes using suitable collection optics [14]. Alternatively, in the coherent case this can be calculated as a pupil-plane mask where it would act as a sophisticated type of Toraldo mask with uniaxial symmetry at high *NA* due to the effects of polarisation. In our experience the image-plane versions are simpler to fabricate and we do not discuss the pupil-plane versions any further.

We will give a brief outline of the definition and calculation of an image-plane mask for a generic imaging operator, A, defined by

$$(Af)(y) = \int A(x, y) f(x) \mathrm{d}x \,, \tag{5.5}$$

where x and y can be multidimensional. We call the object $f(x)$ and the image $g(y)$ and the imaging is described by the operator equation

$$g = Af. \tag{5.6}$$

We shall see a specific example of such an operator shortly in (5.16). (Of course you may peep now!)

The image-plane mask is calculated using the singular value decomposition (SVD) of the operator A, which we consider in the first place to act upon square

integrable functions, i.e. technically it maps functions from one L^2 space into another. The SVD is given by the set of solutions $\{\alpha_k; u_k, v_k\}_{k=0}^{\infty}$ of the coupled integral equations

$$Au_k = \alpha_k v_k \tag{5.7}$$

$$A^* v_k = \alpha_k u_k \,, \tag{5.8}$$

where A^* denotes the adjoint of A. The u_k and v_k are called singular functions and provide orthonormal bases for f and g, respectively, (note that this notation is sometimes reversed). The α_k are called singular values and for a compact operator A they accumulate to zero with increasing k; they represent the strength with which each component is transferred into the image and hence how well it can withstand the addition of noise. Singular value decomposition can be accomplished using standard numerical packages found in many software libraries. Given this decomposition of the imaging operator, by expanding object and image in their respective bases and using their orthonormality it can be seen after a little algebra that the solution of (5.6) is

$$f(x) = \sum_{k=0}^{\infty} \frac{1}{\alpha_k} u_k(x) \int g(y) v_k(y) \mathrm{d}y \,, \tag{5.9}$$

and specifically, on the optical axis

$$f(0) = \sum_{k=0}^{\infty} \frac{1}{\alpha_k} u_k(0) \int g(y) v_k(y) \mathrm{d}y \,. \tag{5.10}$$

For a scanning system the determination of $f(0)$ at each point of the scan is sufficient to reconstruct the image and we utilise this fact for practical reasons, although this does entail some loss of information. Exchanging the sum in (5.10) with the integral we obtain

$$f(0) = \int g(y) M(y) \mathrm{d}y \,, \tag{5.11}$$

where

$$M(y) = \sum_{k=0}^{\infty} \frac{1}{\alpha_k} u_k(0) v_k(y) \,. \tag{5.12}$$

The function $M(y)$ is called an image-plane optical mask and it can be seen that (5.11) explains the object reconstruction process described earlier. In practice the summation is truncated at a value of k which depends on the signal to noise ratio of the detected signal. Using the fact that (5.11) is the scalar product (M, g) in L^2 of M and g and that $g = Af$, we can see that

$$f(0) = (M, g) = (M, Af) = (A^* M, f) \,. \tag{5.13}$$

Thus we can write

$$f(0) = \int T(x) f(x) \mathrm{d}x \,, \tag{5.14}$$

where T is given by

$$T = A^* M \,. \tag{5.15}$$

Thus T is a smoothing operator or "mollifier".

5.4 Optical Systems for Scanning Imaging

We will now try to make some comparisons of the performance of three imaging systems used for scanning, namely, a conventional type I system, a confocal type II system and a system using an image-plane optical mask.

We consider an optical system with two identical aberration-free lenses each with numerical aperture (*N.A.*) for illumination and imaging, respectively, and uniform pupil function. The same lens serves for both in a reflective system. This is a special case of the general partially coherent optics described, for example, in early work of Hopkins [15]. In one dimension the action of each lens is described by the linear integral operator A, relating object $f(x)$ to image $g(y)$:

$$g(y) = (Af)(y) = \int_{-X/2}^{X/2} \frac{\sin \Omega(y-x)}{\pi(y-x)} f(x) \mathrm{d}x \,, \tag{5.16}$$

where the lens accepts light between the transverse wavenumbers $2\pi(N.A.)/\lambda$ of $-\Omega$ and Ω radians/metre (i.e. over a Fourier bandwidth of 2Ω) and the object lies between $-X/2$ and $X/2$. The support, y, of the image is, in theory, infinite but in practice will be defined by the rapid fall-off of the illumination or by a finite detector aperture. The optical system is shown in Fig. 5.12.

We will use optical units for x and y by multiplying actual units by Ω/π and the above equation then has the normalised form:

$$g(y) = \int_{-S/2}^{S/2} \frac{\sin \pi(y-x)}{\pi(y-x)} f(x) \mathrm{d}x \tag{5.17}$$

$$= \int_{-S/2}^{S/2} \mathrm{sinc}\{\pi(y-x)\} f(x) \mathrm{d}x \tag{5.18}$$

$$= \int_{-S/2}^{S/2} S(y-x) f(x) \mathrm{d}x \,, \tag{5.19}$$

where the script $\mathcal{S} = X\Omega/\pi$ is the Shannon number and $S(\cdot)$ is the point spread function. In this form the numerical aperture is taken out of the further calculations and the band of the lens always lies between $-\pi$ and π. We use low-aperture scalar theory to make contact with the work of Braat [16] and Stallinga [5] and one-dimensional

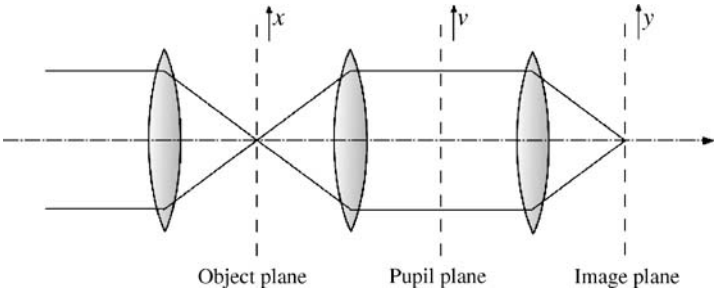


Fig. 5.12. Object, pupil and image planes

integrals for illustration. The one-dimensional results are didactic and more than academic in that they would apply (as a tensor product) to square pupils in two dimensions and they also provide some useful exact analytic formulae. The generalisation to circular pupils in two dimensions requires more lengthy computations but is straightforward (if not the generalisations for high-aperture vector theory, which are dealt with in [2]).

We fill the illumination lens with a plane wave, i.e. $f(x)$ of (5.17) is a point on axis at infinity; the image $g(y)$ (= sinc(πy) in one dimension) then lies in its focal plane, the object plane, and provides the illumination profile of the scanning target (a gaussian beam can be dealt with by modifying the form of the source $f(x)$). We now redefine object and image for the action of the second lens, which again follows (5.17). Using s for the scanning variable, the objective lens sees as its object, $f(x)$, a (complex) field $S(x)R(x-s)$, where $R(x-s)$ is the scanned object transmittance (or reflectance). The field in the image plane of the objective (ignoring magnification) is thus defined by the linear integral operator, K, where

$$g(y, s) = (KR)(y, s) = \int_{-\infty}^{\infty} S(y - x)S(x)R(x - s)\mathrm{d}x \,. \tag{5.20}$$

Here we have taken the limits of the integral to infinity since the support of the object will be defined by the illumination function, which will be effectively zero after a few diffraction rings. The three cases of interest to consider are:

(i) An extended detector in the image plane
(ii) A point detector in the image plane at $y = 0$
(iii) A mask in the image-plane followed by a detection system

A first suggestion as a way to compare these systems would be to use the concepts of Fourier spatial-frequency bandwidth and modulation transfer function and compare these between the systems. A second suggestion would be to calculate the response of each of the systems to an object consisting of a point "impulse" and compare the different object reconstructions. The first method should give just the Fourier transforms of the second. However, for our problem this will not be the case. This is because the above equivalence holds only for linear systems but the detectors respond to the modulus-squared of the field amplitudes g, in different ways and

thus make the systems non-linear in a non-trivial fashion. We will discuss this point, which has led to some confusion in the field, further below but to see what causes the confusion we first go through the formal exercise from both frequency- and impulse-response points of view and then show how they differ and what practical relevence this has.

We shall calculate the spatial-frequency responses in these three cases by defining $R(x)$ as $\cos(lx)$ and the intensity impulse responses by defining $R(x)$ as $\delta(x)$. We first present the analysis for each case and then the results of numerical calculations. For reference it is well known (see, for example, [17]) that a one-dimensional (coherent) system in which the second lens is replaced by a small detector on the optical axis has a uniform frequency response up to a bandwidth l of π and is zero thereafter.

5.4.1 Analytical Results

(i) Extended Detector in the Image Plane This is known as a type I scanning system and its resolution is the same as that of a conventional type I microscope, see for example, Wilson and Sheppard's book [17] pp. 3 and 42. It is the system used in conventional optical disc readout [16,5].

Spatial-frequency response
Using (5.20), substituting $\cos\{l(x-s)\}$ for $R(x-s)$ and following the notation for detected intensity of Stallinga [5], in which the first subscript denotes [i]mage or [p]upil plane and the second a [p]oint or [e]xtended detector, we have

$$I_{i,e}(l,s) = \int_{\text{detector}} \left| \int_{-\infty}^{\infty} S(y-x)S(x)\cos\{l(x-s)\}\mathrm{d}x \right|^2 \mathrm{d}y \,. \tag{5.21}$$

Using our one-dimensional test-bed this becomes

$$I_{i,e}(s) = \int_{\text{detector}} \left| \int_{-\infty}^{\infty} \operatorname{sinc}\{\pi(y-x)\}\operatorname{sinc}(\pi x)\cos\{l(x-s)\}\mathrm{d}x \right|^2 \mathrm{d}y \,. \tag{5.22}$$

This can be written as

$$I_{i,e}(l,s) = \int_{\text{detector}} |\operatorname{sinc}(\pi y) \otimes \{\operatorname{sinc}(\pi y)\cos[l(y-s)]\}|^2 \,\mathrm{d}y \,, \tag{5.23}$$

where $\otimes$ denotes convolution. By Parseval's theorem this is equal to

$$\begin{aligned} I_{p,e}(l,s) &= \int_{\text{pupil}} |\{C(v)\} \\ &\quad \{C(v) \otimes [\exp(\mathrm{i}ls)\delta(v-l) + \exp(-\mathrm{i}ls)\delta(v+l)]\}|^2 \mathrm{d}v \\ &= \int_{-\pi}^{\pi} |\{\exp(\mathrm{i}ls)C(v-l) + \exp(-\mathrm{i}ls)C(v+l)\}|^2 \mathrm{d}v \,, \end{aligned} \tag{5.24}$$

where $C(v) = \operatorname{rect}(\pi v)$ is the characteristic function of the pupil, namely, unity when v is inside it and zero otherwise. This result shows that in this case the extended detector integrates the intensity of a double-slit diffraction pattern. The centres, widths

and phases of the slits are piecewise functions of l depending on the overlap of the rect functions with the limits of the integral. We need to consider separately the regions $0 \le l < \pi$ and $\pi \le l \le 2\pi$.
For $0 \le l < \pi$

$$I_{p,e}(l,s) = \int_{-\pi}^{l-\pi} \mathrm{d}v + 2\int_{-\pi+l}^{\pi-l} |\{\exp(\mathrm{i}ls) + \exp(-\mathrm{i}ls)\}|^2 \mathrm{d}v + \int_{\pi-l}^{\pi} \mathrm{d}v$$
$$= 16(\pi - l)\cos^2(ls) + 2l\,, \tag{5.25}$$

and for $\pi \le l \le 2\pi$

$$I_{p,e}(l,s) = \int_{-\pi}^{\pi-l} \mathrm{d}v + \int_{l-\pi}^{\pi} \mathrm{d}v = 2(2\pi - l)\,. \tag{5.26}$$

This behaviour is shown in Fig. 5.13 which, in fact, is the result of a numerical simulation to be described below. The signal persists at frequencies l beyond π, decreasing to zero at $l = 2\pi$. However, for a pure sinusoidal input probe it carries no modulation beyond $l = \pi$.

Intensity impulse response
Using (5.20) the intensity impulse response is, as a function of scan position,

$$I_{i,e}(s) = \int_{\text{detector}} \left| \int_{-\infty}^{\infty} S(y-x)S(x)\delta(x-s)\mathrm{d}x \right|^2 \mathrm{d}y \tag{5.27}$$
$$= \int_{\text{detector}} |S(y-s)S(s)|^2\,\mathrm{d}y \tag{5.28}$$
$$= |S(s)|^2 \int_{\text{detector}} |S(y-s)|^2\,\mathrm{d}y\,. \tag{5.29}$$

For a detector which is larger than the effective size of $S(y)$ the integral will be a constant, independent of s, and thus

$$I_{i,e}(s) = |S(s)|^2\,. \tag{5.30}$$

As pointed out by Stallinga, this response is the same as that given by integrating *intensity* over the pupil plane and it agrees with (3.7) of Ref. [17]. Note that for a point object the response is the same for both coherent and incoherent illumination.

(ii) Point Detector in the Image Plane This is known as a type II or "confocal" system. We put $y = 0$ in (5.20) and again follow the notation of Stallinga.

Spatial-frequency response
This is given by

$$I_{i,p}(l,s) = \left| \int \mathrm{sinc}^2(\pi x)\cos\{l(x-s)\}\mathrm{d}x \right|^2\,. \tag{5.31}$$

This function has the form shown in Fig. 5.14, obtained by numerical simulation below.

Intensity impulse response
The detected intensity as a function of scan position is

$$I_{i,p}(l,s) = \left| \int_{-\infty}^{\infty} S(x)S(x)\delta(x-s)\mathrm{d}x\,\mathrm{d}y \right|^2 \tag{5.32}$$

$$= |S(s)|^4 . \tag{5.33}$$

This may be shown to be equivalent to integrating the *amplitude* over the pupil plane [17] p. 48 and in the one-dimensional case is 20% narrower at half height than $I_{i,e}$.

(iii) Mask in the Image Plane This is the new detection system developed at King's College, London, using inverse-problem techniques.

Spatial-frequency response
The effect of a mask in the image plane is shown in [14] to have the same bandwidth *for the amplitude image* as the confocal microscope (case (ii) above) but to fill it more effectively at higher frequencies. Explicitly, using (5.15) we find

$$T(x) = S(x) \int S(y-x)M(y)\mathrm{d}y , \tag{5.34}$$

and by taking the Fourier transform of both sides of this equation we find that the transfer function $\hat{T}(\omega)$ is given by

$$\hat{T}(\omega) = \frac{1}{(2\pi)^n} \int \hat{S}(\omega-\omega')S(\omega')\hat{M}(\omega')\mathrm{d}\omega' , \tag{5.35}$$

where $n = 1$ in one dimension and $n = 2$ in two dimensions.

In one dimension the infinite sum for the mask gives [14]

$$\hat{M}(\omega') = 4\pi\{\delta(\omega-\pi) + \delta(\omega+\pi)\} , \tag{5.36}$$

and

$$\hat{T}(\omega) = \frac{1}{(2\pi)} \int_{\omega-\pi}^{\pi} \hat{M}(\omega')\mathrm{d}\omega' \qquad 0 < \omega < 2\pi , \tag{5.37}$$

so that

$$\hat{T}(\omega) = 1, \quad 0 < \omega < 2\pi . \tag{5.38}$$

Intensity impulse response
Application of (5.14) to a δ function gives an impulse response of $\frac{1}{2}\mathrm{sinc}(2\pi s)$ and an intensity impulse response is found of

$$I_{i,m}(s) = \frac{1}{4}|\mathrm{sinc}(2\pi s)|^2 , \tag{5.39}$$

where the extension of Stallinga's notation denotes a mask in the image plane. This result may also be calculated directly using (5.20).

5.4.2 Numerical Results

We have written a software package which calculates the images given by (5.20) and then implements the three detection schemes by numerical integration. The frequency-response functions are found by using a cosinusoidal grating of varying periods up to a bandwidth of 2π and following the modulations of the detector outputs as the grating is translated laterally in the object plane. Figs. 5.13, 5.14 and 5.15 show the output signals at a series of grating translations against the frequency of the grating for each of the three systems and Fig. 5.16 shows the normalised modulation depths at each frequency.

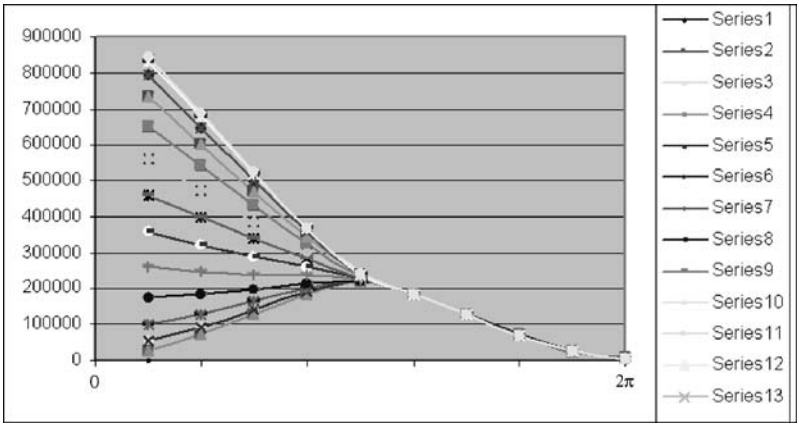


Fig. 5.13. Spatial frequency response of type I system. The plots are for a series of grating translations

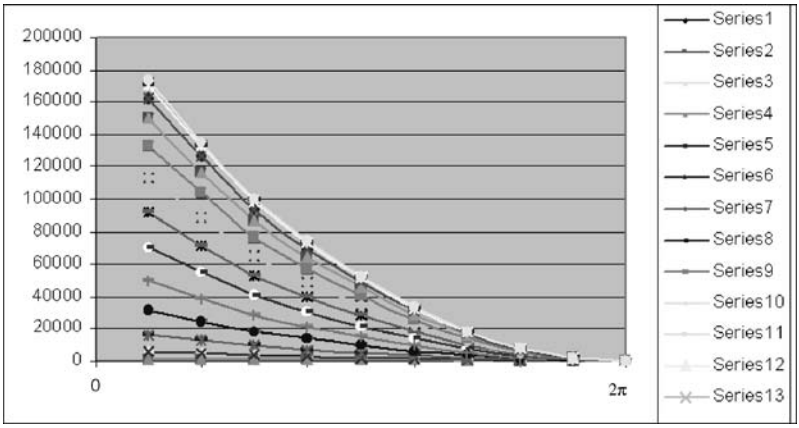


Fig. 5.14. Spatial frequency response of confocal system. The plots are for a series of grating translations

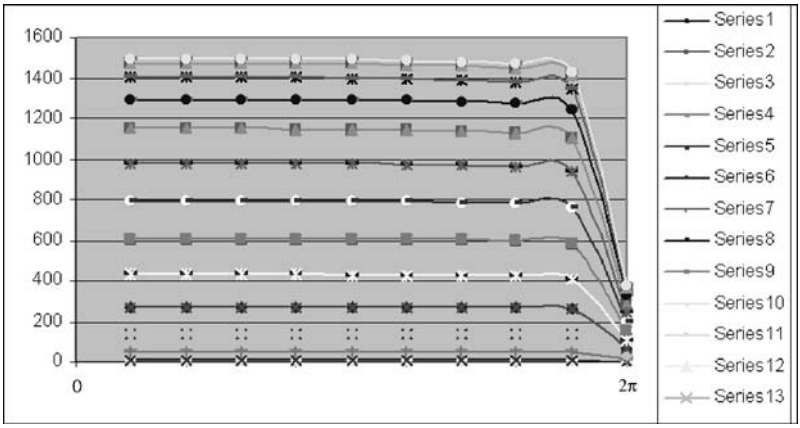


Fig. 5.15. Spatial frequency response of mask system. The plots are for a series of grating translations

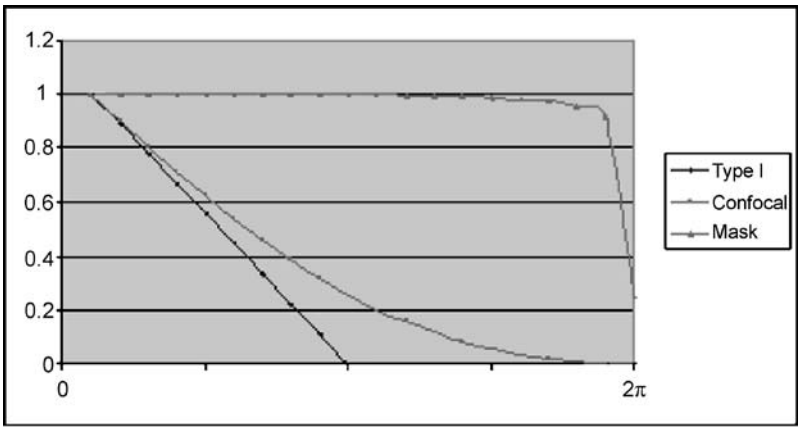


Fig. 5.16. Modulation depth versus spatial frequency (normalised)

It can be seen that the type I system loses the modulation of signal at a bandwidth of π although it has a DC response up to 2π. This is in complete agreement with (5.24), (5.25) and (5.26). There is a full filling of the 2π band in the mask case in agreement with (5.38) save for some numerical effect near the edge of the band. The response of the coherent detector is in line with the literature.

The impulse response functions were calculated by creating one dimensional "pits" of varying widths and scanning them in the object plane. We have also looked at the response to a pair of pits at different spacings with a mark-space ratio of unity. Fig. 5.17 shows the responses of the different detection schemes to a moving narrow pit, normalised to unity at the origin, together with the original sinc function which illuminates the pits, and Fig. 5.18 shows the pair responses. The two-point response function has also been calculated analytically and is shown in Fig. 5.19, it is seen

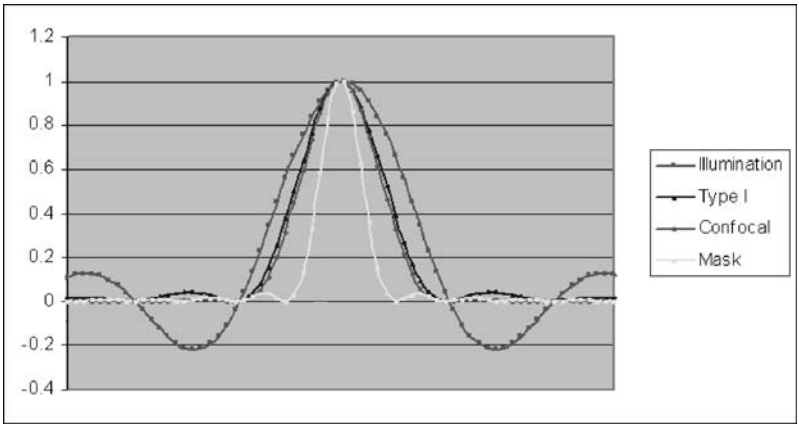


Fig. 5.17. Responses of the three detection systems to a moving narrow pit (normalised). The illumination profile is also shown

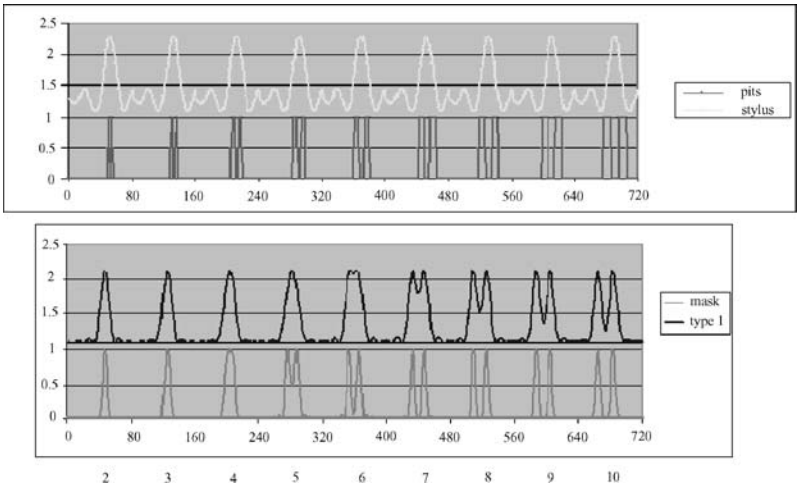


Fig. 5.18. Responses of the type I and mask detection systems to a pair of pits with mark-space ratio of unity at increasing spacings. The illumination profile is also shown above each spacing

that the results support the above numerical work. The absolute values of the response of the mask scheme in both frequency and point response is dependent upon the fact that it squares the integrated field rather than integrating the squares as in the other two cases. We have not had time yet to evaluate the exact relative efficiencies of these coherent systems for practical application but our recent experiments using image-plane masks in confocal scanning fluorescence microscopy, to be discussed below, give encouragingly good images at the same incident laser power.

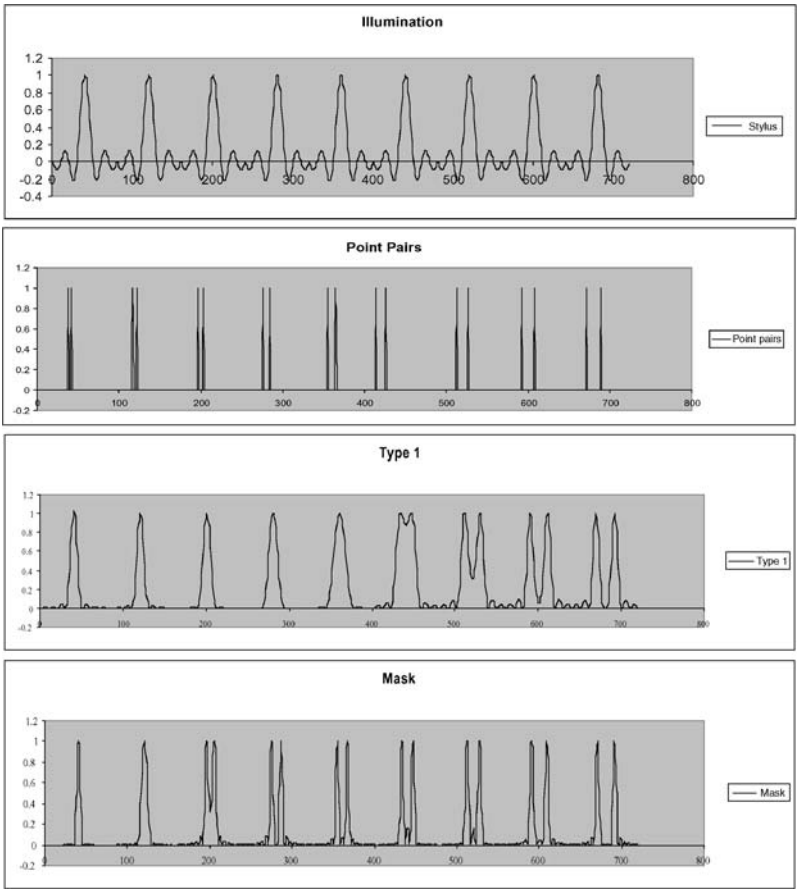


Fig. 5.19. Two-point response of the mask detection scheme. The illumination profile is also shown above each spacing. The spacings are as in Fig. 5.18

5.5 The Comparison of Non-linear Optical Scanning Systems

We seem to have agreement both analytically and numerically on the salient facts of scanning systems. We need to remember that this work only applies to square pupils; the case of circular pupils has yet to be treated in the same way (although the basic physics is unaltered and the results will be qualitatively similar). Our first aim was to resolve uncertainties about the respective (Fourier) band limits of the existing CD/DVD (type I) system and our proposed image-plane mask system but it now seems that this task was misconceived, in the sense that the use of a Fourier decomposition to discuss the performance of an optical (or any other) system is based implicitly upon the assumption that the system is linear. In fact, it is only as

far as the optical fields are concerned that the scanning systems mentioned above are all linear.[2]

The images, $g(y, s)$, in all three cases, are given by (5.20); this is a Fredholm equation of the first kind and is manifestly linear. For a linear system the transfer function in Fourier space can be computed as the amplitude response of the system to sinusoidal input waves of increasing frequency, which will drop to zero at the edge of the "transmission band" or, alternatively, as the Fourier transform of the impulse-response function, which will give the same result. For the second method the input impulse can be decomposed into a uniformly filled band of frequencies (infinite in the theoretical limit but finite in practice) which will appear independently in the output with decreasing amplitudes, exactly the same as if they are presented one by one as in the first method.

However, the electrical signal which is used as output in a CD/DVD system is generated by a photon detector whose output is proportional to the modulus squared of the field amplitude and thus is no longer linearly dependent upon the input. For such a non-linear system these two measures of the "frequency response" are quite different, depending upon the exact nature of the non-linearities. Neither can be said to define an unambiguous "bandwidth" of the system and one sees that the concept of bandwidth loses its fundamental significance for the analysis of the coupled optical-electrical system.[3]

In our case we have seen that the optical-to-electrical impulse response of the type I system to a small single pit is

$$I_{i,e}(s) = |S(s)|^2 \,, \tag{5.40}$$

with a triangular filling of the band to 2π, while the optical-to-electrical impulse response of the mask system is

$$I_{i,m}(s) = |S(2s)|^2 \,, \tag{5.41}$$

with a triangular filling of the band to 4π which thus has double the bandwidth.

On the other hand if the calculation is done by injecting pure sinusoidal waves, the result is a triangular filling of the band to π for type I and a unit filling of the band to 2π for the mask system, which again has double the bandwidth but with different filling factors.

These results are shown in Figs. 5.20 and 5.21.

A third result is also possible for this non-linear system, *viz.*, by adding a DC offset to the sinusoidal input, i.e. imposing a weak signal modulation; we find that the frequency response using sinusoids is a triangle from (0,1) to (2π,0). This is the result found by Braat in "Principles of Optical Disc Systems" in 1985 [16] and according to Stallinga is used in the CD/DVD community.

---

[2] Note that in the case of a fluorescent (or incoherent) object we have the same linear equation (5.20) but with $S(*)$ denoting $\mathrm{sinc}^2(*)$ and g and R this time denoting the *intensities* in the object and image planes.

[3] Of course, the detector output, once generated, can be analysed into Fourier frequencies for discussion of onward linear electronic processing.

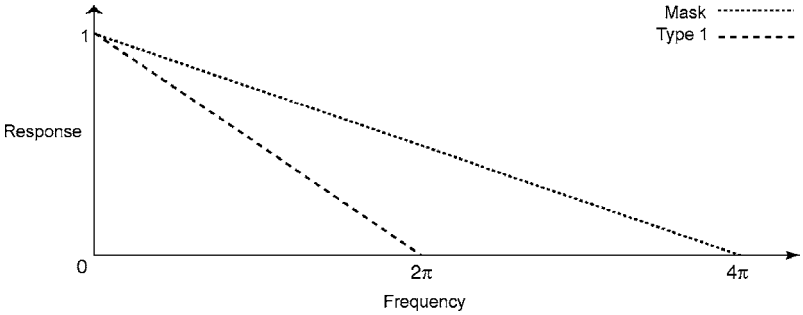


Fig. 5.20. Fourier spectrum of the point response function for type I and mask detection systems

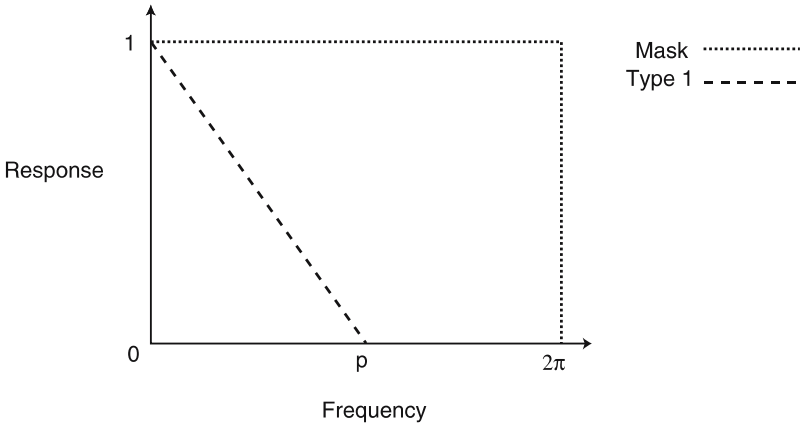


Fig. 5.21. Fourier spectrum of the sinusoidal grating response function for type I and mask detection systems

We should note that in current DVD systems the optical-to-electrical response given by (5.40) is not sufficient to overcome jitter due to adjacent-track interference and an additional "electronic equalisation" must be applied to enhance the tail of the detector response. This would not be necessary with the mask system but could also, in fact, be applied to the mask-detector response given by (5.41) to give a full 4π band in this case.

We have seen, therefore, that it is not a good idea to talk in terms of an optical-to-electrical "frequency response". Fourier analysis is not useful for non-linear systems since the components of the input are not additive in the output and any calculated output "frequency spectra" will be dependent on the particular non-linearities. This applies also to its use for the analysis of different impressed code patterns on optical storage discs since the results would depend in a complicated way on the actual signals. As we have seen above, three different "bandwidths" have already been obtained for three different forms of the input signal. However, whatever the input signal, an image-plane mask system uses all the information available above the

noise level and is fully controlled by the number of terms which are used in the summation of (5.12).

The bottom line for the use of an image-plane mask is that it would seem to promise a near fourfold capacity increase over the present systems in the case where pits are resolved. The reduction in this factor will depend on the actual pattern being scanned and further calculations or simulations and/or experiments are needed to quantify this. For robust codes we would not expect much reduction but our software package is written to handle this problem and we hope to continue our calculations along these lines in the near future by simulating run-length-limited codes. The effects of various levels of noise can also be simulated. The whole exercise will then need to be repeated for circular pupils in two dimensions and we plan to move on to this as soon as possible in parallel with experimental work.

5.6 High-Aperture Image-Plane Masks

In the above discussions we have used simple low-aperture, one-dimensional calculations to illustrate the basic concepts. In reality, of course, this needs to be put into practice in real microscopes and optical storage systems where full vector analysis of high-aperture systems is required. As mentioned in the introduction, we have been making such calculations in recent years and the results can be found in references [1–3] mentioned above. We will just discuss here an example of a mask we have made for the Bio-Rad MRC 600 confocal scanning fluorescence microscope. In fluorescence microscopy we have no phase information and since the implementation of the image-plane mask detection system requires the subtraction of areas of the processed image from the rest we use a mask which both transmits and reflects. Two integrating detectors and an electronic subtractor process the transmitted and reflected components. For fluorescence microscopy circularly polarised light can be used and circular symmetry imposed. The image-plane mask may then be conveniently fabricated as a "binary-coded" version of the calculated continuous mask with aluminium rings deposited on silica flat. The binary ring pattern of the mask is devised to emulate the continuous profile of the calculated mask with sufficient accuracy to preserve the resolving power of the microscope while still being easy to manufacture [18]. The mask is made elliptical to be placed at 45 degrees to the optical axis and present a circular cross section to the incident light. A diagram of the mask is shown in Fig. 5.22.

An image which we have obtained recently of a field of 100 nm diameter fluorescent calibration spheres (Molecular Probes Ltd, Eugene, OR), is shown in Fig. 5.23 using the three detection schemes dicussed above, viz. type I, confocal and the KCL scheme, at an *N.A.* of 1.3 with 488 nm radiation. The inserts show profiles taken across the two spheres near the bottom left corner. The progressive increase in resolution is clearly evident. The leftmost sphere may be out of focus or at the edge of a scanning line. Absolute calibration is not yet available.

We have recently made calculations of high-aperture masks for coherent systems, [2], but have not yet implemented them experimentally. Here cylindrical sym-

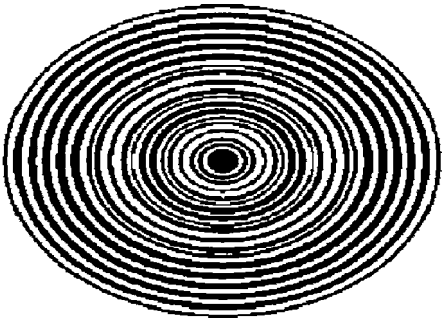


Fig. 5.22. An image-plane mask for fluorescence microscopy

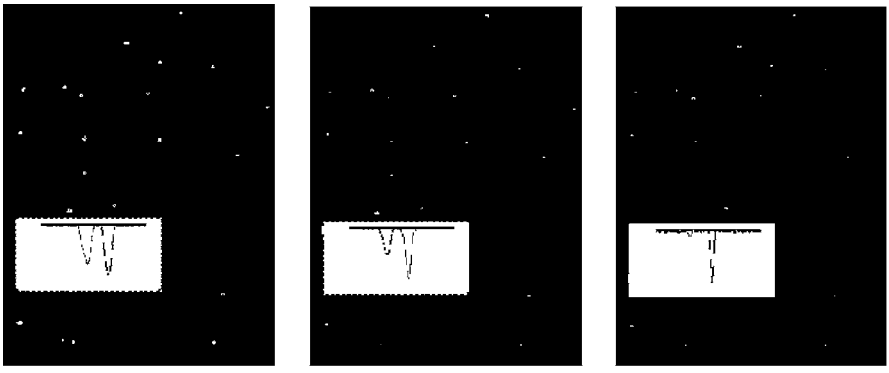


Fig. 5.23. Images of 100 nm calibration spheres in fluorescence microscopy.

metry cannot be invoked and the masks are uniaxial along the direction of the incident polarisation, which must be linear in order for the mask calculation to reduce to a linear inverse problem. An example is shown in Fig. 5.24. This work is in progress and we hope to have results in the near future. It is also possible in the coherent case to implement these masks in the pupil plane and this is also being investigated.

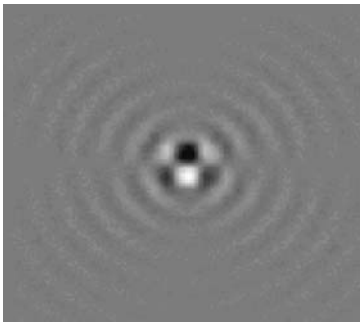


Fig. 5.24. A calculated image-plane mask for a coherent system of *N.A.* 0.6

Acknowledgements

The authors wish to thank the post-graduate students, post-doctoral fellows, academic visitors and staff who have contributed to this work over a number of years at King's College, London and who are now too many to list individually here, save perhaps for picking out Ulli Brand whose superb PhD thesis is devoted entirely to the design and construction of a superresolving confocal scanning fluorescence microscope.

Special thanks are due to Mario Bertero and his colleagues at the University of Genoa and Christine De Mol and her colleagues at the Université Libre de Bruxelles, who have been long-term collaborators in various inverse problems, including especially optical imaging. For help and joint work with us and special lens design in the early days we thank Fred Brakenhof and Hans van der Voort at the University of Amsterdam and Leitz, Wetzlar, respectively. More recently, Colin Sheppard at the University of Sydney and Peter Török at the Imperial College have engaged in helpful discussions on some of the more complex issues. As mentioned in the text, we have also benefitted from joint considerations of scanning systems with Sjoerd Stallinga, Ferry Zijp and Rob Hendriks and other members of the optical storage group at Philips Laboratories, Eindhoven; this has been within an FET programme "Super Laser Array Memory" of the European Union. Finally we acknowledge valuable collaboration with Mark Shipman and Ken Howard of University College, London in the final stages of applying our image-plane masks to biological fluorescence microscopy.

This work has been supported in its various phases by the NATO Scientific Affairs Division, the European Union, the US Army and the UK Engineering and Physical Sciences Research Council and presently, as just mentioned above, by the EU; we have also had help and encouragement from Andrew Dixon and generous support from his company, Bio-Rad Sciences Ltd.

References

1. U. Brand, G. Hester, J. Grochmalicki and E. R. Pike: J. Opt. A. **1**, 794 (1999)
2. J. Grochmalicki and E. R. Pike: Appl. Optics **39**, 6341 (2000)
3. E. R. Pike: 'Superresolution in fluorescence confocal microscopy and in DVD optical data storage'. In: *Confocal and Two-Photon Microscopy, Foundations, Applications and Advances*, ed. by A. Diaspro, (Wiley-Liss Inc., New York, 2002) pp. 499–524
4. E. R. Pike and S.-H. Jiang: J. Phys.: Condensed Matter, to appear (2002)
5. S. Stallinga: Private communications 29th Nov 2001 and 7th Jan 2002
6. J. B. Pendry: Phys. Rev. Lett. **85**, 3966 (2000)
7. E. Shamonina, V. A. Kalinin, K. H. Ringhofer and L. Solymar: Electronics Lett. **37**, 1243 (2001)
8. I. Ichimura, S. Hayashi and G. S. Kino: Appl. Optics **36**, 4339 (1997)
9. T. di Francia: Nuovo Cimento **9**, 426 (1952)
10. M. Kowalczyk, C. J. Zapata-Rodriguez and M. Martinez-Corral: Appl. Optics **37**, 8206 (1998)

11. M. Kowalczyk: private communication
12. M. Martinez-Corral, M. T. Caballero, E. H. K. Stelzer and J. Swoger: Opt. Express, **10**, 98 (2002)
13. T. R. M. Sales and G. M. Morris: J. Opt. Soc. Am. A **14**, 1637 (1997)
14. M. Bertero, P. Boccacci, R. E. Davies, F. Malfanti, E. R. Pike and J. G. Walker: Inverse Problems **8**, 1 (1992)
15. H. Hopkins: Proc. R. Soc. **A217**, 263 (1951)
16. J. Braat: 'Read-out of Optical Discs'. In: *Principles of Optical Disc Systems*, G. Bouwhuis, J. Braat. A. Huijser, J. Pasman, G. van Rosmalen and K. Shouhamer Immink (Adam Hilger, Bristol, 1985) Ch. 2
17. T. Wilson and C. Sheppard: *Theory and Practice of Scanning Optical Microscopy* (Academic Press, London 1984)
18. J. Grochmalicki, E. R. Pike, J. G. Walker, M. Bertero, P. Boccacci and R. E. Davies: J. Opt. Soc. Am. A **10**, 1074 (1993)

6 Depth of Field Control in Incoherent Hybrid Imaging Systems

Sherif Sherif, Thomas Cathey

6.1 Introduction

A hybrid imaging system combines a modified optical imaging system and a digital post-processing step. We define a new metric to quantify the blurring of a defocused image that is more suitable than the defocus parameter for hybrid imaging systems.

We describe a spatial-domain method to design a pupil phase plate to extend the depth of field of an incoherent hybrid imaging system with a rectangular aperture. We use this method to obtain a pupil phase plate to extend the depth of field, which we refer to as the logarithmic phase plate. By introducing a logarithmic phase plate at the exit pupil and digitally processing the output of the detector, the depth of field is extended by an order of magnitude more than the Hopkins defocus criterion [1]. We compare the performance of the logarithmic phase plate with other extended-depth-of-field phase plates in extending the depth of field of incoherent hybrid imaging systems with rectangular and circular apertures.

We use our new metric for defocused image blurring to design a pupil phase plate to reduce the depth of field, thereby increasing the axial resolution, of an incoherent hybrid imaging systems with a rectangular aperture. By introducing this phase plate at the exit pupil and digitally processing the output of the detector output, the depth of field is reduced by more than a factor of two.

Finally, we examine the effect of using a noisy charge-coupled device (CCD) optical detector, instead of an ideal optical detector, on the control of the depth of field.

6.2 Hybrid Imaging Systems

A hybrid imaging system is different from a system obtained by cascading a standard imaging system and a digital post-processing step. In a hybrid system, both the optical and digital modules are parts of a *single* system, and the imaging process is divided between them. Thus the final image in a hybrid system is obtained by digitally processing an intermediate optical image, as shown in Fig. 6.1. The additional digital degrees of freedom in a hybrid system can be used to improve its imaging performance beyond the best feasible performance of a similar standard system [2,3].

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

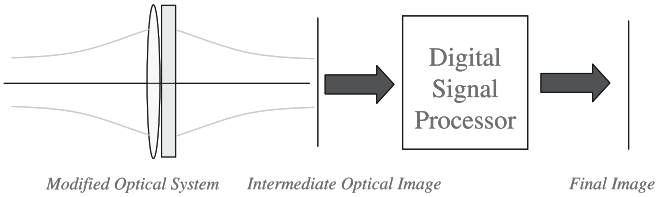


Fig. 6.1. Hybrid imaging system

In this chapter, we control the depth of field of an incoherent hybrid imaging system by introducing a phase plate at its exit pupil and digitally processing the intermediate optical image. We use a phase plate instead of an amplitude plate or a complex plate to avoid any decrease of optical power at the image plane.

6.2.1 Digital Post-Processing

The addition of a phase plate at the exit pupil of an optical system attenuates the magnitude of optical transfer function (OTF) of the original system, thereby attenuating most frequency components in the image. The addition of a phase plate at the exit pupil may also distort the phase of the original systems OTF. Thus in all the hybrid imaging systems described in this chapter, the digital post-processing steps involve a restoration digital filter, which amplifies the attenuated frequency components of the intermediate optical image and, if necessary corrects the phase of those frequency components.

For simplicity, we amplify the intermediate images attenuated frequency components and, if necessary correct their phase with a linear frequency-domain inverse filter whose frequency response is given by [4]

$$\boldsymbol{H}_{\text{inverse}}\left(f_x, f_y\right) = \begin{cases} \dfrac{\boldsymbol{H}_{\text{clear-aperture}}\left(f_x, f_y\right)}{\boldsymbol{H}_{\text{phase-plate}}\left(f_x, f_y\right)} & : \boldsymbol{H}_{\text{phase-plate}}\left(f_x, f_y\right) \neq 0 \\ 0 & : \boldsymbol{H}_{\text{phase-plate}}\left(f_x, f_y\right) = 0\,, \end{cases} \tag{6.1}$$

where $\boldsymbol{H}_{\text{clear-aperture}}$ is the in-focus OTF of the optical module with a clear aperture, without a phase plate at its exit pupil, and $\boldsymbol{H}_{\text{phase-plate}}$ is the in-focus OTF of the optical module with a phase plate at its exit pupil. Since the inverse filter, $\boldsymbol{H}_{\text{inverse}}$, is a high-pass filter, it will reduce the overall signal-to-noise ratio (SNR) of the system. This reduction in the overall SNR is one of the main drawbacks of hybrid imaging systems. We will quantify the effect of optical detector noise on hybrid imaging systems systems later in this chapter.

6.2.2 New Metric for Defocused Image Blurring

We define a new metric to quantify the blurring of a defocused image that is more suitable than the defocus parameter for hybrid imaging systems [5]. As mentioned in the previous section, a hybrid imaging system with a pupil phase plate has a blurred

intermediate optical image. A specific restoration digital filter is used to amplify the attenuated frequency components of an intermediate optical image, and, if necessary correct the phase of those frequency components, thereby obtaining a final image. The degree of digital restoration of an out-of-focus object depends on the similarity between the in-focus digital filter used, and the out-of-focus digital filter required.

The angle in Hilbert space between any two functions is a measure of the similarity between these two functions. The smaller the angle between the two functions, the more similar are the two functions and vice versa. Thus as a new metric to quantify the defocused image blurring, we choose the angle in Hilbert space between a defocused point spread function (PSF) and the in-focus PSF. This angle, $0 \le \theta \le \frac{\pi}{2}$, is defined, for any defocus parameter value, ψ, as [6]

$$\cos\theta = \frac{\left\langle |h(u,0)|^2 , |h(u,\psi)|^2 \right\rangle}{|||h(u,0)|^2|, |||h(u,\psi)|^2||} , \tag{6.2}$$

where the inner-product of the in-focus PSF, $|h(u,0)|^2$, and a defocused PSF, $|h(u,\psi)|^2$, is defined as

$$\langle |h(u,0)|^2, |h(u,\psi)|^2 \rangle = \int_{-\infty}^{\infty} |h(u,0)|^2 |h(u,\psi)|^2 \mathrm{d}u , \tag{6.3}$$

and the length in Hilbert space of the PSF for any defocus parameter value is defined as

$$|||h(u,\cdot)|^2|| = \left(\int_{-\infty}^{\infty} |h(u,\cdot)|^2 |h(u,\cdot)|^2 \mathrm{d}u \right)^{1/2} . \tag{6.4}$$

Being a measure of the similarity between a defocused PSF and the in-focus PSF, the angle θ is a more general metric to quantify the blurring of a defocused image than the defocus parameter, ψ, which, for a given imaging system, is a measure of the defocus distance only.

6.3 Extended Depth of Field

Extended depth of field (EDF) in optical imaging systems has been the goal of many researchers over the last few decades. To solve this problem, most researchers used apodization techniques on standard imaging systems. Usually, an absorbing plate with a possible $\pm\pi$ phase plate was introduced at the exit pupil to extend the depth of field [7–11]. All of these apodization-based methods share two drawbacks: a decrease of optical power at the image plane and a possible decrease in image resolution. Another method for extending the depth of field of a standard imaging system without apodization was described in [12]. In this method, the focus must be varied during exposure; hence, it is not always practical. To extend the depth of

field of an incoherent hybrid imaging system, one method used an absorbing plate at the exit pupil; hence, this method suffered from the above drawbacks [13].

In [14], a cubic phase plate was introduced at the exit pupil of an incoherent hybrid imaging system, and a depth of field extension of an order of magnitude more than the Hopkins defocus criterion [1] was achieved, without loss of optical power or image resolution at the image plane. A less than ten times increase in the depth of field of a nonparaxial hybrid imaging system was achieved using a logarithmic asphere lens [15]. A logarithmic asphere is a lens that is divided into annular rings of different focal lengths; hence, it has a continuous radial variation in its focal length. This logarithmic asphere lens differs from the rectangular logarithmic phase plate that we describe in this chapter in two fundamental ways: it is a circularly symmetric lens and it is designed by applying Fermat's principle [16].

In this section, we describe a spatial-domain method to design a pupil phase plate to extend the depth of field of an incoherent hybrid imaging system with a rectangular aperture. We use this method to obtain a pupil phase plate to extend the depth of field, which we refer to as the logarithmic phase plate. To verify that a logarithmic phase plate can be used to extend the depth of field of an incoherent diffraction-limited imaging system with a rectangular aperture, we show that the PSF and the OTF of such system with a logarithmic phase plate at its exit pupil are invariant with defocus. To demonstrate the extended depth of field, we compare two sets of computer-simulated images of a chirp pattern for different defocus parameter values. The first set is obtained using a standard incoherent system and the second set is obtained using a similar system, but with a logarithmic phase plate at its exit pupil. Finally, we compare the performance of the logarithmic phase plate with other extended-depth-of-field phase plates in extending the depth of field of hybrid imaging systems with rectangular and circular apertures.

6.3.1 Design of a Rectangular EDF Phase Plate

Mathematical Separability of PSF The PSF of a defocused paraxial imaging system with a rectangular aperture, is given by [17]

$$\left|h\left(u,v,w_x,w_y\right)\right|^2 = \left|\kappa \int_{-y_{\max}}^{y_{\max}} \int_{-x_{\max}}^{x_{\max}} \exp\left\{jk\left[\left(\frac{w_x x'^2}{x_{\max}^2}+\frac{w_y y'^2}{y_{\max}^2}\right)-\left(\frac{ux'}{z_i}+\frac{vy'}{z_i}\right)\right]\right\} \mathrm{d}x'\mathrm{d}y'\right|^2 , \tag{6.5}$$

where (u, v) and (x', y') are rectangular coordinates in the image and exit pupil planes, respectively, w_x and w_y are the defocus coefficients in the x' and y' directions, respectively, $x_{\max}$ and $y_{\max}$ are the half-widths of the aperture in the directions of x' and y', respectively, k is the propagation constant, z_i is the image distance and κ is a constant. Since this defocused PSF is mathematically separable, we can restrict our analysis to a one-dimensional (1-D) defocused PSF. We define a normalized

defocus coefficient, $w = \frac{w_x}{x_{\max}^2}$, introduce the normalized co-ordinates, $x = \frac{x'}{x_{\max}}$, introduce a phase plate, $f(x, y)$, at the exit pupil and drop all multiplicative constants, yielding

$$|h(u,w)|^2 = \left| \int_{-1}^{1} \exp\left[jk\left(w x_{\max}^2 x^2 - f(x) - \frac{u x_{\max} x}{z_i} \right) \right] \mathrm{d}x \right|^2 . \tag{6.6}$$

Axial Symmetry Condition of PSF For an extended depth of field, the PSF of an imaging system must be invariant with defocus. Any defocus-invariant PSF is also symmetric about the image plane; hence, any phase plate, $f(x)$, that extends the depth of field must satisfy the PSF axial symmetry condition,

$$|h(u,w)|^2 = |h(u,-w)|^2 . \tag{6.7}$$

It can be shown [21] that Eq. (6.7) is satisfied if and only if the phase plate, $f(x)$, is an odd function, which implies that it must not have any focusing power. If $f(x)$ has any focusing power, it will shift the location of the image plane.

Asymptotic approximation of a defocused PSF To simplify the mathematical analysis, we obtain the asymptotic approximation of the integral in a defocused PSF, Eq. (6.6), as $k \to \infty$. We use the stationary phase method [16] and [18–20], which is an excellent approximation for any large value of k (small wavelength λ). Optical wavelengths are small enough to justify the stationary phase approximation, which yields results similar to geometrical optics since both assume infinitely small wavelength. It can be shown that the asymptotic approximation of a defocused PSF of a system with a rectangular aperture, Eq. (6.6), using the stationary phase method, is given by [21]

$$|h(u,w)|^2 \simeq \left| \frac{\lambda}{2w x_{\max}^2 - f''(x_s)} \right| . \tag{6.8}$$

A more useful expression for this approximate defocused PSF is given by

$$|h(u,w)|^2 \simeq \left| \frac{\lambda\left(\frac{\partial x_s}{\partial w}\right)}{2 x_{\max}^2 x_s} \right| , \tag{6.9}$$

where $x_s(w,u)$, is a stationary point of the integral in Eq. (6.6).

Condition for Extended Depth of Field For an extended depth of field, we seek the stationary point, $x_s(w,u')$, from which we can obtain $f(x)$, such that the approximate defocused PSF, Eq. (6.9), does not change with respect to w at an arbitrary point in the image plane, u'

$$\frac{\mathrm{d}}{\mathrm{d}w}\left|\frac{\lambda\left(\frac{\partial x_s}{\partial w}\right)}{2x_{\max}^2 x_s}\right| = 0\ . \tag{6.10}$$

We solve Eq. (6.10) and choose the solution that makes $f(x)$ an odd function [21], yielding

$$f(x) = \mathrm{sgn}(x)\alpha x_{\max}^2 x^2\left(\log|x| + \beta\right) - \frac{u' x_{\max} x}{z_i}\ . \tag{6.11}$$

Because of mathematical separability, our desired two-dimensional phase plate to extend the depth of field, $f(x, y)$, is given by

$$\boxed{\begin{aligned} f(x,y) =& \mathrm{sgn}(x)\alpha x_{\max}^2 x^2\left(\log|x| + \beta\right) - \frac{u' x_{\max} x}{z_i} + \\ & \mathrm{sgn}(y)\alpha y_{\max}^2 y^2\left(\log|y| + \beta\right) - \frac{u' y_{\max} y}{z_i}\ . \end{aligned}} \tag{6.12}$$

We refer to $f(x, y)$ as the logarithmic phase plate.

Optimum Logarithmic Phase Plate We obtain the optimum values of the logarithmic phase plate parameters, α, β and u', for an arbitrary paraxial imaging system, using Zemax: an optical design program [22]. Using Zemax, we specify an incoherent imaging system, whose parameters are shown in Table 6.1, and with a logarithmic phase plate at its exit pupil.

Table 6.1. Imaging system parameters

λ (nm)	f(mm)	$F\#$	z_i (mm)	M
587.56	20.0	4.0	40.0	-1

We use the numerical optimization routine of Zemax to obtain the optimum values of α, β and u' such that an objective function is minimized.

The first part of our objective function is defined as the difference between various points of the in-focus OTF and their corresponding points of different defocused OTFs. If this part of our objective function is minimized, the systems OTF will be as defocus invariant as possible. The second part of our objective function is defined as the difference between various points of the in-focus OTF and their corresponding points of the in-focus diffraction-limited OTF. If this part of the objective function is minimized, the systems OTF will have the highest values possible. In essence, the minimization of our objective function yields the highest value of the OTF for a system that is as invariant with defocus as possible.

The optimum values of the logarithmic phase plate parameters, which correspond to the imaging system parameters shown in Table 6.1, are shown in Table 6.2.

Table 6.2. Logarithmic phase plate optimum parameters

α (mm)	β	u' (mm)
4.23×10^{-4}	0.57	-2.18×10^{-4}

The initial value of our objective function, which corresponds to $\alpha = 0, \beta = 0$ and $u' = 0$ is 0.51221. The optimum value of our objective function, which corresponds to the optimum parameters in Table 6.2 is 0.04000. The value of our optimum objective function increases from 0.0400 to 0.04031, which is a change of only 0.06% of the initial value of our objective function, when we set $u' = 0$. Thus u' has a negligible effect in extending the depth of field and will be ignored. Thus the practical values of the logarithmic phase plate parameters, which correspond to the imaging system parameters shown in Table 6.1, are shown in Table 6.3.

Table 6.3. Logarithmic phase plate practical parameters

α (mm)	β	u' (mm)
4.23×10^{-4}	0.57	0.0

The profile of the logarithmic phase plate, whose parameters are shown in Table 6.3, is shown in Fig. 6.2.

6.3.2 Performance of a Logarithmic Phase Plate

Defocused PSF Using a Logarithmic Phase Plate The PSF of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, and with a logarithmic phase plate, whose parameters are shown in Table 6.3, at its exit pupil is shown in Fig. 6.3 for different values of the defocus parameter ψ.

We note that, apart from a slight lateral shift with defocus, the shape and the intensity of the PSF shown in Fig. 6.3 are invariant with defocus over a wide range of defocus parameter values, compared to the PSF of a similar defocused standard diffraction-limited imaging system, shown in Fig. 6.4.

A quantitative way to show that the PSF of a diffraction-limited system with a logarithmic phase plate at its exit pupil is invariant with defocus, compared to the PSF of a similar standard imaging system, is to evaluate the angle in Hilbert space between the in-focus PSF and different defocused PSFs, Eq. (6.2).

In Fig. 6.5, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited standard imaging system, whose parameters are shown in Table 6.1. On the same figure, Fig. 6.5, we also show the angle in Hilbert space between the in-focus PSF and defocused PSFs of the same imaging

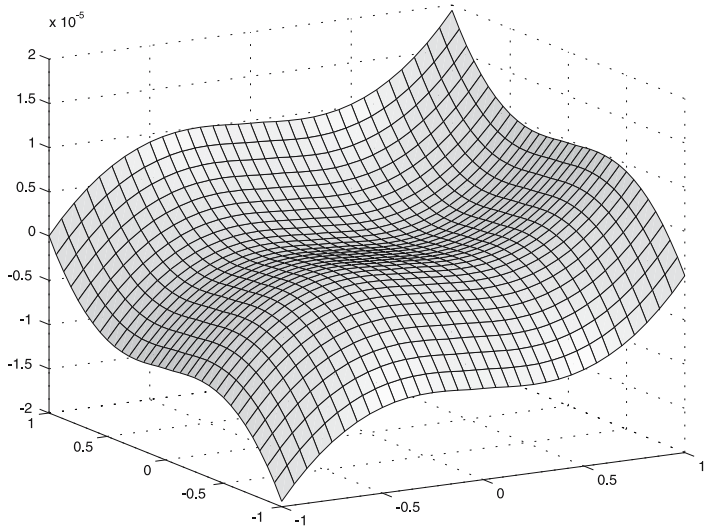


Fig. 6.2. Profile of a logarithmic phase plate

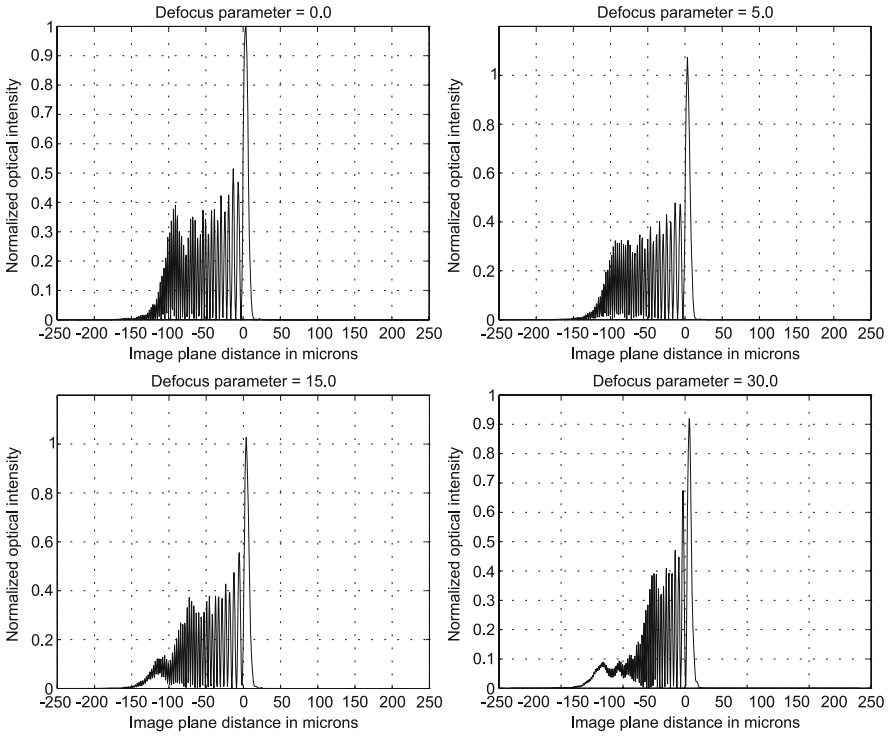


Fig. 6.3. Defocused diffraction-limited PSF using a logarithmic phase plate

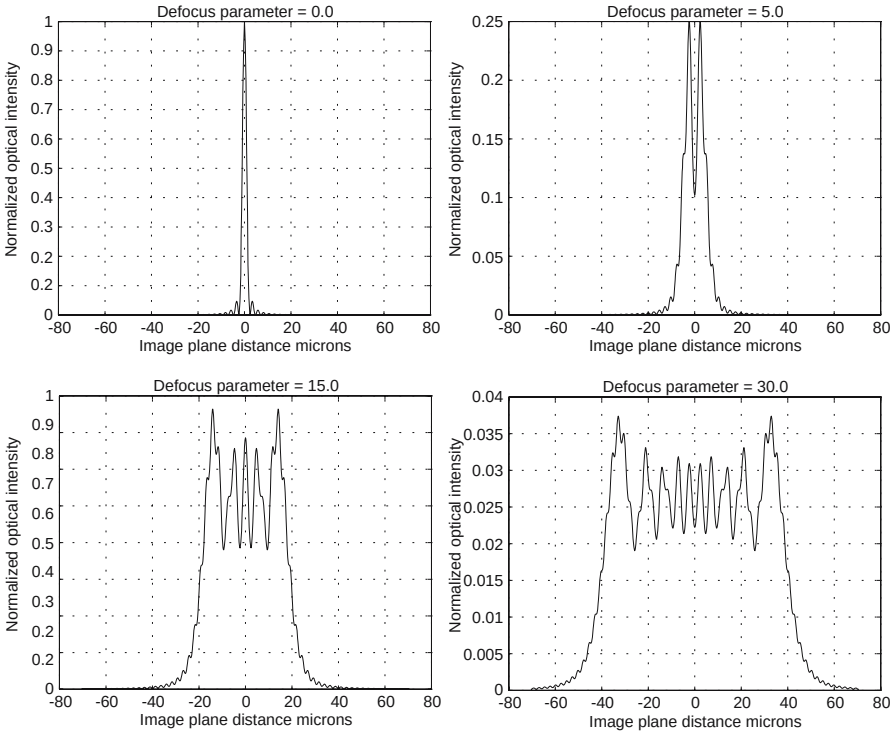


Fig. 6.4. Defocused diffraction-limited PSF using a clear rectangular aperture

system, but with a logarithmic phase plate, whose parameters are shown in Table 6.3, at its exit pupil.

From Fig. 6.5, we note that, for all shown defocus parameter values, the angle in Hilbert space between the in-focus PSF and defocused PSFs of a system with a logarithmic phase plate at its exit pupil has a smaller value than the corresponding angle in Hilbert space between the in-focus PSF and defocused PSFs of a standard system. Thus the PSF of a diffraction-limited imaging system with a logarithmic phase plate at its exit pupil is invariant with defocus, compared to the PSF of a similar standard diffraction-limited imaging system.

Defocused OTF Using a Logarithmic Phase Plate The OTF of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, and with a logarithmic phase plate, whose parameters are shown in Table 6.3, at its exit pupil is shown in Fig. 6.6 for different values of the defocus parameter ψ.

We note that the OTF shown in Fig. 6.6 is invariant with defocus over a wide range of defocus parameter values, compared to the OTF of a similar defocused imaging system, shown in Fig. 6.7. We also note that for all defocus parameter

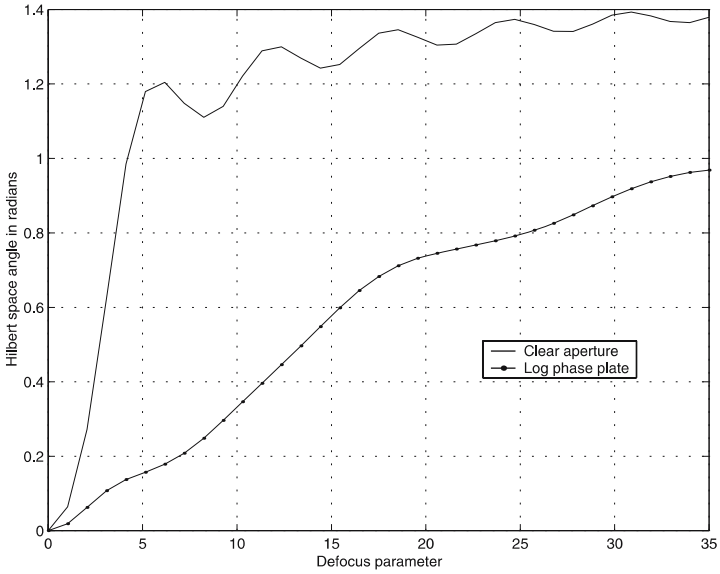


Fig. 6.5. Diffraction-limited Hilbert space angles using a clear rectangular aperture and using a logarithmic phase plate

values, the OTF shown in Fig. 6.6 has no nulls; hence, there is no loss of spatial frequencies in the image.

The Woodward function of the pupil function of an imaging system, with a rectangular aperture, represents a polar display of the systems OTFs for different values of defocus [23]. The Woodward function of the pupil function of a diffraction-limited imaging system with a logarithmic phase plate at its exit pupil is shown in Fig. 6.8.

On comparing the Woodward function shown in Fig. 6.8 with the Woodward function of the pupil function of a standard imaging system, shown in Fig. 6.9, we note that radial lines through the origin of the Woodward function shown in Fig. 6.8 have nearly the same values as a function of angle, for a broad range of angles.

We also note that the Woodward function shown in Fig. 6.8 has non-zero values uniformly distributed along the normalized frequency axis. Thus the OTF of an imaging system with a logarithmic phase plate at its exit pupil is invariant with defocus over a wide range of defocus parameter values compared to the OTF of a defocused standard imaging system.

Simulated Imaging Example To demonstrate the extended depth of field, we compare two sets of computer-simulated images of a chirp pattern for different defocus parameter values. On the left column of Fig. 6.10, we show computer-simulated images of a chirp pattern, for different defocus parameter values, that we obtained using an incoherent standard diffraction-limited imaging system, whose parame-

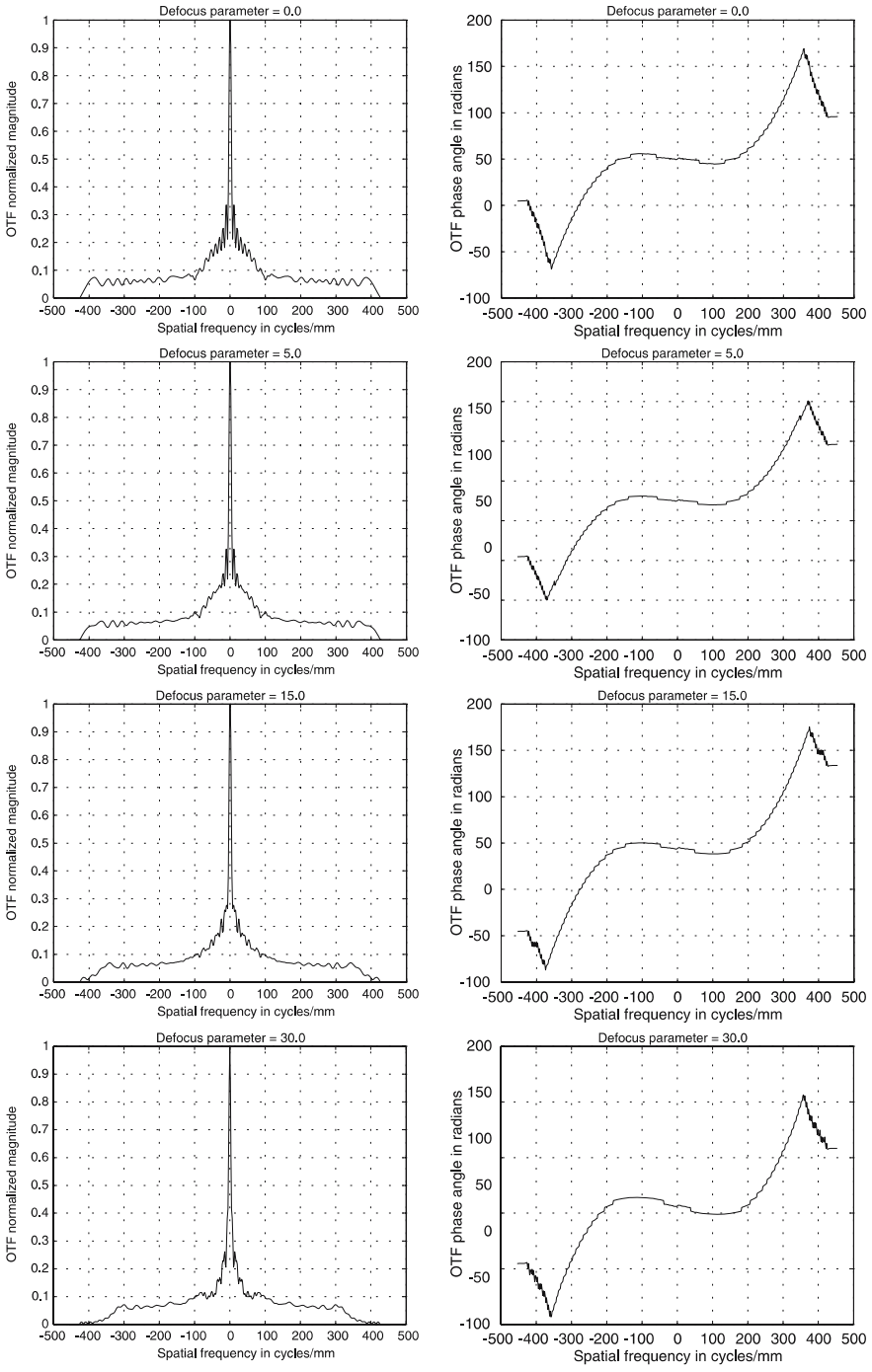


Fig. 6.6. Defocused diffraction-limited OTF using a logarithmic phase plate

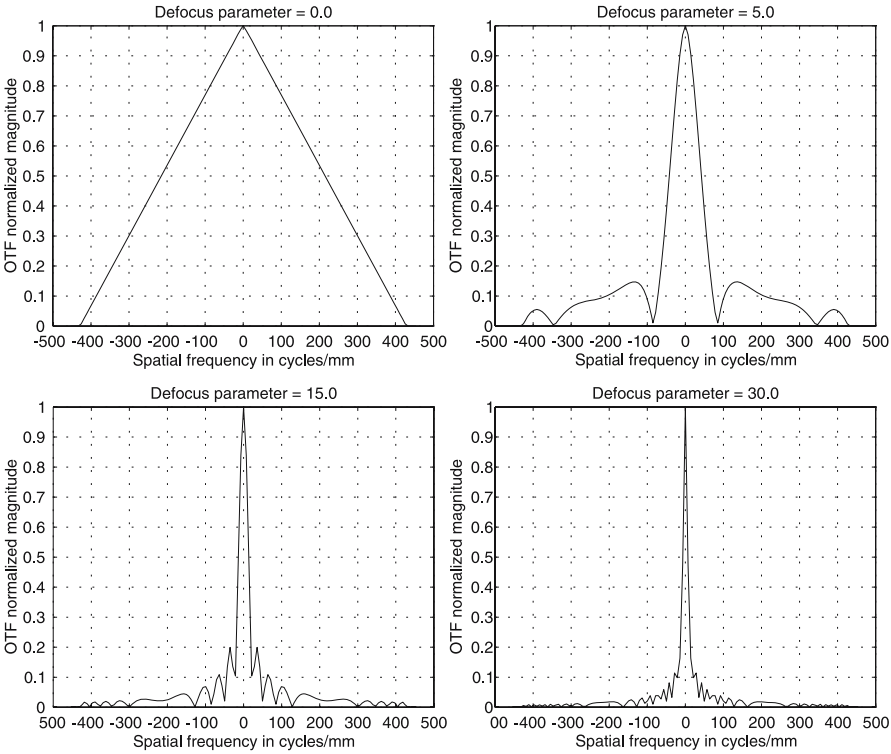


Fig. 6.7. Defocused diffraction-limited OTF using a clear rectangular aperture

ters are shown in Table 6.1. On the right column of Fig. 6.10, we show computer-simulated images of the same chirp pattern, for different defocus parameter values, that we obtained using a similar imaging system with a logarithmic phase plate, whose parameters are shown in Table 6.3, at its exit pupil. In each column, the value of the defocus parameter is changed from 0 (in-focus) to 30.

We obtained these images by using an inverse filter whose frequency response is given by [4]

$$\boldsymbol{H}_{\text{inverse}}\left(f_x, f_y\right) = \begin{cases} \frac{\boldsymbol{H}_{\text{clear-aperture}}\left(f_x, f_y\right)}{\boldsymbol{H}_{\text{log-plate}}\left(f_x, f_y\right)} & : \boldsymbol{H}_{\text{log-plate}}\left(f_x, f_y\right) \neq 0 \\ 0 & : \boldsymbol{H}_{\text{log-plate}}\left(f_x, f_y\right) = 0\,, \end{cases} \tag{6.13}$$

where $\boldsymbol{H}_{\text{clear-aperture}}$ is the in-focus OTF of the diffraction-limited imaging system with a clear aperture, without a phase plate at its exit pupil, and $\boldsymbol{H}_{\text{log-plate}}$ is the in-focus OTF of the diffraction-limited imaging system with the logarithmic phase plate at its exit pupil.

From Fig. 6.10, we note that the imaging system with a logarithmic phase plate at its exit pupil has a depth of field that is an order of magnitude more than the Hopkins defocus criterion.

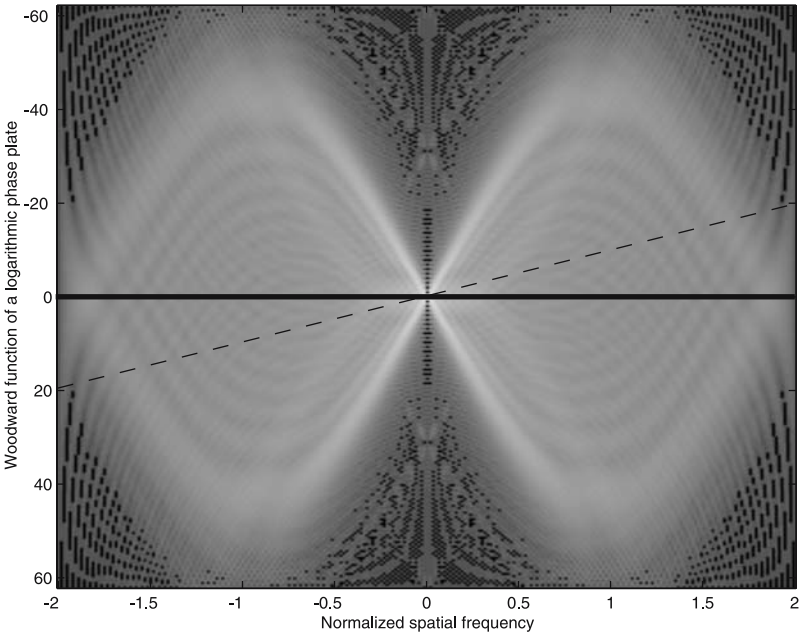


Fig. 6.8. Woodward function of a logarithmic phase plate

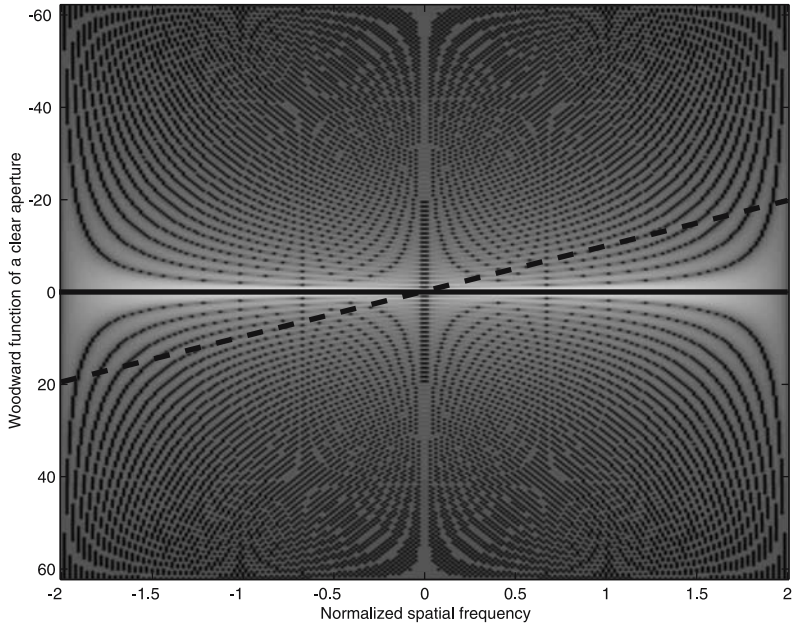


Fig. 6.9. Woodward function of a clear rectangular aperture

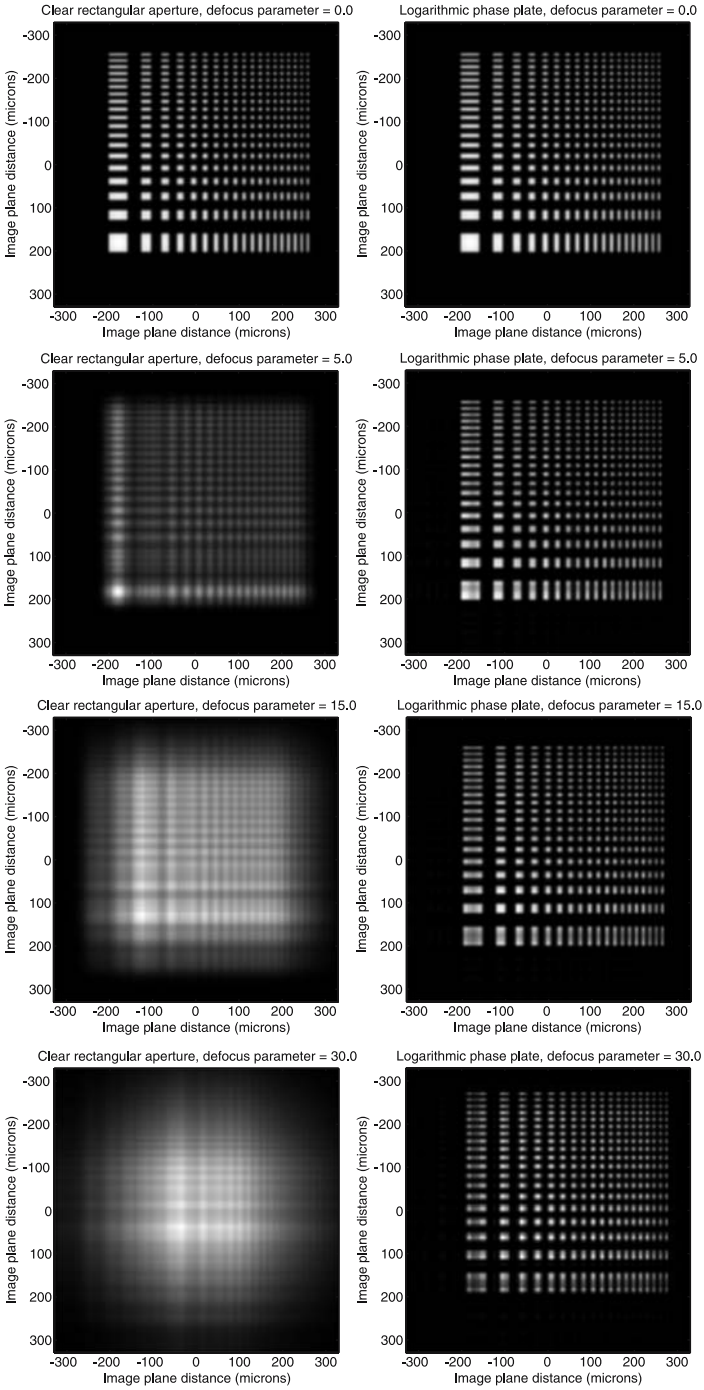


Fig. 6.10. Defocused diffraction-limited images using a clear rectangular aperture and using a logarithmic phase plate

6.3.3 Performance Comparison of Different EDF Phase Plates

Imaging System with a Rectangular Aperture In this section, we compare the performance of a logarithmic phase plate with a cubic phase plate [14] in extending the depth of field of a diffraction-limited imaging system with a rectangular aperture. A cubic phase plate is mathematically written as [14]

$$f(x, y) = \alpha(x^3 + y^3) \, . \tag{6.14}$$

We obtain the optimum value of the cubic phase plate parameter, α, using Zemax. We use the numerical optimization routine of Zemax to minimize the same objective function that we used to obtain the optimum values of the logarithmic phase plate parameters shown in Table 6.3. For the diffraction-limited imaging system whose parameters are shown in Table 6.1, the optimum value of the cubic phase plate parameter, α, is 0.005 mm.

In Fig. 6.11, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited standard imaging system, whose parameters are shown in Table 6.1, with a cubic phase plate with $\alpha = 0.005$ mm. On the same figure, Fig. 6.11, we also show the angle in Hilbert space between the in-focus PSF and defocused PSFs of the same imaging system, with a logarithmic phase plate, whose parameters are shown in Table 6.3, at its exit pupil.

From Fig. 6.11, we note that, for all defocus parameter values less than 15, the angle in Hilbert space between the in-focus PSF and defocused PSFs of a system

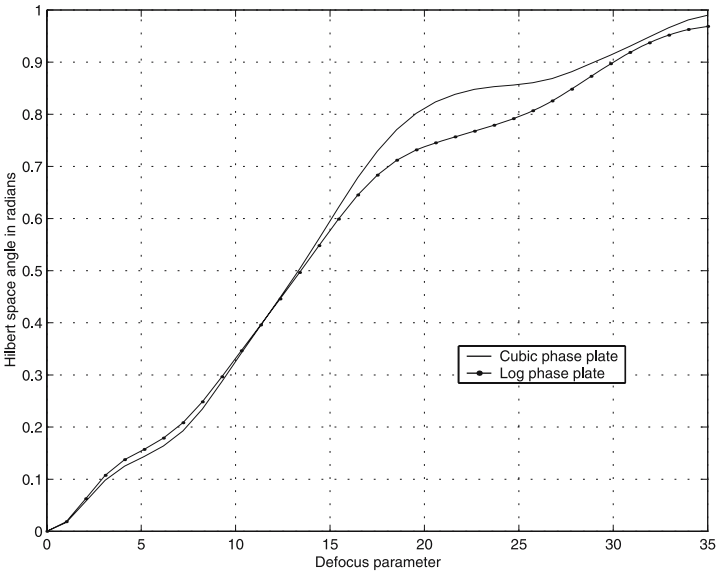


Fig. 6.11. Diffraction-limited Hilbert space angles using a cubic phase plate and using a logarithmic phase plate

with a cubic phase plate at its exit pupil has an approximately equal value to the corresponding angle in Hilbert space between the in-focus PSF and defocused PSFs of a system with a logarithmic phase plate at its exit pupil. For all defocus parameter values between 15 and 30, the angle in Hilbert space between the in-focus PSF and defocused PSFs of a system with a cubic phase plate at its exit pupil is slightly larger than the corresponding angle in Hilbert space between the in-focus PSF and defocused PSFs of a system with a logarithmic phase plate at its exit pupil.

Thus the performances of the logarithmic phase plate and the cubic phase plate in extending the depth of field of an imaging system with a rectangular aperture are very similar, with a slight advantage of the logarithmic phase plate for higher defocus parameter values. However, it is important to note that the values of the parameters of a plate are very critical to the its performance.

Imaging System with a Circular Aperture In this section, we compare the performance of a rectangular logarithmic phase plate, a rectangular cubic phase plate, a logarithmic asphere [15], and an EDF circular phase plate [26] in extending the depth of field of a diffraction-limited hybrid imaging system with a circular aperture. A logarithmic asphere, $f(r)$, is a circularly symmetric, stand-alone phase imaging element and is given by [15]

$$f(r) = \left\{ \left[\sqrt{r^2 + z_i^2} - z_i \right] + \frac{r_{\max}^2}{2(z_2 - z_1)} \times \left[\log \left(2 \frac{z_2 - z_1}{r_{\max}^2} \left\{ \sqrt{r^2 + \left(z_1 + \frac{z_2 - z_1}{r_{\max}^2} \right)^2} + \left(z_1 + \frac{z_2 - z_1}{r_{\max}^2} \right) \right\} + 1 \right) - \log \left(4 \frac{z_2 - z_1}{r_{\max}^2} z_1 + 1 \right) \right] \right\} \tag{6.15}$$

where z_i is the in-focus image distance, z_1 and z_2 are defocused image distances and $r_{\max}$ is the radius of the aperture. An EDF circular plate, $f(r, \theta)$, is a non-circularly symmetric pupil phase plate and is given by [24–26]

$$f(r, \theta) = \begin{cases} f_R(r, \theta) & : \quad -\frac{\pi}{2} \le \theta \le \frac{\pi}{2} \\ f_L(r, \theta) & : \quad -\frac{\pi}{2} \ge \theta \ge \frac{\pi}{2} \, , \end{cases} \tag{6.16}$$

where

$$f_L(r, \theta) = \alpha r^{5.33} \theta^3 \tag{6.17}$$

and

$$f_L(r, \theta) = -f_R(r, \theta - \pi) \tag{6.18}$$

In Fig. 6.12, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited standard imaging system, whose parameters are

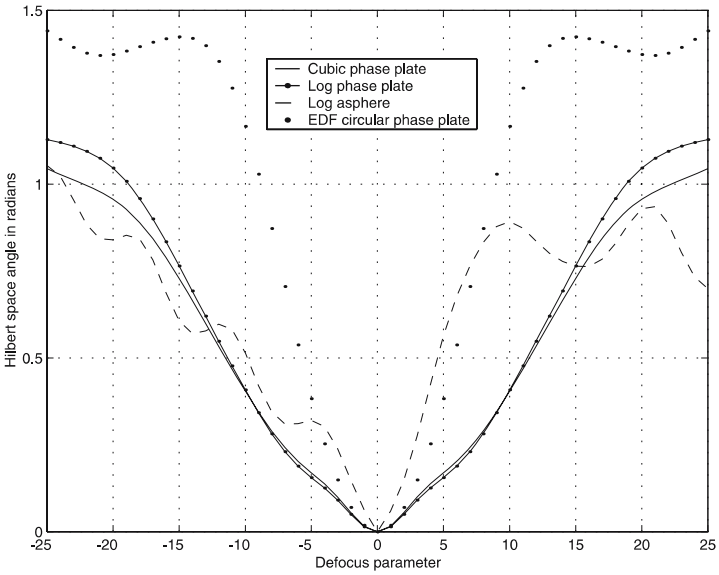


Fig. 6.12. Diffraction-limited Hilbert space angles using different EDF phase plates

shown in Table 6.1, with different EDF phase plates at its exit pupil. The first EDF phase plate is a logarithmic phase plate, whose parameters are shown in Table 6.3. The second EDF phase plate is a cubic phase plate with $\alpha = 0.005\,\text{mm}$. The third EDF phase plate is an EDF circular plate with $\alpha = 0.0031\,\text{mm}$. On the same figure, Fig. 6.12, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, which uses a logarithmic asphere, Eq. (6.15), instead of a lens. We choose $z_1 = 39.64\,\text{mm}$ and $z_2 = 40.36\,\text{mm}$ which, for the given imaging system, are the distances equivalent to defocus parameters $\psi = -30$ and $\psi = 30$, respectively.

From Fig. 6.12, we note that the angle in Hilbert space between the in-focus and defocused PSFs of an imaging system that uses a logarithmic asphere is not symmetric about the image plane. We also note that, in general, the angle of a system with a cubic phase plate or a logarithmic phase plate at its exit pupil is lower than the corresponding angle of a system with an EDF circular phase plate at its exit pupil or the corresponding angle of a diffraction-limited imaging system which uses a logarithmic asphere.

Thus both performances of the logarithmic phase plate and the cubic phase plate are better than both performances of the EDF circular phase plate and the logarithmic asphere in extending the depth of field of an imaging system with a circular aperture. The performance of the cubic phase plate has a slight advantage over the logarithmic phase plate for higher defocus parameter values. However, it is important to note that the values of the parameters of a plate are very critical to the its performance.

6.4 Reduced Depth of Field

Reduced depth of field (RDF), thereby increased axial resolution, in optical imaging systems is of great importance in three-dimensional (3-D) imaging. If the magnitudes of the defocused optical transfer functions (OTFs) of a hybrid imaging system have no nulls within the pass-band of the system, there will be no irrecoverable loss of information (e.g. missing cone) in the axial direction. When there is no irrecoverable loss of information in the axial direction, the lower the depth of field, the higher the optical-sectioning capacity of an imaging system [27,28].

However, compared to extended depth of field, not too much attention has been given to this problem. Continuously varying amplitude pupil plates [29], and annular binary pupil plates, [30,31], were used to reduce the length of the central lobe of the axial intensity PSF of a standard imaging systems. Similar to the EDF problem, these apodization-based methods to reduce the depth of field share two drawbacks: a decrease of optical power at the image plane and a possible decrease in image resolution. In [32], a phase-only pupil plate was used to reduce the axial spot size of a confocal scanning microscope. Another method to reduce the depth of field of a standard optical system used structured illumination [33]. In this method, three images taken at different positions are processed to produce a single reduced depth of field image; hence, it is not always efficient.

In this section, we use our new metric for defocused image blurring to design a pupil phase plate to reduce the depth of field, thereby increasing the axial resolution, of an incoherent hybrid imaging system with a rectangular aperture. By introducing this phase plate at the exit pupil and digitally processing the output of the detector, the depth of field is reduced by more than a factor of two.

6.4.1 Design of a Rectangular RDF Phase Plate

1-D Defocused PSF Similar to our approach to extend the depth of field, we reduce the depth of field by introducing a phase plate at the exit pupil of the imaging system and digitally processing the intermediate optical image. Following our analysis in Sect. 3.1, the 1-D PSF of a defocused paraxial imaging system with a phase plate, $f(x)$, at its exit pupil can be written as

$$|h(u,w)|^2 = \left| \int_{-1}^{1} \exp\left\{ j\psi_x x^2 - jk\left[f(x) + \frac{ux_{\max}x}{z_i} \right] \right\} dx \right|^2 , \tag{6.19}$$

where $\psi_x = kw_x$ is the defocus parameter in the x direction.

Condition for Reduced Depth of Field For a reduced depth of field, we seek a phase plate which results in maximum image blurring at a slightly defocused plane that is specified by a relatively small defocus parameter value. To obtain our desired phase plate, $f(x)$, which reduces the depth of field, we substitute Eq. (6.19) into Eq. (6.2) and solve the optimization problem,

$$\min_{f} \frac{\left\langle |h(u,0)|^2, |h(u,\psi)|^2 \right\rangle}{|||h(u,0)|^2||, |||h(u,\psi)|^2||} , \tag{6.20}$$

for a relatively small value of the defocus parameter ψ_x.

Rectangular RDF phase plate We solve the optimization problem, Eq. (6.20), for $\psi_x = 1$ by assuming that our desired phase plate, $f(x)$, is a periodic phase grating [5]. We assume $f(x)$ to be a periodic grating so that the PSF of the imaging system is an array of narrowly spaced spots. An array of narrowly spaced spots would change its overall shape considerably, as its spots spread out and overlap due to propagation. As mentioned in Sect. 3.1, to avoid any shifting in the location of the image plane, $f(x)$ must not have any focusing power. Thus we represent $f(x)$ by an odd and finite Fourier series,

$$f(x) = \sum_{n=1}^{N} b_n \sin(2\pi n\nu x) . \tag{6.21}$$

After substituting Eq. (6.21) into Eq. (6.20), we obtain the optimum values of the fundamental spatial frequency, ν, and the coefficients, b_n, numerically by using the method of steepest descent [34].

The initial value of θ which corresponds to a system with a clear aperture is 0.0601 radians. The optimum value of θ which corresponds to a system with a periodic phase grating, Eq. (6.21), and with a number of coefficients $N = 5$ is 0.1355 radians. We note that an increase in the number of coefficients from $N = 4$ to $N = 5$ results in a negligible 0.15%, in the optimal value of θ. Thus we restrict the number of phase grating coefficients to $N = 5$. The optimum values of the fundamental frequency, ν, and the coefficients, b_n, which correspond to an F/4 imaging system are shown in 6.4.

Table 6.4. Rectangular RDF phase grating optimum parameters

ν(cycles/$x_{\max}$)	b_1(μm)	b_2(μm)	b_3(μm)	b_4(μm)	b_5(μm)
1.04	1.1705	-0.0437	0.0271	-0.0325	-0.007

Because of mathematical separability, our desired 2-D phase plate to reduce the depth of field, $f(x,y)$, is given by

$$\boxed{f(x,y) = \sum_{n=1}^{N} b_n \sin(2\pi n\nu x) + \sum_{n=1}^{N} b_n \sin(2\pi n\nu y)} . \tag{6.22}$$

We refer to $f(x,y)$, whose coefficients are shown in 6.4, as the rectangular RDF phase grating. The profile of the rectangular RDF phase grating is shown in Fig. 6.13

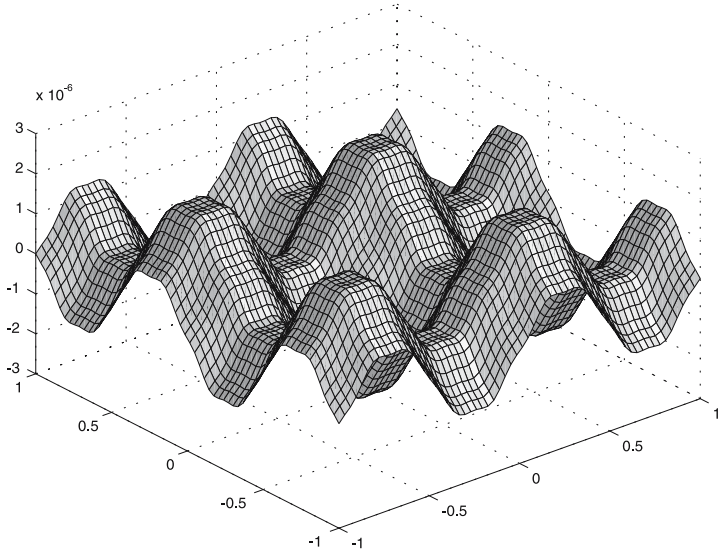


Fig. 6.13. Profile of a rectangular RDF phase grating

6.4.2 Performance of a Rectangular RDF Phase Grating

Defocused PSF Using a Rectangular RDF Phase Grating The PSF of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, and with an RDF phase grating, whose parameters are shown in Table 6.4, at its exit pupil is shown in Fig. 6.14, for different values of the defocus parameter ψ.

We note that the variation with defocus in the shape of the PSF shown in Fig. 6.14 is greater than the variation with defocus in the shape of the PSF of a similar standard diffraction-limited system, shown in Fig. 6.15.

In Fig. 6.16, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited standard imaging system, whose parameters are shown in Table 6.1. On the same figure, Fig. 6.16, we also show the angle in Hilbert space between the in-focus PSF and defocused PSFs of the same imaging system, but with an RDF phase plate, whose parameters are shown in Table 6.4, at its exit pupil.

From Fig. 6.16, we note that, for all shown defocus parameter values, the angle in Hilbert space between the in-focus PSF and defocused PSFs of a system with a rectangular RDF phase grating at its exit pupil, has a greater value than the corresponding angle in Hilbert space between the in-focus PSF and defocused PSFs of a standard system. Thus the shape of the PSF of a diffraction-limited imaging system, with a rectangular RDF phase grating at its exit pupil, varies with defocus more than the shape of the PSF of a similar standard diffraction-limited imaging system. Furthermore, we note that, for lower defocus parameter values, the angle in Hilbert space is larger than the corresponding angle between the in-focus PSF and defocused PSFs of a standard system by more than a factor of two. Thus a diffraction-

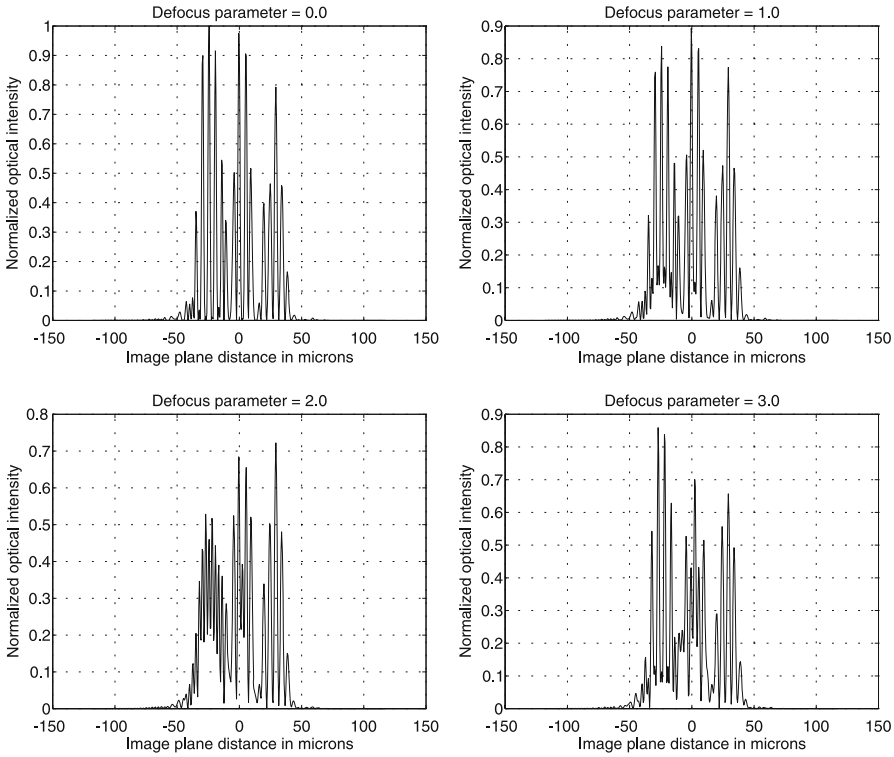


Fig. 6.14. Defocused diffraction-limited PSF using a rectangular RDF phase grating

limited system with a rectangular RDF phase grating at its exit pupil has less than half the depth of field of a similar diffraction-limited system.

Defocused OTF Using a Rectangular RDF Phase Grating The OTF of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, and with an RDF phase grating, whose parameters are shown in Table 6.4, at its exit pupil is shown in Fig. 6.17 for different values of the defocus parameter ψ.

We note that there is rapid variation with defocus in the phase of the OTF shown in Fig. 6.17. Thus the variation with defocus in this OTF is greater than the variation with defocus in the OTF of a similar standard diffraction-limited system, shown in Fig. 6.18. We also note that the in-focus OTF shown in Fig. 6.17 has no nulls; hence, there is no loss of spatial frequencies in the image.

Simulated Imaging Example To demonstrate the reduced depth of field, we compare two sets of computer-simulated images of a chirp pattern for different defocus parameter values. On the left column of Fig. 6.19, we show computer-simulated images of a chirp pattern, for different defocus parameter values, that we obtained

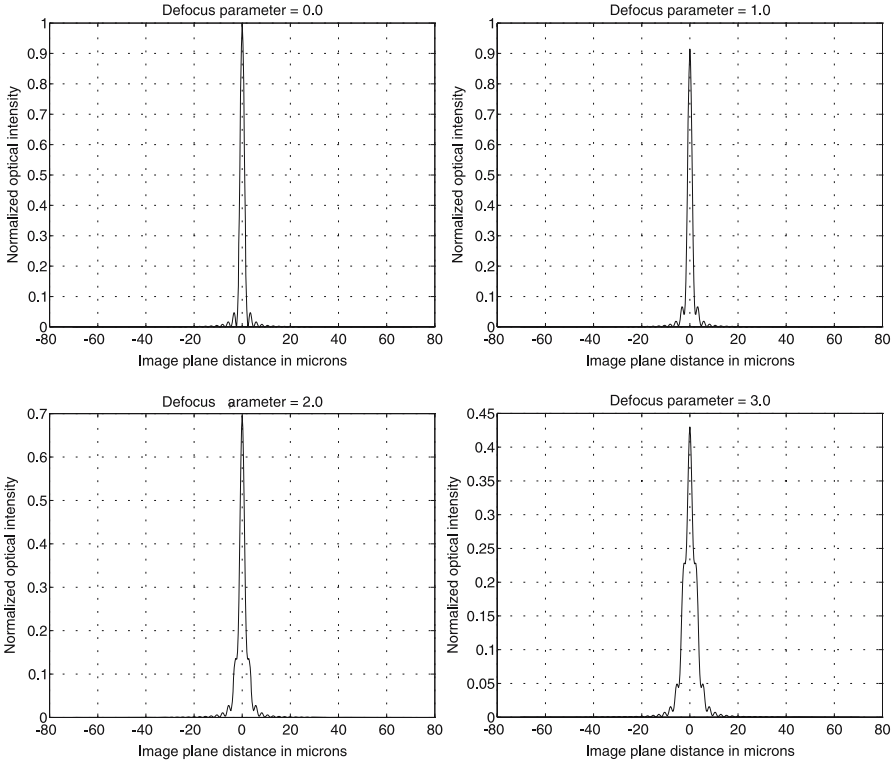


Fig. 6.15. Defocused diffraction-limited PSF using a clear rectangular aperture

using an incoherent standard diffraction-limited imaging system, whose parameters are shown in Table 6.1. On the right column of Fig. 6.19, we show computer-simulated images of the same chirp pattern, for different defocus parameter values, that we obtained using a similar imaging system with an RDF phase grating, whose parameters are shown in Table 6.4, at its exit pupil. In each column, the value of the defocus parameter is changed from 0.0 (in-focus) to 3.0.

We obtained these images by using an inverse filter whose frequency response is given by [4]

$$\boldsymbol{H}_{\text{inverse}}\left(f_x, f_y\right) = \begin{cases} \dfrac{\boldsymbol{H}_{\text{clear-aperture}}\left(f_x, f_y\right)}{\boldsymbol{H}_{\text{RDF-grt}}\left(f_x, f_y\right)} & : \boldsymbol{H}_{\text{RDF-grt}}\left(f_x, f_y\right) \neq 0 \\ 0 & : \boldsymbol{H}_{\text{RDF-grt}}\left(f_x, f_y\right) = 0\,, \end{cases} \tag{6.23}$$

where $\boldsymbol{H}_{\text{clear-aperture}}$ is the in-focus OTF of the diffraction-limited imaging system with a clear aperture, without a phase plate at its exit pupil, and $\boldsymbol{H}_{\text{RDF-grt}}$ is the in-focus OTF of the diffraction-limited imaging system with the EDF phase grating at its exit pupil.

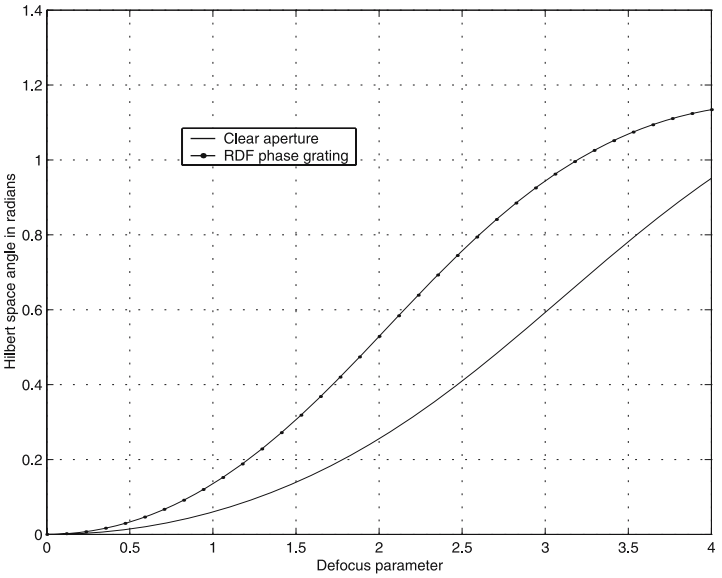


Fig. 6.16. Diffraction-limited Hilbert space angles using a clear rectangular aperture and using a rectangular RDF phase grating

From Fig. 6.19, we note that the depth of field of the imaging system with a rectangular RDF phase grating at its exit pupil is reduced, compared to the depth of field of the standard system.

6.5 Effect of Optical Detector on Depth of Field Control

6.5.1 Effect of Additive White Noise at the Optical Detector

As mentioned earlier, a pupil-coded hybrid imaging system needs a digital filter to restore the attenuated frequency components of the intermediate optical image and, if necessary, to correct the phase. This restoration filter will be usually a high-pass digital filter, so it will increase the noise power in the final image compared the optical intermediate image. Assuming additive white noise at the detector, the increased noise power in the final image is equal to the detector noise power multiplied by a noise factor [35]

$$F_{\text{noise}} = \frac{\sum_{i=1}^{M} \sum_{j=1}^{N} h_{\text{digital}}^2(i) h_{\text{digital}}^2(j)}{\left[\sum_{i=1}^{M} h_{\text{digital}}^2(i)\right]^4} , \tag{6.24}$$

where h_{digital} is the impulse response of the restoration filter. Thus linear signal processing will typically decrease the SNR of the final image of the hybrid system by $1/F_{\text{noise}}$ compared to the intermediate image.

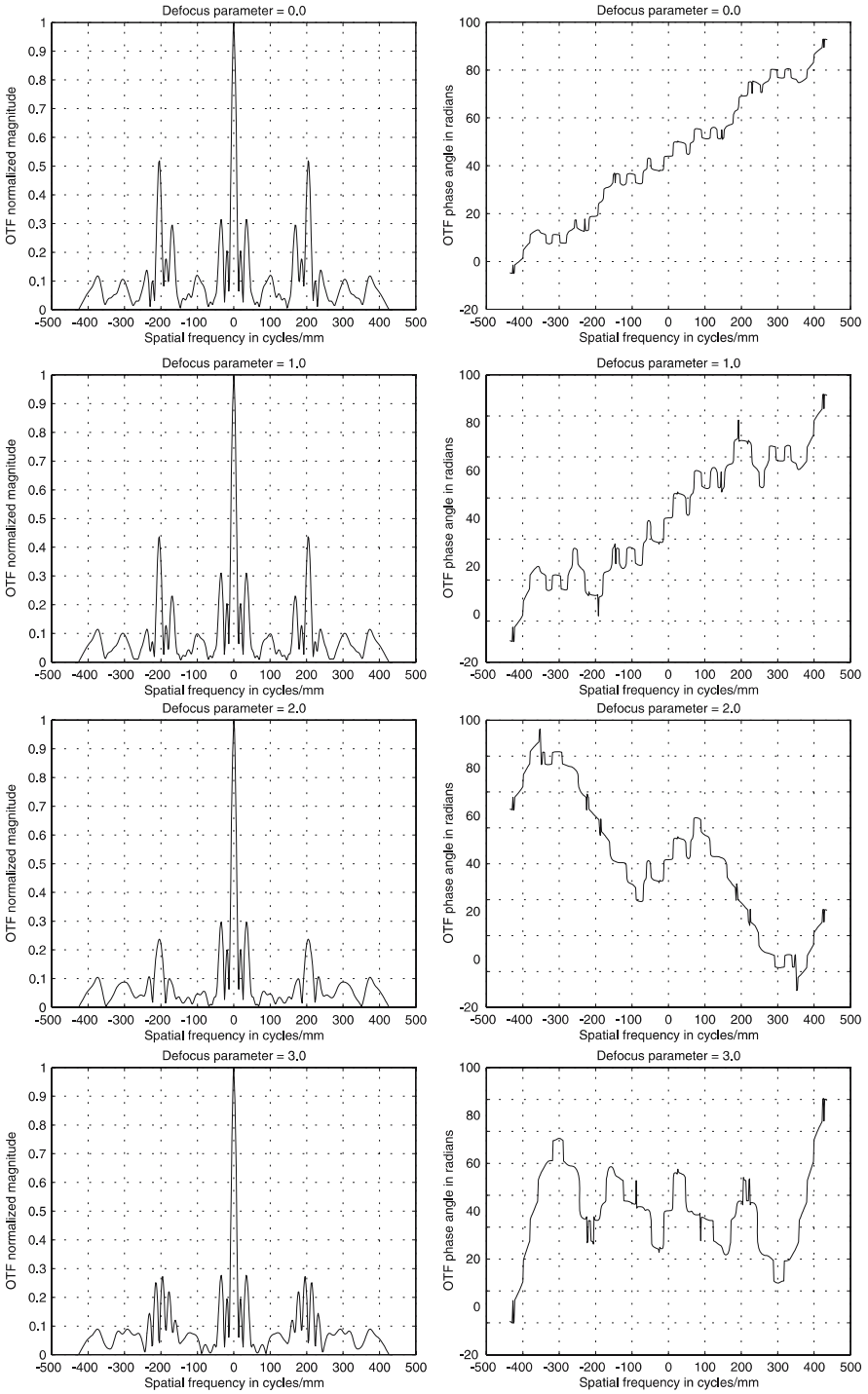


Fig. 6.17. Defocused diffraction-limited OTF using a rectangular RDF phase grating

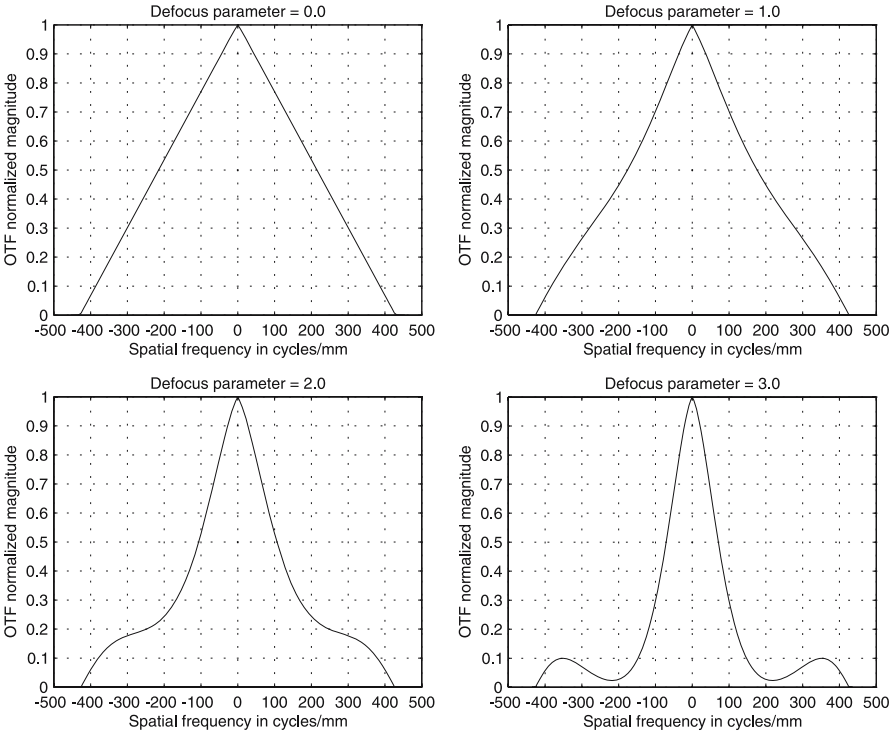


Fig. 6.18. Defocused diffraction-limited OTF using a clear rectangular aperture

In Fig. 6.20 we show the variation with defocus of the angle in Hilbert space for different logarithmic phase plates with relative thickness $a = \frac{\alpha}{\alpha_{\max}}$, where $\alpha_{\max}$ corresponds to the plate thickness which results in the maximum extension of depth of field without any loss of resolution. We also include the reduction of the SNR in the final image $\left(\frac{1}{F_{\text{noise}}}\right)_{\text{dB}} = -10 \log_{10} F_{\text{noise}}$ for each phase plate, so one could immediately realize the noise related cost for a certain increase in the depth of field.

For example, the uppermost curve in Fig. 6.20 corresponds to a system with a clear aperture; hence, there is no extension of the depth of field or decrease in the SNR of the final image. The lowermost curve on the same figure corresponds to a system with the thickest possible logarithmic phase plate that would cause the maximum extension of depth of field without any loss in resolution. In this case the SNR of the final image is decreased by 15.87 dB.

6.5.2 Charge-Coupled Device-Limited PSF

A charge-coupled device (CCD) detector can be modeled as an array of rectangular pixels. Assuming uniform responsivity across each and every pixel, the 1-D PSF of

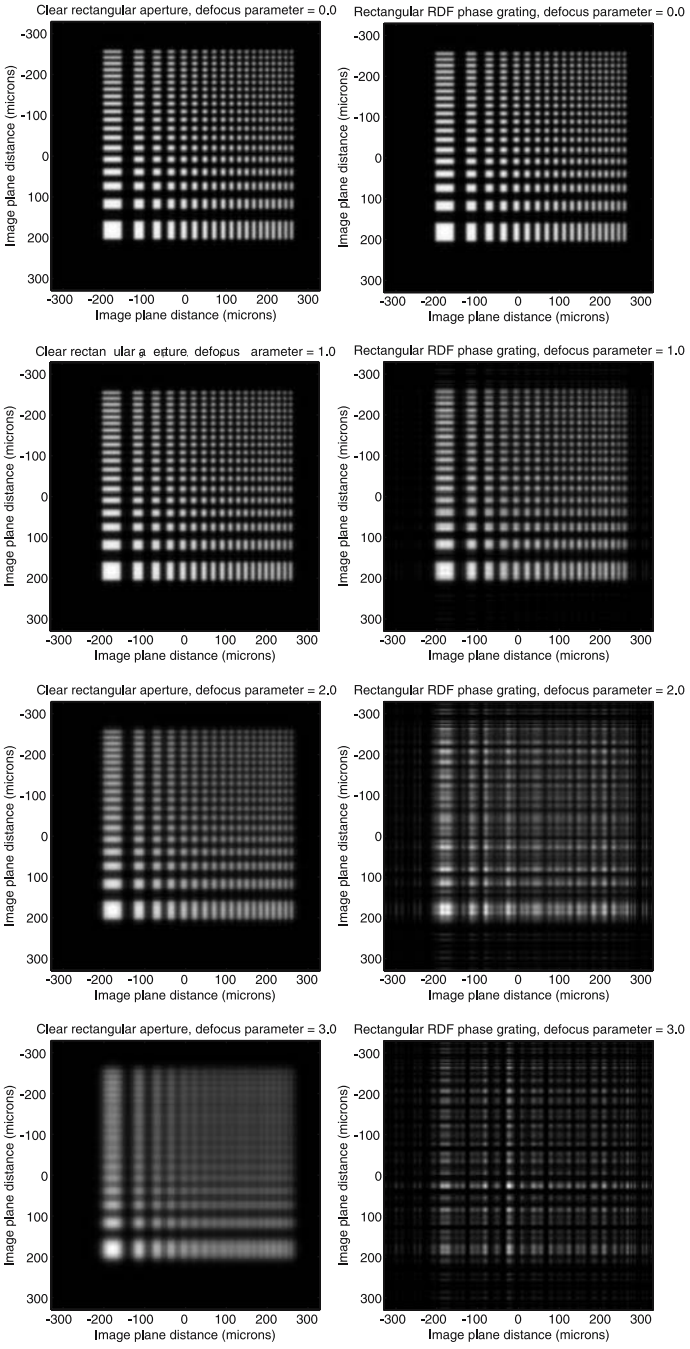


Fig. 6.19. Defocused diffraction-limited images using a clear rectangular aperture and using a rectangular RDF phase grating

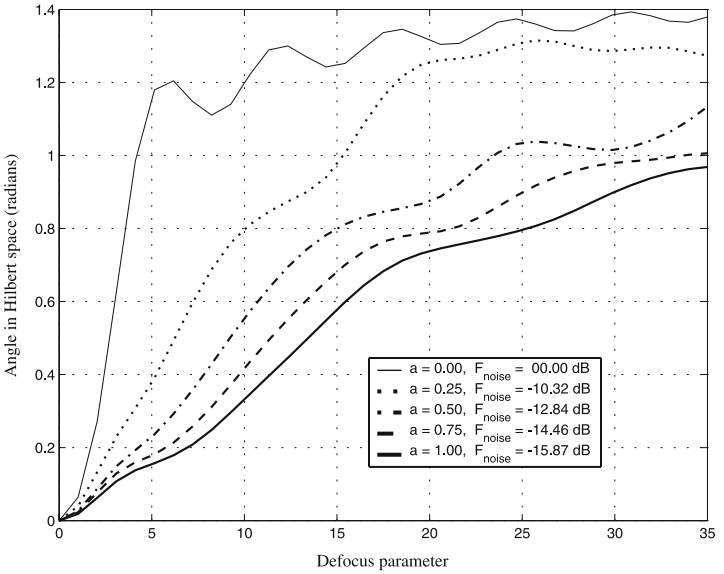


Fig. 6.20. EDF due to logarithmic phase plates of different thickness and the corresponding SNR reduction in the final image

a defocused CCD-limited imaging system can be written as [36],

$$\left[|h(u,w)|^2\right]_{\text{CCD-limited}} = \left\{|h(u,w)|^2 * \text{rect}\left(\frac{u}{a}\right)\right\} \text{comb}\left(\frac{u}{u_s}\right) \tag{6.25}$$

where a is the 1-D pixel size, $\text{comb}\left(\frac{u}{u_s}\right)$ is a train of Dirac delta functions with spacing u_s and $*$ is a 1-D convolution operator.

6.5.3 CCD Effect on Depth of Field Extension

In Fig. 6.21, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, with a logarithmic phase plate, whose parameters are shown in Table 6.3, at its exit pupil. On the same figure, Fig. 6.21, we also show the angle in Hilbert space between the in-focus PSF and defocused PSFs of the same imaging system, but with a CCD of pixel size $a = 2.0\,\mu\text{m}$.

From Fig. 6.21, we note that, for all shown defocus parameter values, the angle in Hilbert space between the in-focus PSF and defocused PSFs of a CCD-limited system with a logarithmic phase plate at its exit pupil has a smaller value than the corresponding angle in Hilbert space of a similar diffraction-limited system. Thus the PSF of a CCD-limited imaging system varies even less with defocus, compared to the PSF of a similar diffraction-limited imaging system.

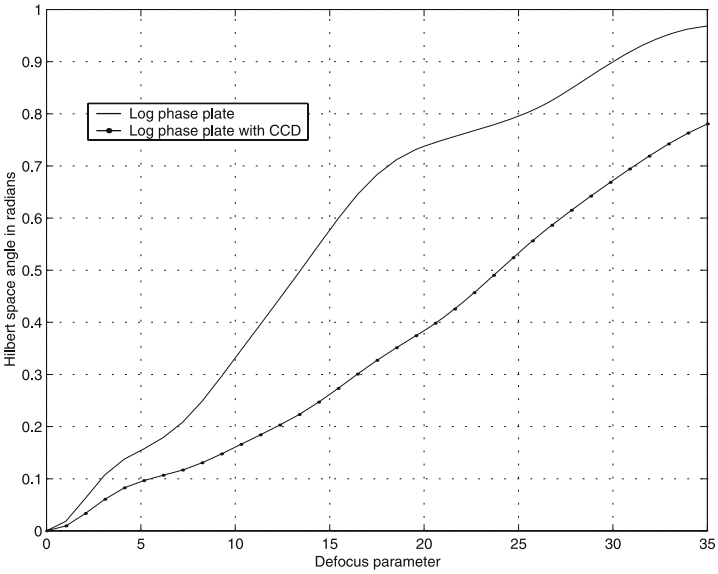


Fig. 6.21. Diffraction-limited and CCD-limited Hilbert space angles using a logarithmic phase plate

In general, the use of a CCD optical detector with a diffraction-limited system, which has a logarithmic phase plate at its exit pupil, helps in extending the depth of field.

6.5.4 CCD Effect on Depth of Field Reduction

In Fig. 6.22, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, with an RDF phase grating, whose parameters are shown in Table 6.4, at its exit pupil. On the same figure, Fig. 6.22, we also show the angle in Hilbert space between the in-focus PSF and defocused PSFs of the same imaging system, but with a CCD of pixel size $a = 1.0\,\mu\text{m}$.

From Fig. 6.22, we note that, for all shown defocus parameter values, the angle in Hilbert space between the in-focus PSF and defocused PSFs of a CCD-limited system with an RDF phase grating at its exit pupil has a smaller value than the corresponding angle in Hilbert space of a similar diffraction-limited system. Thus the PSF of a CCD-limited imaging system does not vary as much with defocus, compared to the PSF of a similar diffraction-limited imaging system.

In Fig. 6.23, we show the angle in Hilbert space between the in-focus PSF and defocused PSFs of a diffraction-limited imaging system, whose parameters are shown in Table 6.1, with an RDF phase grating, whose parameters are shown in Table 6.4, at its exit pupil. On the same figure, Fig. 6.23, we also show the angle in

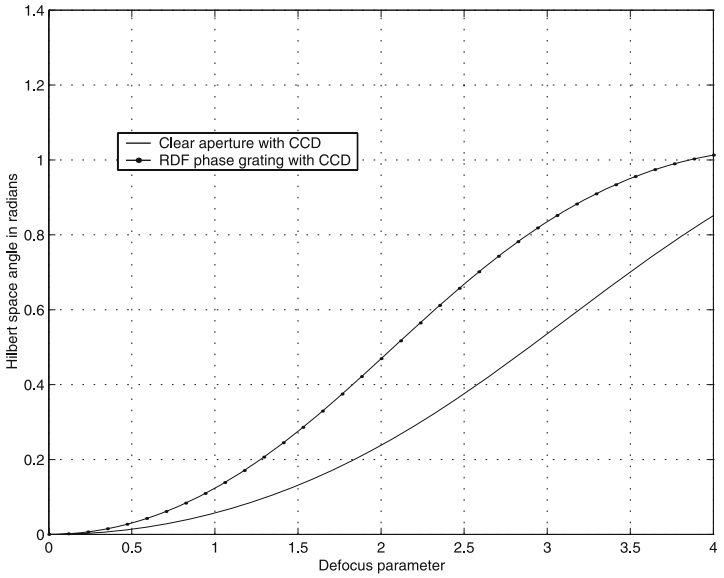


Fig. 6.22. CCD-limited Hilbert space angles using a clear rectangular aperture and a rectangular RDF phase grating

Hilbert space between the in-focus PSF and defocused PSFs of the same imaging system, but with a CCD of pixel size $a = 1.0\,\mu\text{m}$.

From Fig. 6.23, we note that, for lower defocus parameter values, the angle in Hilbert space between the in-focus PSF and defocused PSFs of a system with a rectangular RDF phase grating at its exit pupil, is larger than the corresponding angle in Hilbert space of a standard system by more than a factor of two. Thus a CCD-limited system with a rectangular RDF phase grating at its exit pupil also has less than ha lf the depth of field of a similar standard CCD-limited system.

In general, the use of a CCD optical detector with a diffraction-limited system, which has an RDF phase grating at its exit pupil, has a negative effect on the reduction of the depth of field. However, the use of an RDF phase grating to reduce the depth of field of a CCD-limited imaging system by a factor of 2 is possible.

6.6 Conclusions

We defined a new metric to quantify the blurring of a defocused image that is more suitable than the defocus parameter for hybrid imaging systems.

We described a spatial-domain method to design a pupil phase plate to extend the depth of field of an incoherent hybrid imaging system with a rectangular aperture. We used this method to obtain a pupil phase plate to extend the depth of field, which we referred to as the logarithmic phase plate. By introducing a logarithmic phase plate at the exit pupil and digitally processing the output of the detector, the depth

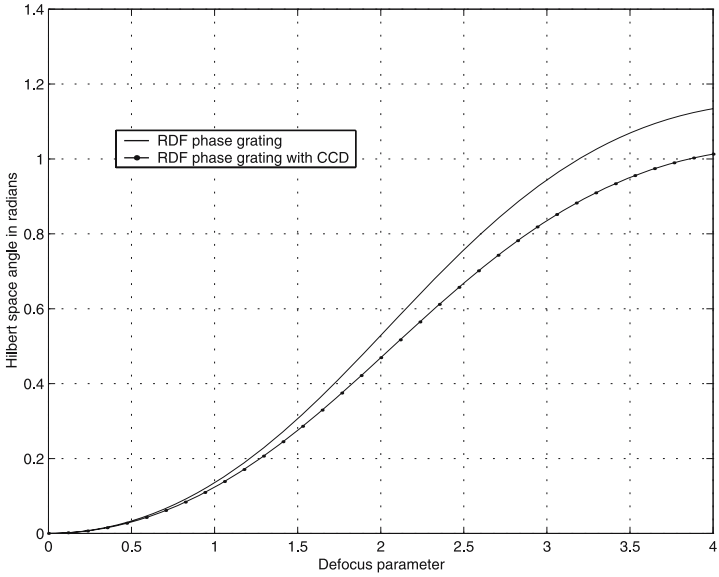


Fig. 6.23. Diffraction-limited and CCD-limited Hilbert space angles using a rectangular RDF phase grating

of field was extended by an order of magnitude more than the Hopkins defocus criterion. We compared the performance of the logarithmic phase plate with other extended-depth-of-field phase plates in extending the depth of field of incoherent hybrid imaging systems with rectangular and circular apertures.

We used our new metric for defocused image blurring to design a pupil phase plate to reduce the depth of field, thereby increasing the axial resolution, of an incoherent hybrid imaging systems with a rectangular aperture. By introducing this phase plate at the exit pupil and digitally processing the output of the detector output, the depth of field was reduced by more than a factor of two.

Finally, we examined the effect of using a noisy CCD optical detector, instead of an ideal optical detector, on the control of the depth of field. We found that the SNR of the final image is typically reduced by the digital post-processing filter. We also found that the use of a CCD with a diffraction-limited system helps in extending the depth of field and the use of a CCD has a negative effect on the reduction of the depth of field. However, the use of an RDF phase grating to reduce the depth of field of a CCD-limited imaging system by a factor of 2 is possible.

Acknowledgements

This material is based upon work supported by, or in part by, the U.S. Army Research Laboratory and the U.S. Army Research Office under contract/grant number DAAD 19-00-1-0514. S. Sherif thanks Dr. Costel Flueraru for his support through

the Optics Group at the Institute of Microstructural Sciences, National Research Council Canada.

References

1. H.H. Hopkins : Proc. Roy. Soc. A **231**, 91 (1955)
2. W.T. Cathey, B.R. Frieden, W.T. Rhodes and C.K. Rushford: J. Opt. Soc. Amer. A **1**, 241 (1984)
3. W.T. Cathey and E.R. Dowski: Appl. Opt. **41**, 6080 (2002)
4. A.K. Jain: *Fundamentals of Digital Image Processing*. (Prentice Hall, New Jersey 1989)
5. S.S. Sherif and W.T. Cathey: Appl. Opt. **41**, 6062 (2002)
6. L.E. Franks: *Signal Theory*, Revised edn. (Dowden and Culver, Stroudsburg 1981)
7. W.T. Welford: J. Opt. Soc. Amer. **50**, 794 (1960)
8. M. Mino and Y. Okano: Appl. Opt. **10**, 2219 (1971)
9. J. Ojeda-Castenada, P. Andres and A. Diaz: Opt. Letters **11**, 478 (1986)
10. J. Ojeda-Castenada, E. Tepichin and A. Diaz: Appl. Opt. **28**, 266 (1989)
11. J. Ojeda-Castenada and L.R. Berriel-Valdos: Appl. Opt. **29**, 994 (1990)
12. G. Hausler: Opt. Comm. **6**, 994 (1972)
13. J. Ojeda-Castenada, R. Ramos and A. Noyola-Isgleas: Appl. Opt. **27**, 2583 (1988)
14. E.R. Dowski and W.T. Cathey: Appl. Opt. **34**, 1859 (1995)
15. W. Chi and N. George: Opt. Letters **26**, 875 (2001)
16. M. Born and E. Wolf: *Principles of Optics*, 6th edn. (Cambridge University Press, Cambridge 1997)
17. J.W. Goodman: *Introduction to Fourier Optics*, 2nd edn. (McGraw-Hill, New York 1996)
18. E.T. Copson: *Asymptotic Expansions*. (Cambridge University Press, Cambridge 1967)
19. E.L. Key, E.N. Fowle and R.D. Haggarty: IRE Int. Conv. Rec. **4**, 146 (1961)
20. E.N. Fowle: IEEE Trans. Inf. Theory **10**, 61 (1964)
21. S.S. Sherif, W.T. Cathey and E.R. Dowski: Appl. Opt. **43**, 2709 (2004)
22. Focus Software, Inc.: *Zemax: Optical Design Program, Users Guide*, (Tucson, Arizona 2000)
23. K. Brenner, A. Lohmann and J. Ojeda-Castenada: Opt. Comm. **44**, 323 (1983)
24. S.S. Sherif, E.R. Dowski and W.T. Cathey: J. Mod. Opt. **51**, 1191 (2004)
25. S.S. Sherif, G. Muyo and A.R. Harvey: J. Mod. Opt. **52**, 1783 (2005)
26. S.S. Sherif: Depth of Field Control in Incoherent Hybrid Imaging Systems. Ph.D. Dissertation, University of Colorado, Boulder (2002)
27. J.R. Swedlow, J.W. Sedat and D.A. Agard: 'Deconvolution in Optical Microscopy'. In: *Deconvolution of Images and Spectra*, 2nd edn. ed. by P.A. Janssen (Academic Press, San Diego 1997) pp. 284–309
28. A. Erhardt, G. Zinser, D. Komitowski and J. Bille: Appl. Opt. **24**, 194 (1985)
29. C.J. Sheppard and Z.S. Hegedus: J. Opt. Soc. Amer. A **5**, 643 (1988)
30. M. Martinez-Corral, P. Andres, J. Ojeda-Castenada and G. Saavedra: Opt. Comm. **119**, 491 (1995)
31. M. Martinez-Corral, P. Andres, C. Zapata-Rodriguez and M. Kowalczyk: Opt. Comm. **165**, 267 (1999)
32. T.R.M. Sales and G.M. Morris: Opt. Comm. **156**, 227 (1998)
33. M. Neil, R. Juskaitis and T. Wilson: Opt. Lett. **22**, 1905 (1997)
34. G.S. Beveridge and R.S. Schechter: *Optimization: Theory and Practice*. (McGraw-Hill, New York 1970)
35. S.S. Sherif, E.R. Dowski and W.T. Cathey: Opt. Lett. **30**, 2566 (2005)
36. J.E. Greivenkamp and A.E. Lowman: Appl. Opt. **33**, 5029 (1994)

7 Wavefront Coding Fluorescence Microscopy Using High Aperture Lenses

Matthew R. Arnison, Carol J. Cogswell, Colin J.R. Sheppard, Peter Török

7.1 Extended Depth of Field Microscopy

In recent years live cell fluorescence microscopy has become increasingly important in biological and medical studies. This is largely due to new genetic engineering techniques which allow cell features to grow their own fluorescent markers. A popular example is green fluorescent protein. This avoids the need to stain, and thereby kill, a cell specimen before taking fluorescence images, and thus provides a major new method for observing live cell dynamics.

With this new opportunity come new challenges. Because in earlier days the process of staining killed the cells, microscopists could do little additional harm by squashing the preparation to make it flat, thereby making it easier to image with a high resolution, shallow depth of field lens. In modern live cell fluorescence imaging, the specimen may be quite thick (in optical terms). Yet a single 2D image per time-step may still be sufficient for many studies, as long as there is a large depth of field as well as high resolution.

Light is a scarce resource for live cell fluorescence microscopy. To image rapidly changing specimens the microscopist needs to capture images quickly. One of the chief constraints on imaging speed is the light intensity. Increasing the illumination will result in faster acquisition, but can affect specimen behaviour through heating, or reduce fluorescent intensity through photobleaching.

Another major constraint is the depth of field. Working at high resolution gives a very thin plane of focus, leading to the need to constantly "hunt" with the focus knob while viewing thick specimens with rapidly moving or changing features. When recording data, such situations require the time-consuming capture of multiple focal planes, thus making it nearly impossible to perform many live cell studies.

Ideally we would like to achieve the following goals:

- use all available light to acquire images quickly,
- achieve maximum lateral resolution,
- and yet have a large depth of field.

However, such goals are contradictory in a normal microscope.

For a high aperture aplanatic lens, the depth of field is [1]

$$\Delta z = 1.77\lambda \,/\, \left[4\sin^2\frac{\alpha}{2}\left(1-\frac{1}{3}\tan^4\frac{\alpha}{2}\right)\right], \tag{7.1}$$

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

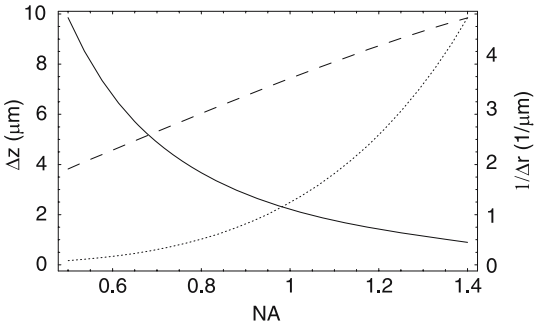


Fig. 7.1. Depth of field (*solid line*), lateral resolution (*dashed line*) and peak intensity at focus (*dotted line* – arbitrary units) for an oil immersion ($n_{\text{oil}} = 1.518$) aplanatic microscope objective with a typical range of NA and $\lambda_0 = 0.53$ μm as the vacuum wavelength

where Δz is defined as the distance along the optical axis for which the intensity is more than half the maximum. Here the focal region wavelength is λ and the aperture half-angle is α. A high aperture value for the lateral resolution can be approximated from the full-width at half-maximum (FWHM) of the unpolarised intensity point spread function (PSF) [2]. We can use the same PSF to find the peak intensity at focus, as a rough indication of the high aperture light collection efficiency,

$$I_{\text{focus}} \propto \left[1 - \frac{5}{8}(\cos^{\frac{3}{2}} \alpha)(1 + \frac{3}{5}\cos\alpha)\right]^2 . \tag{7.2}$$

These relationships are plotted in Fig. 7.1 for a range of numerical apertures (NA),

$$\text{NA} = n_1 \sin\alpha \tag{7.3}$$

where n_1 is the refractive index of the immersion medium. Clearly maximising the depth of field conflicts with the goals of high resolution and light efficiency.

7.1.1 Methods for Extending the Depth of Field

A number of methods have been proposed to work around these limitations and produce an extended depth of field (EDF) microscope.

Before the advent of charge-coupled device (CCD) cameras, Häusler [3] proposed a two step method to extend the depth of focus for incoherent microscopy. First, an axially integrated photographic image is acquired by leaving the camera shutter open while the focus is smoothly changed. The second step is to deconvolve the image with the integration system transfer function. Häusler showed that as long as the focus change is more than twice the thickness of the object, the transfer function does not change for parts of the object at different depths – effectively the transfer function is invariant with defocus. The transfer function also has no zeros, providing for easy single-step deconvolution.

This method could be performed easily with a modern microscope, as demonstrated recently by Juškaitis et al. [4]. However, the need to smoothly vary the focus

is a time-consuming task requiring some sort of optical displacement within the microscope. This is in conflict with our goal of rapid image acquisition.

A similar approach is to simply image each plane of the specimen, stepping through focus, then construct an EDF image by taking the axial average of the 3D image stack, or some other more sophisticated operation which selects the best focused pixel for each transverse specimen point. This has been described in application to confocal microscopy [5], where the optical sectioning makes the EDF post-processing straightforward. Widefield deconvolution images could also be used. In both cases the focal scanning and multiple plane image capture are major limitations on overall acquisition speed.

Potuluri et al. [6] have demonstrated the use of rotational shear interferometry with a conventional widefield transmission microscope. This technique, using incoherent light, adds significant complexity, and reduces the signal-to-noise ratio (SNR). However the authors claim an effectively infinite depth of field. The main practical limit on the depth of field is the change in magnification with depth (perspective projection) and the rapid drop in image contrast away from the imaging lens focal plane.

Another approach is to use a pupil mask to increase the depth of field, combined with digital image restoration. This creates a digital–optical microscope system. Designing with such a combination in mind allows additional capabilities not possible with a purely optical system. We can think of the pupil as encoding the optical wavefront, so that digital restoration can decode a final image, which gives us the term *wavefront coding*.

In general a pupil mask will be some complex function of amplitude and phase. The function might be smoothly varying, and therefore usable over a range of wavelengths. Or it might be discontinuous in step sizes that depend on the wavelength, such as a binary phase mask.

Many articles have explored the use of amplitude pupil masks [7–9], including for high aperture systems [10]. These can be effective at increasing the depth of field, but they do tend to reduce dramatically the light throughput of the pupil. This poses a major problem for low light fluorescence microscopy.

Wilson et al. [11] have designed a system which combines an annulus with a binary phase mask. The phase mask places most of the input beam power into the transmitting part of the annular pupil, which gives a large boost in light throughput compared to using the annulus alone. This combination gives a ten times increase in depth of field. The EDF image is laterally scanned in x and y, and then deconvolution is applied as a post-processing step.

Binary phase masks are popular in lithography where the wavelength can be fixed. However, in widefield microscopy any optical component that depends on a certain wavelength imposes serious restrictions. In epi-fluorescence, the incident and excited light both pass through the same lens. Since the incident and excited light are at different wavelengths, any wavelength dependent pupil masks would need to be imaged onto the lens pupil from beyond the beam splitter that separates

the incoming and outgoing light paths. This adds significant complexity to the optical design of a widefield microscope.

The system proposed by Wilson et al. [11] is designed for two-photon confocal microscopy. Optical complexity, monochromatic light, and scanning are issues that confocal microscopy needs to deal with anyway, so this method of PSF engineering adds relatively little overhead.

Wavefront coding is an incoherent imaging technique that relies on the use of a smoothly varying phase-only pupil mask, along with digital processing. Two specific functions that have been successful are the cubic [12,13] and logarithmic [14] phase masks, where the phase is a cubic or logarithmic function of distance from the centre of the pupil, in either radial or rectangular co-ordinates. The logarithmic design is investigated in detail in Chap. 6.

The cubic phase mask (CPM) was part of the first generation wavefront coding systems, designed for general EDF imaging. The CPM has since been investigated for use in standard (low aperture) microscopy [15]. The mask can give a ten times increase in the depth of field without loss of transverse resolution.

Converting a standard widefield microscope to a wavefront coding system is straightforward. The phase mask is simply placed in the back pupil of the microscope objective. The digital restoration is a simple single-step deconvolution, which can operate at video rates. Once a phase mask is chosen to match a lens and application, an appropriate digital inverse filter can be designed by measuring the PSF. The resulting optical–digital system is specimen independent.

The main trade off is a lowering of the SNR as compared with normal widefield imaging. The CPM also introduces an imaging artefact where specimen features away from best focus are slightly laterally shifted in the image. This is in addition to a perspective projection due to the imaging geometry, since an EDF image is obtained from a lens at a single position on the optical axis. Finally, as the CPM is a rectangular design, it strongly emphasises spatial frequencies that are aligned with the CCD pixel axes.

High aperture imaging does produce the best lateral resolution, but it also requires more complex theory to model accurately. Yet nearly all of the investigations of EDF techniques reviewed above are low aperture. In this chapter we choose a particular EDF method, wavefront coding with a cubic phase plate, and investigate its experimental and theoretical performance for high aperture microscopy.

7.2 High Aperture Fluorescence Microscopy Imaging

A wavefront coding microscope is a relatively simple modification of a modern microscope. A system overview is shown in Fig. 7.2.

The key optical element in a wavefront coding system is the waveplate. This is a transparent molded plastic disc with a precise aspheric height variation. Placing the waveplate in the back focal plane of a lens introduces a phase aberration designed to create invariance in the optical system against some chosen imaging parameter.

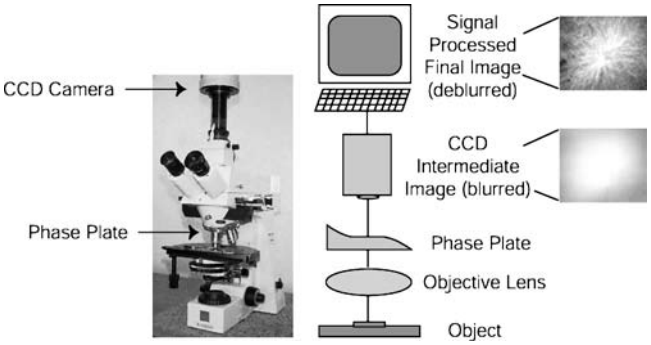


Fig. 7.2. An overview of a wavefront coding microscope system. The image-forming light from the object passes through the objective lens and phase plate and produces an intermediate encoded image on the CCD camera. This blurred image is then digitally filtered (decoded) to produce the extended depth of field result. Examples at right show the fluorescing cell image of Fig. 7.7(c) at each stage of the two-step process. At lower left an arrow shows where the phase plate is inserted into the microscope

A cubic phase function on the waveplate is useful for microscopy, as it makes the low aperture optical transfer function (OTF) insensitive to defocus.

While the optical image produced is quite blurry, it is uniformly blurred over a large range along the optical axis through the specimen (Fig. 7.3). From this blurred intermediate image, we can digitally reconstruct a sharp EDF image, using a measured PSF of the system and a single step deconvolution. The waveplate and digital filter are chosen to match a particular objective lens and imaging mode, with the digital filter further calibrated by the measured PSF. Once these steps are carried out, wavefront coding works well for any typical specimen.

The EDF behaviour relies on modifying the light collection optics only, which is why it can be used in other imaging systems such as photographic cameras, without needing precise control over the illumination light. In epi-fluorescence both the illumination light and the fluorescent light pass through the waveplate. The CPM provides a beneficial effect on the illumination side, by spreading out the axial range of stimulation in the specimen, which will improve the SNR for planes away from best focus.

7.2.1 Experimental Method

The experimental setup followed the system outline shown in Fig. 7.2. We used a Zeiss Axioplan microscope with a Plan Neofluar 40× 1.3 NA oil immersion objective. The wavefront coding plate was a rectangular cubic phase function design (CPM 127-R60 Phase Mask from CDM Optics, Boulder, CO, USA) with a peak to valley phase change of 56.6 waves at 546 nm across a 13 mm diameter optical surface. This plate was placed in a custom mount and inserted into the differential interference contrast slider slot immediately above the objective, and aligned so that it was centred with the optical axis, covering the back pupil.

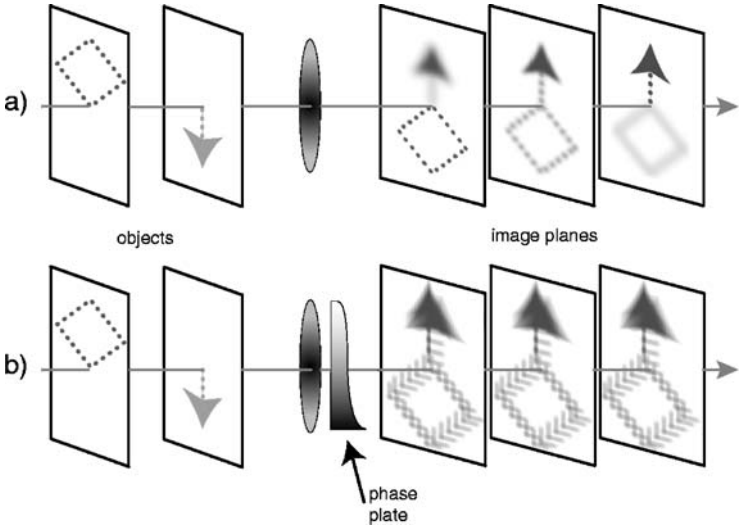


Fig. 7.3. How points are imaged in standard versus wavefront coding systems: (**a**) Conventional (small depth of field) system with two axially-separated objects to the left of a lens. Because each object obtains best focus at a different image plane, the arrow object points decrease in diameter toward their plane of best focus (far right), while the object points of the diamond are increasingly blurred. (**b**) Inserting a CPM phase plate causes points from both objects to be equivalently blurred over the same range of image planes. Signal processing can be applied to any one of these images to remove the constant blur and produce a sharply-focused EDF image

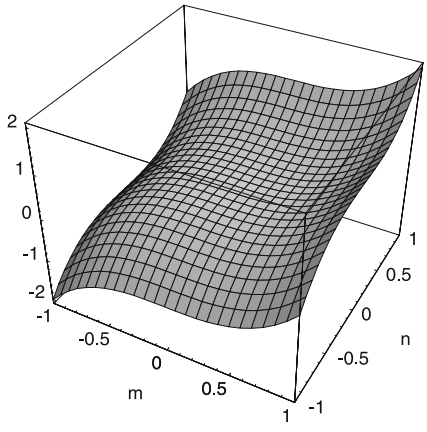


Fig. 7.4. Height variation across the cubic phase mask given in (7.4), for $A = 1$

A custom square aperture mask was inserted into an auxiliary slot 22 mm above the lens, with the square mask cut to fit inside the 10 mm circular pupil of the objective lens. This square mask is needed due to the rectangular nature of the CPM

function,

$$\varphi(m,n) = A(m^3 + n^3)\ , \tag{7.4}$$

where m and n are the Cartesian co-ordinates across the pupil and A is the strength of the phase mask (see Fig. 7.4). The square mask was rotated to match the mn axes of the CPM. For comparison, standard widefield fluorescence imaging was performed without the CPM or the square aperture mask in place.

Images were taken in epi-fluorescence mode with a mercury lamp (HBO 50 W) and fluorescein isothiocyanate (FITC) fluorescence filters in place. Images were recorded with a Photometrics cooled camera (CH250) with a Thomson TH 7895 CCD at 12-bit precision. To ensure we were sampling at the maximum resolution of the 1.3 NA lens, a 2.5× eyepiece was inserted just before the camera inside a custom camera mount tube. This tube also allowed precise rotational alignment of the camera, in order to match the CCD pixel array axes with the CPM mn axes.

With 100× total magnification and 19 μm square CCD pixels, this setup gave a resolution of 0.19 μm per pixel. This is just below the theoretical maximum resolution of 0.22 μm for a 1.3 NA lens (see Fig. 7.1), for which critical sampling would be 0.11 μm per pixel, so the results are slightly under sampled.

The PSF was measured using a 1 μm diameter polystyrene bead stained with FITC dye. The peak emission wavelength for FITC is 530 nm. Two dimensional PSF images were taken over a focal range of 10 μm in 1 μm steps. This PSF measurement was used to design an inverse filter to restore the EDF image. The OTF was obtained from the Fourier transform of the 2D PSF.

Each wavefront coding intermediate image was a single exposure on the CCD camera. A least squares filter was incorporated into the inverse filter to suppress noise beyond the spatial frequency cutoff of the optical system. A final wavefront coding image was obtained by applying the inverse filter to a single intermediate image.

7.2.2 PSF and OTF Results

The measured PSFs and derived OTFs for the focused and 4 μm defocused cases are shown in Fig. 7.5, comparing standard widefield microscopy with wavefront coding using a CPM.

The widefield PSF shows dramatic change with defocus as expected for a high aperture image of a 1 μm bead. But the wavefront coding PSF shows very little change after being defocused by the same amount.

The OTF measurements emphasise this focus independence for the wavefront coding system. While the in-focus OTF for the widefield system has the best overall response, the OTF quickly drops after defocusing. The widefield defocused OTF also has many nulls before the spatial frequency cutoff, indicated in these results by a downward spike. These nulls make it impossible in widefield to use the most straightforward method of deconvolution – division of the image by the system OTF in Fourier space. Time consuming iterative solutions must be used instead.

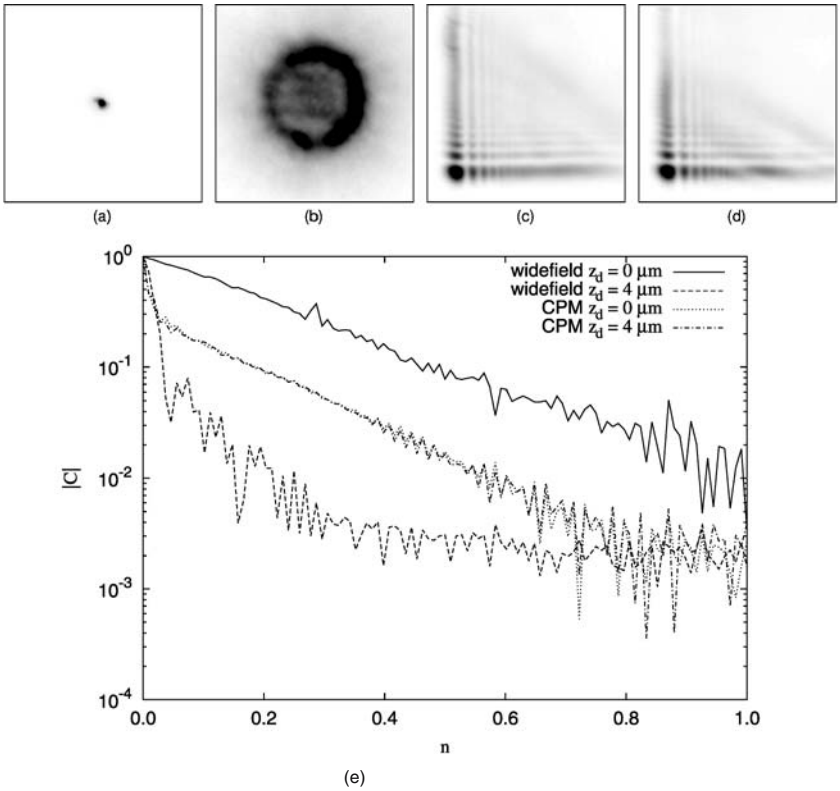


Fig. 7.5. Experimental PSFs and OTFs for the widefield and wavefront coding systems as measured using a 1 μm fluorescent bead and a NA = 1.3 oil objective. For each type of microscope, a PSF from the plane of best focus is followed by one with 4 μm defocus. The upper images (**a**-**d**) show the intensity of a central region of the PSF whilst the lower graph (**e**) gives the magnitude of the OTF for a line $m = 0$ through the OTF for each case: (**a**) widefield $z_d = 0$ μm (*solid line*), (**b**) widefield defocused $z_d = 4$ μm (*dashed line*), (**c**) CPM $z_d = 0$ μm (*dotted line*), (**d**) CPM defocused $z_d = 4$ μm (*dash-dotted line*). The spatial frequency n has been been normalised so that $n = 1$ lies at the CCD camera spatial frequency cutoff. The PSFs have area 13 μm × 13 μm

The wavefront coding system OTF shows a reduced SNR compared with the in-focus widefield OTF. Yet the same SNR is maintained through a wide change in focus, indicating a depth of field at least 8 times higher than the widefield system. The CPM frequency response extends to 80% of the spatial frequency cutoff of the widefield case before descending into the noise floor. This indicates that the wavefront coding system has maintained much of the transverse resolution expected from the high aperture lens used. Because there are no nulls in the CPM OTF at spatial frequencies below the SNR imposed cutoff, deconvolution can be performed using a single-pass inverse filter based on the reciprocal of the system OTF.

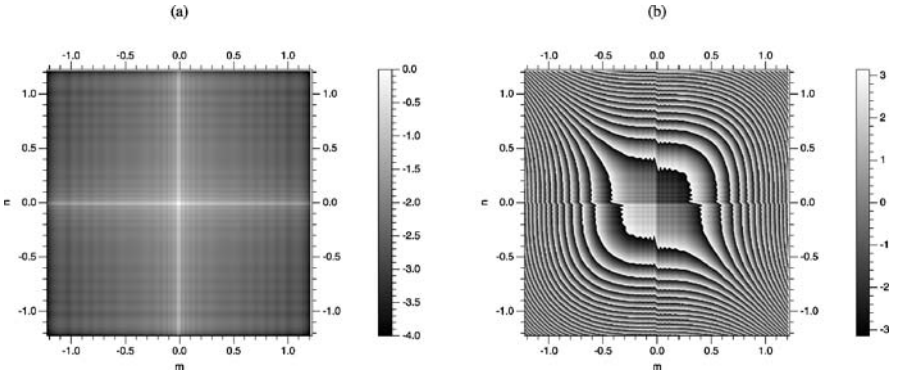


Fig. 7.6. The measured CPM in-focus 2D OTF: (**a**) is the magnitude of the OTF in $\log_{10}$ scale, and (**b**) is the wrapped OTF phase in radians. The spatial frequencies m and n have been been normalised so that $|m|, |n| = 1$ lies at the CCD camera spatial frequency cutoff

A limiting factor on the SNR, and therefore the wavefront coding system resolution, is the CCD camera dynamic range of 12 bits, giving a noise floor of at least 2.4×10^{-4}. From Fig. 7.5(e) the effective noise floor seems to be a bit higher at 10^{-3}. This has a greater impact on the off-axis spatial frequencies, where a higher SNR is required to maintain high spatial frequency response, an effect which is clearly seen in the measured 2D OTF in Fig. 7.6.

7.2.3 Biological Imaging Results

In order to experimentally test high resolution biological imaging using the CPM wavefront coding system in epi-fluorescence, we imaged an anti-tubulin / FITC-labeled HeLa cell. For comparison, we also imaged the same mitotic nucleus in both a standard widefield fluorescence microscope and a confocal laser scanning system (Fig. 7.7). The first widefield image, Fig. 7.7(a), shows a mitotic nucleus with one centriole in sharp focus, while a second centriole higher in the specimen is blurred. This feature became sharp when the focus was altered by 6 μm, as shown in Fig. 7.7(b). The wavefront coding system image in Fig. 7.7(c) shows a much greater depth of field, with both centrioles in focus in the same image. We observed a depth of field increase of at least 6 times compared with the widefield system, giving a 6 μm depth of field for the wavefront coding system for the NA = 1.3 oil objective.

For further comparison, we imaged the same specimen using a confocal microscope. A simulated EDF image is shown in Fig. 7.7(d), obtained by averaging 24 planes of focus. This gives an image of similar quality to the wavefront coding image. However, the confocal system took over 20 times longer to acquire the data for this image, due to the need to scan the image point in all three dimensions. There is also a change in projection geometry between the two systems. The confocal EDF image has orthogonal projection, whereas the wavefront coding EDF image has perspective projection.

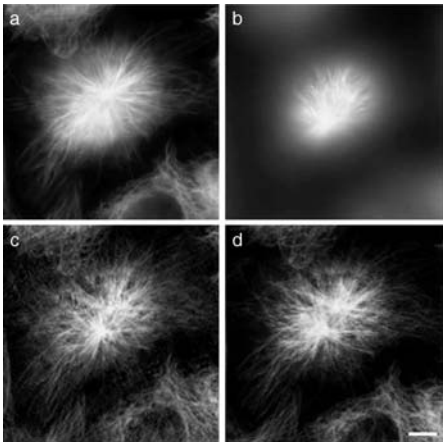


Fig. 7.7. Comparison images of an antitubulin / FITC-labeled HeLa cell nucleus obtained using three kinds of microscope. (**a**-**b**) Conventional widefield fluorescence images of the same mitotic nucleus acquired at two different focal planes, 6 μm apart in depth. Misfocus blurring is prevalent, with only one of the two centrioles in focus in each image. (**c**) A CPM wavefront coding image of this nucleus greatly increases focal depth so that now both centrioles in the mitotic spindle are sharply focused. (**d**) An equivalent confocal fluorescence EDF image obtained by averaging 24 separate planes of focus, spaced 0.5 μm apart. The resolutions of the wavefront coding and confocal images are comparable but the confocal image took over 20 times longer to produce. Note that wavefront coding gives a perspective projection and confocal gives an isometric projection, which chiefly accounts for their slight difference in appearance. Objective NA=1.3 oil, scale bar: 6 μm

7.3 Wavefront Coding Theory

In this section we will investigate theoretical models for wavefront coding microscopy. We present a summary of the development of the cubic phase function and the paraxial theory initially used to model it. We then analyse the system using vectorial high aperture theory, as is normally required for accuracy with a 1.3 NA lens.

High aperture vectorial models of the PSF for a fluorescence microscope are well developed [16,17]. The Fourier space equivalent, the OTF, also has a long history [18–20]. However, the CPM defined in (7.4) is an unusual microscope element:

1. Microscope optics usually have radial symmetry around the optical axis, which the CPM does not.
2. The CPM gives a very large phase aberration of up to 60 waves, whilst most aberration models are oriented towards phase strengths on the order of a wave at most.
3. In addition, the CPM spreads the light over a very long focal range, whilst most PSF calculations can assume the energy drops off very rapidly away from focus.

These peculiarities have meant we needed to take particular care with numerical computation in order to ensure accuracy, and in the case of the OTF modeling the radial asymmetry has motivated a reformulation of previous symmetric OTF theory.

7.3.1 Derivation of the Cubic Phase Function

There are various methods that may be used to derive a pupil phase function which has the desired characteristics for EDF imaging. The general form of a phase function in Cartesian co-ordinates is

$$T(m,n) = \exp\left[\mathrm{i}k\varphi(m,n)\right] \ , \tag{7.5}$$

where m, n are the lateral pupil co-ordinates and $k = 2\pi/\lambda$ is the wave-number. The cubic phase function was found by Dowski and Cathey [13] using paraxial optics theory by assuming the desired phase function is a simple 1D function of the form

$$\varphi(m) = Am^{\gamma}, \ \gamma \neq \{0,1\}, \ A \neq 0 \ . \tag{7.6}$$

By searching for the values of A and γ which give an OTF which does not change through focus, they found, using the stationary phase approximation and the ambiguity function, that the best solution was for $A \gg 20/k$ and $\gamma = 3$. Multiplying out to 2D, this gives the cubic phase function in (7.4).

7.3.2 Paraxial Model

Using the Fraunhofer approximation, as suitable for low NA, we can write down a 1D pupil transmission function encompassing the effects of cubic phase (7.4) and defocus,

$$T(m) = \exp\left[\mathrm{i}k\varphi(m)\right]\exp\left(\mathrm{i}m^2\psi\right) \ , \tag{7.7}$$

where ψ is a defocus parameter. We then find the 1D PSF is

$$E(x) = \int_{-1}^{1} T(m)\exp\left(\mathrm{i}xm\right)\,\mathrm{d}m \ , \tag{7.8}$$

where x is the lateral co-ordinate in the PSF . The 1D OTF is

$$C(m) = \int_{-1}^{1} T\left(m' + m/2\right)\,T^{*}(m' - m/2)\,\mathrm{d}m' \ . \tag{7.9}$$

The 2D PSF is simply $E(x)E(y)$.

Naturally this 1D CPM gives behaviour in which, in low aperture systems at least, the lateral x and y imaging axes are independent of each other. This gives significant speed boosts in digital post-processing. Another important property of the CPM is that the OTF does not reach zero below the spatial frequency cutoff which means that deconvolution can be carried out in a single step. The lengthy

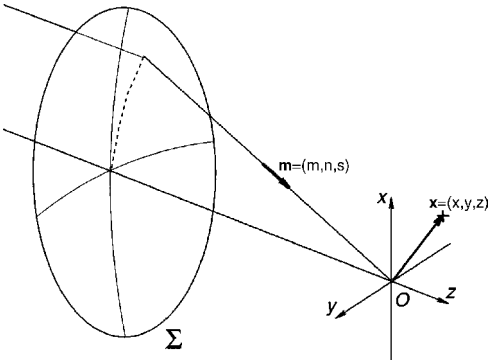


Fig. 7.8. Diagram of the light focusing geometry used in calculating the high NA PSF, indicating the focal region co-ordinate $\boldsymbol{x}$ and the pupil co-ordinate $\boldsymbol{m}$, the latter of which may also be thought of as a unit vector aligned with a ray from the pupil to the focal point O

iterative processing of widefield deconvolution is largely due to the many zeros in the conventional defocused OTF. Another important feature of Fraunhofer optics is that PSF changes with defocus are limited to scaling changes. Structural changes in the PSF pattern are not possible.

This paraxial model for the cubic phase mask has been thoroughly verified experimentally for low NA systems [12,15].

7.3.3 High Aperture PSF Model

We now explore the theoretical behaviour for a high NA cubic phase system. Normally we need high aperture theory for accurate modeling of lenses with NA > 0.5. However large aberrations like our cubic phase mask can sometimes overwhelm the high NA aspects of focusing. By comparing the paraxial and high NA model results we can determine the accuracy of the paraxial approximation for particular wavefront coding systems.

The theory of Richards and Wolf [2] describes how to determine the electric field near to the focus of a lens which is illuminated by a plane polarised quasi-monochromatic light wave. Their analysis assumes very large values of the Fresnel number, equivalent to the Debye approximation. We can then write the equation for the vectorial amplitude PSF $\boldsymbol{E}(\boldsymbol{x})$ of a high NA lens illuminated with a plane polarised wave as the Fourier transform of the complex vectorial pupil function $\boldsymbol{Q}(\boldsymbol{m})$ [19],

$$\boldsymbol{E}(\boldsymbol{x}) = -\frac{ik}{2\pi} \iiint \boldsymbol{Q}(\boldsymbol{m}) \exp(\mathrm{i}k\boldsymbol{m} \cdot \boldsymbol{x}) \, \mathrm{d}\boldsymbol{m} \, . \tag{7.10}$$

Here $\boldsymbol{m} = (m, n, s)$ is the Cartesian pupil co-ordinate, and $\boldsymbol{x} = (x, y, z)$ is the focal region co-ordinate. The z axis is aligned with the optical axis, and s is the corresponding pupil co-ordinate, as shown in Fig. 7.8. The vectorial pupil function $\boldsymbol{Q}(\boldsymbol{m})$ describes the effect of a lens on the polarisation of the incident field, the complex value of any amplitude or phase filters across the aperture, and any additional aberration in the lens focusing behaviour from that which produces a perfect spherical wavefront converging on the focal point.

From the Helmholtz equation for a homogeneous medium, assuming constant refractive index in the focal region, we know that the pupil function is only non-zero on the surface of a sphere with radius k,

$$\boldsymbol{Q}(\boldsymbol{m}) = \boldsymbol{P}(\boldsymbol{m})\,\delta\left(|\boldsymbol{m}| - k^2\right) . \tag{7.11}$$

Because the pupil function only exists on the surface of a sphere, we can slice it along the $s = 0$ plane into a pair of functions

$$\boldsymbol{Q}(\boldsymbol{m}) = \boldsymbol{Q}(\boldsymbol{m})\,\frac{k}{s}\,\delta\left(s - \sqrt{k^2 - l^2}\right) + \boldsymbol{Q}(\boldsymbol{m})\,\frac{k}{s}\,\delta\left(s + \sqrt{k^2 - l^2}\right) , \tag{7.12}$$

representing forward and backward propagation [21,22]. Here we have introduced a radial co-ordinate $l = \sqrt{m^2 + n^2}$. Now we take the axial projection $\boldsymbol{P}_+(m, n)$ of the forward propagating component of the pupil function,

$$\boldsymbol{P}_+(m, n) = \int_0^\infty \boldsymbol{Q}(\boldsymbol{m})\,\frac{k}{s}\,\delta\left(s - \sqrt{k^2 - l^2}\right) \mathrm{d}s \tag{7.13}$$

$$= \boldsymbol{Q}\,(m, n, s_+)\;\frac{1}{s_+} , \tag{7.14}$$

where we have normalised the radius to $k = 1$ and indicated the constraint on s to the surface of the sphere with

$$s_+ = \sqrt{1 - l^2} . \tag{7.15}$$

For incident light which is plane-polarised along the x axis, we can derive a vectorial strength function $\boldsymbol{a}(m, n)$, from the strength factors used in the vectorial point spread function integrals [2,22,23]

$$\boldsymbol{a}(m, n) = \begin{pmatrix} (m^2 s_+ + n^2)/l^2 \\ -mn(1 - s_+)/l^2 \\ -m \end{pmatrix} \tag{7.16}$$

where we have converted from the spherical polar representation in Richards and Wolf to Cartesian co-ordinates.

We can now model polarisation, apodisation and aperture filtering as amplitude and phase functions over the projected pupil,

$$\boldsymbol{P}_+(m, n) = \frac{1}{s_+}\,\boldsymbol{a}(m, n) S\,(m, n) T(m, n) \tag{7.17}$$

representing forward propagation only ($\alpha \leq \pi/2$), where $S\,(m, n)$ is the apodisation function, and $T(m, n)$ is any complex transmission filter applied across the aperture of the lens. T can also be used to model aberrations.

Microscope objectives are usually designed to obey the sine condition, giving aplanatic imaging [24], for which we write the apodisation as

$$S\,(m, n) = \sqrt{s_+} . \tag{7.18}$$

By applying low angle and scalar approximations, we can derive from (7.17) a paraxial pupil function,

$$P_+(m,n) \cong T(m,n)\ . \tag{7.19}$$

Returning to the PSF, we have

$$\boldsymbol{E}(\boldsymbol{x}) = -\frac{\mathrm{i}k}{2\pi}\iint_{\Sigma} \boldsymbol{P}_+(m,n)\exp\left(\mathrm{i}k\boldsymbol{m}_+\cdot\boldsymbol{x}\right)\,\mathrm{d}m\,\mathrm{d}n\ , \tag{7.20}$$

integrated over the projected pupil area Σ. The geometry is shown in Fig. 7.8. We use $\boldsymbol{m}_+ = (m,n,s_+)$ to indicate that $\boldsymbol{m}$ is constrained to the pupil sphere surface.

For a clear circular pupil of aperture half-angle α, the integration area Σ_{circ} is defined by

$$0 \le l \le \sin\alpha\ , \tag{7.21}$$

while for a square pupil which fits inside that circle, the limits on Σ_{sq} are

$$\begin{aligned} |m| &\le \sin\alpha/\sqrt{2} \\ |n| &\le \sin\alpha/\sqrt{2} \end{aligned}\ . \tag{7.22}$$

The transmission function T is unity for a standard widefield system with no aberrations, while for a cubic phase system (7.4) and (7.5) give

$$T_{\mathrm{c}}(m,n) = \exp\left[\mathrm{i}kA\left(m^3+n^3\right)\right]\ . \tag{7.23}$$

7.3.4 High Aperture OTF Model

A high aperture analysis of the OTF is important, because the OTF has proven to be more useful than the PSF for design and analysis of low aperture wavefront coding systems. For full investigation of the spatial frequency response of a high aperture microscope, we would normally look to the 3D OTF [18–20,25]. We have recently published a method for calculating the 3D OTF suitable for arbitrary pupil filters [21] which can be applied directly to find the OTF for a cubic phase plate. But since an EDF system involves recording a single image at one focal depth, a frequency analysis of the 2D PSF at that focal plane is more appropriate. This can be performed efficiently using a high NA vectorial adaptation of 2D Fourier optics [22].

This adaptation relies on the Fourier projection–slice theorem [26], which states that a slice through real space is equivalent to a projection in Fourier space:

$$f(x,y,0) \Longleftrightarrow \int F(m,n,s)\,\mathrm{d}s \tag{7.24}$$

where $F(m,n,s)$ is the Fourier transform of $f(x,y,z)$. We have already obtained the projected pupil function $\boldsymbol{P}_+(m,n)$ in (7.17). Taking the 2D Fourier transform and applying (7.24) gives the PSF in the focal plane

$$\boldsymbol{E}(x,y,0) \Longleftrightarrow \boldsymbol{P}_+(m,n)\ . \tag{7.25}$$

Since fluorescence microscopy is incoherent, we then take the intensity and 2D Fourier transform once more to obtain the OTF of that slice of the PSF

$$C(m,n) \Longleftrightarrow |\boldsymbol{E}\,(x,y,0)|^2\ . \tag{7.26}$$

We can implement this approach using 2D fast Fourier transforms to quickly calculate the high aperture vectorial OTF for the focal plane.

7.3.5 Defocused OTF and PSF

To investigate the EDF performance, we need to calculate the defocused OTF. Defocus is an axial shift z_d of the point source being imaged relative to the focal point. By the Fourier shift theorem, a translation z_d of the PSF is equivalent to a linear phase shift in the 3D pupil function,

$$\boldsymbol{E}(x,y,0+z_\mathrm{d}) \Longleftrightarrow \exp(\mathrm{i}ksz_\mathrm{d})\,\boldsymbol{Q}(m,n,s)\ . \tag{7.27}$$

Applying the projection-slice theorem as before gives a modified version of (7.25)

$$\boldsymbol{E}(x,y,z_\mathrm{d}) \Longleftrightarrow \int \exp(\mathrm{i}ksz_\mathrm{d})\,\boldsymbol{Q}(m,n,s)\,\mathrm{d}s\ . \tag{7.28}$$

allowing us to isolate a pupil transmission function that corresponds to a given defocus z_d,

$$T_\mathrm{d}(m,n,z_\mathrm{d}) = \exp(\mathrm{i}ks_+z_\mathrm{d})\ , \tag{7.29}$$

which we incorporate into the projected pupil function $\boldsymbol{P}_+(m,n)$ from (7.17), giving

$$\boldsymbol{P}_+(m,n,z_\mathrm{d}) = \frac{1}{s_+}\,\boldsymbol{a}(m,n)\,S\,(m,n)\,T_\mathrm{d}(m,n,z_\mathrm{d})\,T_\mathrm{c}(m,n)\ . \tag{7.30}$$

If we assume a low aperture pupil, we can approximate (7.15) to second order, giving the well known paraxial aberration function for defocus

$$T_\mathrm{d}(m,n,z_\mathrm{d}) \cong \exp\left(-\mathrm{i}kz_\mathrm{d}\frac{l^2}{2}\right)\ . \tag{7.31}$$

Finally, using $\mathcal{F}$ to denote a Fourier transform, we write down the full algorithm for calculating the OTF of a transverse slice through the PSF:

$$C(m,n,z_\mathrm{d}) = \mathcal{F}^{-1}\left\{|\mathcal{F}\,[\boldsymbol{P}_+(m,n,z_\mathrm{d})]|^2\right\}\ . \tag{7.32}$$

Table 7.1. Optical parameters used for PSF and OTF simulations

Optical parameter	Simulation value
Wavelength	530 nm
Numerical aperture	NA = 1.3 oil
Oil refractive index	$n_1 = 1.518$
Aperture half angle	$\alpha = \pi/3$
Pupil shape	Square
Pupil width	7.1 mm
Cubic phase strength	25.8 waves peak to valley

It is convenient to calculate the defocused PSF using the first step of the same approach:

$$\boldsymbol{E}(x, y, z_\mathrm{d}) = \mathcal{F}\left[\boldsymbol{P}_+(m, n, z_\mathrm{d})\right] \ . \tag{7.33}$$

7.3.6 Simulation Results

We have applied this theoretical model to simulate the wavefront coding experiments described earlier, using the parameters given in Table 7.1. The theoretical assumption that the incident light is plane polarised corresponds to the placement of an analyser in the microscope beam path. This polarisation explains some xy asymmetry in the simulation results.

Due to the large phase variation across the pupil, together with the large defocus distances under investigation, a large number of samples of the cubic phase function were required to ensure accuracy and prevent aliasing. We created a 2D array with 1024^2 samples of the pupil function $\boldsymbol{P}+$ from (7.30) using (7.22) for the aperture cutoff. We then padded this array out to 4096^2 to allow for sufficient sampling of the resulting PSF, before employing the algorithms in (7.33) and (7.32) to calculate the PSF and OTF respectively. Using fast Fourier transforms, each execution of (7.32) took about 8 minutes on a Linux Athlon 1.4 GHz computer with 1 GB of RAM.

The wavefront coding inverse filter for our experiments was derived from the theoretical widefield (no CPM) OTF and the measured CPM OTF. The discrepancy in the focal plane theoretical widefield OTF between the paraxial approximation and our vectorial high aperture calculation is shown in Fig. 7.9(a). We show a similar comparison of the defocused widefield OTF in Fig. 7.9(b). We can see there is a major difference in the predictions of the two models, especially at high frequencies. The discrepancy between the models increases markedly with defocus. This implies that the best deconvolution accuracy will be obtained by using the vectorial widefield OTF when constructing the digital inverse filter for a high aperture system.

We now investigate the simulated behaviour of a CPM system according to our vectorial theory. Figures 7.10 and 7.11 show the vectorial high aperture PSF and OTF for the focal plane with a strong CPM. The defocused $z_\mathrm{d} = 4$ μm vectorial CPM OTF (not shown) and the paraxial in-focus and defocused $z_\mathrm{d} = 4$ μm CPM PSFs and OTFs (not shown) are all qualitatively similar to the vectorial CPM results shown in Figs. 7.10 and 7.11.

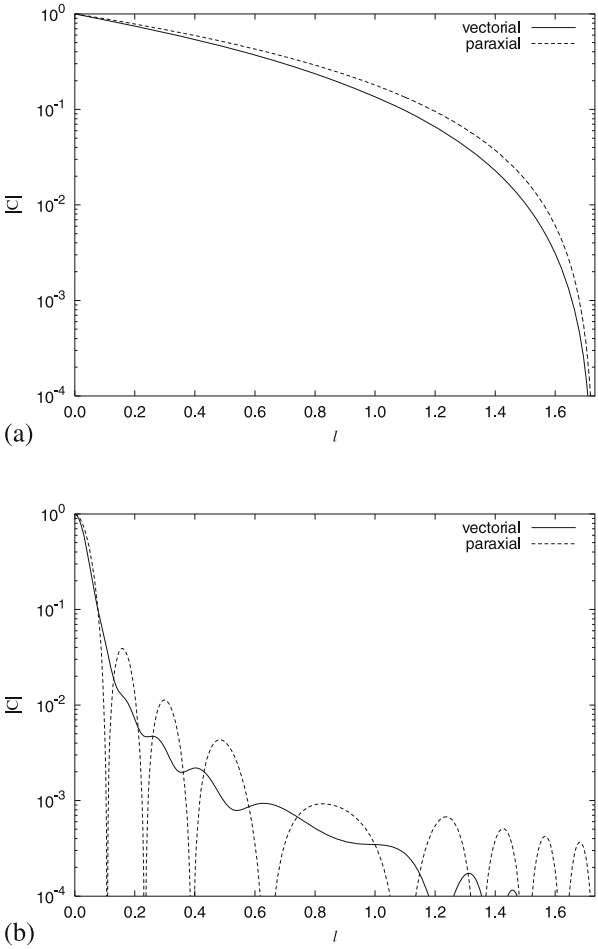


Fig. 7.9. A comparison of widefield (no CPM) OTFs using our vectorial (*solid line*) and paraxial (*dashed line*) simulations: (**a**) in-focus at $z_d = 0$ μm and (**b**) defocused to $z_d = 4$ μm. For a diagonal line through the OTF along $m = n$, we have plotted the value of the 2D projected OTF for each case. While the structure of the in-focus OTF curves is similar for the two models, the relative difference between them increases with spatial frequency, reaching over 130% at the cutoff. Once defocus is applied, the two models predict markedly different frequency response in both structure and amplitude

However, if we perform a quantitative comparison we see that there are marked differences. Figure 7.12 shows the relative strength of the CPM OTF for a diagonal cross section. The differences between the models are similar to the widefield OTF in Fig. 7.9(a) for the in-focus case, with up to 100% difference at high spatial frequencies. However, as the defocus increases, the structure of the vectorial CPM OTF

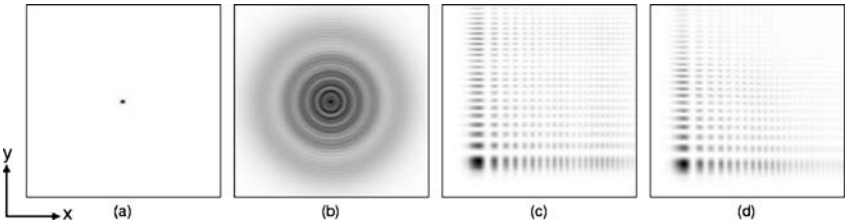


Fig. 7.10. The simulated vectorial high aperture PSF for widefield and wavefront coding, showing the effect of defocus: (**a**) widefield in-focus $z_d = 0$ μm, (**b**) widefield defocused $z_d = 4$ μm, (**c**) CPM in-focus $z_d = 0$ μm, (**d**) CPM defocused $z_d = 4$ μm. This amount of defocus introduces very little discernible difference between the CPM PSFs. Indeed paraxial CPM simulations (not shown here) are also similar in structure. The PSFs shown have the same area as Fig. 7.5 (13 μm × 13 μm). The incident polarisation is in the x direction. The images are normalised to the peak intensity of each case. Naturally the peak intensity decreases with defocus, but much less rapidly in the CPM system

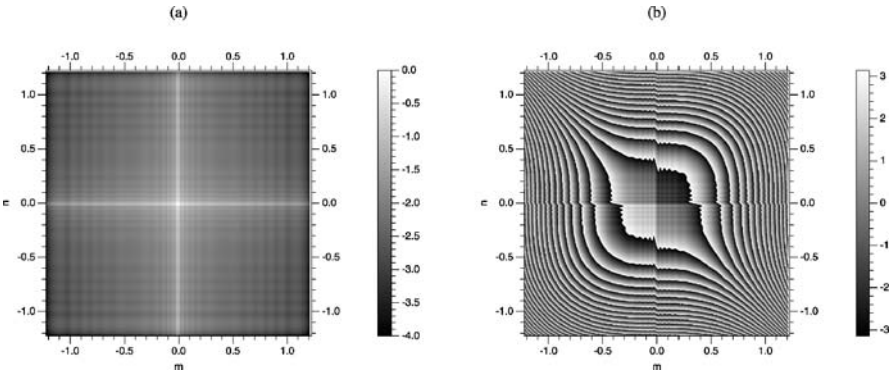


Fig. 7.11. The simulated vectorial high aperture in-focus CPM OTF: (**a**) is the magnitude of the OTF in $\log_{10}$ scale, and (**b**) is the wrapped phase in radians. While the frequency response is much stronger along the m and n axes, the magnitude remains above 10^{-3} throughout the spatial frequency cutoff. The phase of the OTF is very similar to the cubic phase in the pupil. Compensating for the OTF phase is important in digital restoration. The $z_d = 4$ μm defocused OTF (not shown) has a similar appearance to this in-focus case. See Fig. 7.6 to compare with the measured OTFs

begins to diverge from the paraxial model, as well as the point where the strength drops below 10^{-4}. This is still a much lower discrepancy than the widefield model for similar amounts of defocus, as is clear by comparison with Fig. 7.9.

These plots allow us to assess the SNR requirements for recording images with maximum spatial frequency response. For both widefield and CPM systems, the experimental dynamic range will place an upper limit on the spatial frequency response. In widefield a 10^3 SNR of will capture nearly all spatial frequencies up to the cutoff (see Fig. 7.9(a)), allowing for good contrast throughout. Further increases

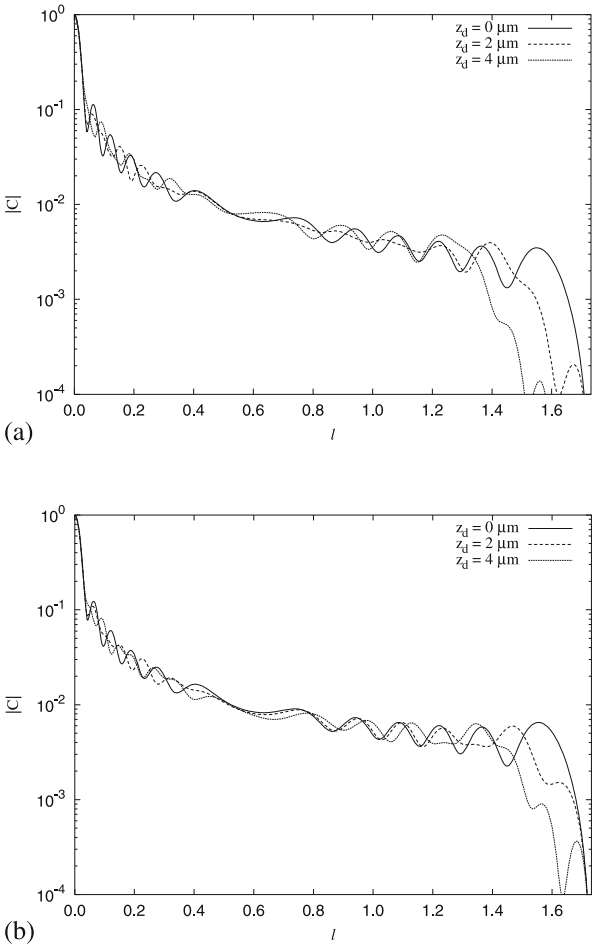


Fig. 7.12. The magnitude of the wavefront coding OTF for the (**a**) vectorial and (**b**) paraxial models, plotted along a diagonal line $m = n$ through the OTF, with different values of defocus: in-focus $z_d = 0$ μm (*solid line*), defocused $z_d = 2$ μm (*dashed line*), defocused $z_d = 4$ μm (*dotted line*). In common with the the widefield system, the models differ the most at high spatial frequencies, up to 300% for the in-focus case. As defocus increases, the differences become more extreme, with the vectorial simulation predicting a quicker reduction in effective cutoff

in SNR will bring rapidly diminishing returns, only gradually increasing the maximum spatial frequency response.

For CPM imaging the same 10^3 SNR will produce good contrast only for low spatial frequencies, with the middle frequencies lying less than a factor of ten above the noise floor, and the upper frequencies dipping below it. However, a SNR of 10^4 will allow a more reasonable contrast level across the entire OTF. For this reason,

a 16-bit camera, together with other noise control measures, is needed for a CPM system to achieve the full resolution potential of high aperture lenses. This need for high dynamic range creates a trade off for rapid imaging of living specimens – faster exposure times will reduce the SNR and lower the resolution.

Arguably the most important OTF characteristic used in the EDF digital deconvolution is the phase. As can be seen from Fig. 7.11 the CPM OTF phase oscillates heavily due to the strong cubic phase. This corresponds to the numerous contrast reversals in the PSF. The restoration filter is derived from the OTF, and therefore accurate phase in the OTF is needed to ensure that any contrast reversals are correctly restored.

A comparison of the amount of OTF phase difference between focal planes for the vectorial and paraxial models is shown in Fig. 7.13. We calculated this using the unwrapped phase, obtained by taking samples of the OTF phase along a line $m = n$, then applying a 1D phase unwrapping algorithm to those samples. After finding the unwrapped phases for different focal planes, $z_d = 2$ μm and $z_d = 4$ μm, we then subtracted them from the in focus case at $z_d = 0$ μm.

Ideally the OTF phase difference between planes within the EDF range should be very small. It is clear however that there are some notable changes with defocus. Both paraxial and vectorial models show a linear phase ramp, with oscillations.

This linear phase ramp is predicted by the stationary phase approximation to the 1D CPM OTF, Eq. (A12) in Dowski and Cathey [13]. Since the Fourier transform of a phase ramp is a lateral displacement, this gives a lateral motion of the PSF for different focal planes. In practice this has the effect of giving a slightly warped projection. A mismatch between the microscope OTF and the inverse filter of this sort will simply result in a corresponding offset of image features from that focal plane. Otherwise spatial frequencies should be recovered normally.

The oscillations will have a small effect; they are rapid and not overly large in amplitude: peaking at $\pi/2$ for both vectorial and paraxial models. This will effectively introduce a source of noise between the object and the final recovered image. Whilst these oscillations are not predicted by the stationary phase approximation, they are still evident for the paraxial model.

The most dramatic difference between the two models is in the curvature of the vectorial case, which is particularly striking in the $z_d = 4$ μm plane, and not discernible at all in the paraxial case (Fig.7.13). The primary effect of this curvature will be to introduce some additional blurring of specimen features in the $z_d = 4$ μm plane, which the inverse filter will not be able to correct. The total strength of this curvature at $z_d = 4$ μm is about 2π across the complete $m = n$ line, or one wave, which is a significant aberration.

7.3.7 Discussion

The CPM acts as a strong aberration which appears to dominate both the effects of defocus and of vectorial high aperture focusing. The paraxial approximation certainly loses accuracy for larger values of defocus, but not nearly so much as in the defocused widefield case. Yet significant differences remain between the two

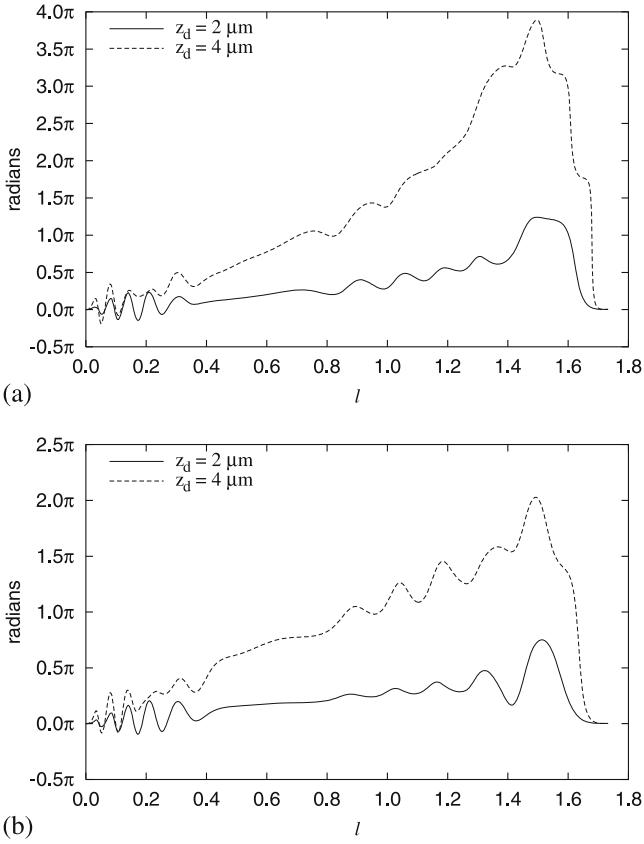


Fig. 7.13. The relative OTF phase angle between focal planes, along a diagonal line $m = n$ through the CPM OTF, for (**a**) the vectorial model, and (**b**) the paraxial model. For both (a) and (b) we show two cases, the unwrapped phase difference between the $z_d = 0$ μm and $z_d = 2$ μm OTF (*solid line*) and the unwrapped phase difference between $z_d = 0$ μm and $z_d = 4$ μm (*dashed line*). All cases show a linear phase ramp with an oscillation of up to $\pi/2$. This phase ramp corresponds to a lateral shift of the PSF. The vectorial case shows an additional curvature and larger overall phase differences of up to 4π radians (or 2 waves) across the spectrum

models, notably a one wave curvature aberration in the vectorial case, and this suggests that vectorial high aperture theory will be important in the future design of high aperture wavefront coding systems.

We can also look at the two models as providing an indication of the difference in performance of CPM wavefront coding between low aperture and high aperture systems. The curvature aberration in the high aperture case varies with defocus, which means that it cannot be incorporated into any single-pass 2D digital deconvolution scheme. This effectively introduces an additional blurring of specimen features in

planes away from focus, lowering the depth of field boost achieved with the same CPM strength in a low aperture wavefront coding system.

In general the CPM performs a little better at low apertures for EDF applications. But the high aperture CPM system still maintains useful frequency response across the full range of an equivalent widefield system, especially for on-axis frequencies.

7.4 Conclusion

Wavefront coding is a new approach to microscopy. Instead of avoiding aberrations, we deliberately create and exploit them. The aperture of the imaging lens still places fundamental limits on performance. However wavefront coding allows us to trade off those limits between the different parameters we need for a given imaging task. Focal range, signal to noise, mechanical focus scanning speed and maximum frequency response are all negotiable using this digital–optical approach to microscopy.

The high aperture experimental results presented here point to the significant promise of wavefront coding. The theoretical simulations predict an altered behaviour for high apertures, which will become more important with higher SNR imaging systems. For large values of defocus, these results predict a tighter limit on the focal range of EDF imaging than is the case for paraxial systems, as well as additional potential for image artefacts due to aberrations.

The fundamental EDF behaviour remains in force at high apertures, as demonstrated by both experiment and theory. This gives a solid foundation to build on. The CPM was part of the first generation wavefront coding design. Using simulations, new phase mask designs can be tested for performance at high apertures before fabrication. With this knowledge, further development of wavefront coding techniques may be carried out, enhancing its use at high apertures.

Acknowledgments

We would like to thank W. Thomas Cathey and Edward R. Dowski Jr. of CDM Optics Inc, Boulder, CO, USA. Experimental assistance was provided by Eleanor Kable, Theresa Dibbayawan, David Philp and Janey Lin at the University of Sydney and Claude Rosignol at Colorado University. Peter Török acknowledges the partial support from the European Union within the framework of the Future and Emerging Technologies-SLAM program and the EPSRC, UK.

References

1. C. J. R. Sheppard: J. Microsc. **149**, 73 (1988)
2. B. Richards, E. Wolf: Proc. Roy. Soc. A **253**, 358 (1959)
3. G. Häusler: Opt. Commun. **6**, 38 (1972)
4. R. Juškaitis, M. A. A. Neil, F. Massoumian, T. Wilson: 'Strategies for wide-field extended focus microscopy'. In: *Focus on microscopy conference* (Amsterdam, 2001)
5. C. J. R. Sheppard, D. K. Hamilton, I. J. Cox: Proc. R. Soc. Lond. A **A387**, 171 (1983)
6. P. Potuluri, M. Fetterman, D. Brady: Opt. Express **8**, 624 (2001)
7. J. Ojeda-Castañeda, R. Ramos, A. Noyola-Isgleas: Appl. Opt. **27**, 2583 (1988)
8. J. Ojeda-Castañeda, E. Tepichin, A. Diaz: Appl. Opt. **28**, 2666 (1989)
9. W. T. Welford: J. Opt. Soc. Am. **50**, 749 (1960)
10. J. Campos, J. C. Escalera, C. J. R. Sheppard, M. J. Yzuel: J. Mod. Optics **47**, 57 (2000)
11. T. Wilson, M. A. A. Neil, F. Massoumian: 'Point spread functions with extended depth of focus'. In: *Proc. SPIE, volume 4621* (San Jose, CA, 2002) pp. 28–31
12. S. Bradburn, W. T. Cathey, E. R. Dowski, Jr.: App. Opt. **36**, 9157 (1997)
13. E. R. Dowski, Jr., W. T. Cathey: App. Opt. **34**, 1859 (1995)
14. W. Chi, N. George: Opt. Lett. **26**, 875 (2001)
15. S. C. Tucker, W. T. Cathey, E. R. Dowski, Jr.: Optics Express **4**, 467 (1999)
16. P. D. Higdon, P. Török, T. Wilson: J. Microsc. **193**, 127 (1999)
17. P. Török, P. Varga, Z. Laczik, G. R. Booker: J. Opt. Soc. Am. A **12**, (325) (1995)
18. B. R. Frieden: J. Opt. Soc. Am. **57**, 56 (1967)
19. C. W. McCutchen: J. Opt. Soc. Am. **54**, 240 (1964)
20. C. J. R. Sheppard, M. Gu, Y. Kawata, S. Kawata: J. Opt. Soc. Am. A **11**, 593 (1994)
21. M. R. Arnison, C. J. R. Sheppard: Opt. Commun. **211**, 53 (2002)
22. C. J. R. Sheppard, K. G. Larkin: Optik **107** 79 (1997)
23. M. Mansuripur: J. Opt. Soc. Am. A **3**, 2086 (1986)
24. H. H. Hopkins: Proc. Phys. Soc. **55**, 116 (1943)
25. C. J. R. Sheppard, C. J. Cogswell: 'Three-dimensional imaging in confocal microscopy'. In: *Confocal microscopy* (Academic Press, London 1990) pp. 143–169.
26. R. N. Bracewell: *Two-dimensional imaging* (Prentice Hall, Englewood Cliffs, NJ 1995)

Part II

Nonlinear Techniques in Optical Imaging

8 Total Internal Reflection Fluorescence Microscopy

Daniel Axelrod

8.1 Features and Applications

Total internal reflection fluorescence (TIRF) microscopy (also called "evanescent wave microscopy") provides a means to selectively excite fluorophores in an aqueous environment very near a solid surface (within $\leq$ 100 nm). In wide-field microscopy, fluorescence excitation by this thin zone of electromagnetic energy (called an "evanescent field") results in images with very low background fluorescence from out-of-focus planes. In contrast to confocal microscopy, TIRF involves minimal exposure of the sample to excitation light in out-of-focus planes. Figure 8.1 shows an example of TIRF on intact living cells in culture, compared with standard epifluorescence (EPI). The unique features of TIRF, often in simultaneous combination with other fluorescence techniques, have enabled numerous applications in chemistry, biochemistry, and cell biology [1]. In the following partial list, we mainly (but not exclusively) emphasize those TIRF applications that have been adapted to microscopy.

(a) *Cell/substrate contact regions.* TIRF can be used qualitatively to observe the position, extent, composition, and motion of contact regions even in samples in which fluorescence elsewhere would otherwise obscure the fluorescent pattern [2–4]. A variation of TIRF to identify cell-substrate contacts involves doping the solution surrounding the cells with a nonadsorbing and nonpermeable fluorescent volume marker; focal contacts then appear relatively dark [5,6].
(b) *Cytoplasmic filament structure and assembly.* Although TIRF cannot view deeply into thick cells, it can display with high contrast the fluorescence-marked submembrane filament structure at the substrate contact regions [7,8]. In combination with very low levels of specific filament labeling, which produces fluorescent speckles, TIRF illumination can visualize filament subunit turnover [9].
(c) *Single molecule fluorescence near a surface* [10–20]. The purpose of single molecule detection is to avoid the ensemble averaging inherent in standard spectroscopies on bulk materials. This enables the detection of spectroscopic features that give evidence of intermediate or transient states that otherwise are obscured. TIRF provides the very dark background needed to observe single fluorophores. Polarized TIRF illumination and detection can provide information about surface-bound single molecule orientation and rotational dynamics [21–23]. Related to single molecule detection is the capability of seeing fluorescence

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

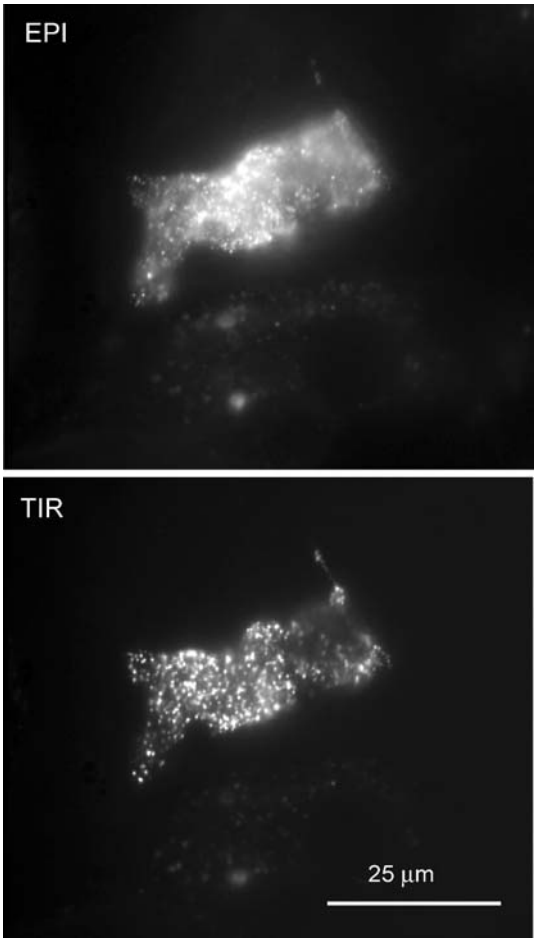


Fig. 8.1. EPI vs. objective-based TIR fluorescence digital images, excited with an argon laser beam of wavelength 488 nm entering the side illumination port of an Olympus IX-70 microscope and viewed through an Olympus 1.45NA 60X objective. This bovine chromaffin cells contains secretory granules marked with GFP-atrial naturetic protein

fluctuations as fluorescent molecules enter and leave the thin evanescent field region in the bulk. These fluctuations (which are visually obvious in TIRF) can be quantitatively autocorrelated in a technique called fluorescence correlation spectroscopy (FCS) to obtain kinetic information about the molecular motion [24,25]. Single quantum dots can be easily located by TIRF and their peculiar flickering behavior studied individually [26].

(d) *Secretory granule tracking and exocytosis.* Numerous types of subcellular organelles encase materials destined for exocytotic release from cells. These organelles (also called granules or vesicles) often reside through the whole depth of the cell and are difficult to distinguish individually in standard epi-illumination when they are fluorescence-marked. TIRF provides very distinct images of only those vesicles closest to the membrane and thereby perhaps closest to being exocytosed. The thin evanescent field allows small motions of individual fluorescence-marked secretory granules to manifest as small intensity changes

arising from small motions in the direction normal to the substrate and plasma membrane. The precision of such tracking can be as small as 2 nm, much smaller than the light microscope resolution limit. In some cases, dispersal of granule contents can be observed and interpreted as exocytosis [27–52].

(e) *Proteins at cell membranes, cell surface receptors and artificial membranes*. TIR combined with fluorescence recovery after photobleaching (FRAP) or FCS can examine the specific or nonspecific binding kinetics of proteins to cellular or artificial membranes [53–66]. TIR/FRAP additionally can be used to measure lateral surface diffusion coefficients along with on/off kinetics of reversibly adsorbed fluorescent molecules [53,54,57,62,65]. Interactions among channel or receptor proteins at surfaces can be studied by TIR/fluorescence resonance energy transfer [67].

(f) *Micromorphological structures and dynamics on living cells*. By utilizing the unique polarization properties of the evanescent field of TIR, endocytotic or exocytotic sites, ruffles and other submicroscopic irregularities can be highlighted [68,69].

(g) *Long-term fluorescence movies of cultured cells.* Since the cells are exposed to TIR excitation light only at their cell/substrate contact regions but not through their bulk, they tend to survive longer under observation, thereby enabling time-lapse recording of a week in duration. During this time, newly appearing cell surface receptors can be immediately marked by fluorescent ligand that is continually present in the full cell culture medium while maintaining a low fluorescence background [70].

(h) *Membrane ion channels and ions at surfaces*. TIRF can be used to visualize single Ca^{2+} channels with good spatial and temporal resolution [71]. TIRF is completely compatible with standard epi-fluorescence, bright field, dark field, or phase contrast illumination. These methods of illumination can be switched back and forth rapidly with TIRF by electro-optic devices to compare membrane-proximal ionic transients and deeper cytoplasmic transients [72]. The spatio-temporal evolution of artificial induced pH gradients at surfaces has been studied by combining TIRF with scanning electrochemical microscopy [73].

(i) *Sensors.* The application of TIRF for biosensors and environmental detectors is an active and growing field in both microscope and non-microscope configurations [74–77].

(j) *Adsorption at liquid/liquid interfaces.* Spectroscopy of molecules adsorbed at liquid hydrophobic/hydrophilic interfaces can be studied with TIRF (both microscopic and nonmicroscopic) provided there is an ample refractive index difference between the two liquids [78–80].

(k) *Brownian motion at surfaces.* A nearby surface can affect the local freedom of motion of a solute for a variety of reasons, including local fields, local viscosity, tethering, steric hindrance, and spatial patterning on the surface The motions can be studied by single molecule or FCS techniques. Larger particles can be followed by nonfluorescent light scattering of evanescent illumination [81–88].

8.2 Theoretical Principles

The thin layer of illumination is an "evanescent field" produced by an excitation light beam in a solid (e.g., a glass coverslip or tissue culture plastic) that is incident at a high angle upon the solid/liquid surface at which the sample (e.g., single molecules or cells) adheres. The incidence angle θ measured from the normal, must be greater than some "critical angle" for the beam to totally internally reflect rather than refract through the interface. TIR generates a very thin electromagnetic field in the liquid with the same frequency as the incident light, exponentially decaying in intensity with distance from the surface. This field is capable of exciting fluorophores near the surface while avoiding excitation of a possibly much larger number of fluorophores farther out in the liquid.

8.2.1 Infinite Plane Waves

The simplest case of TIR is that of an "infinitely"-extended plane wave incident upon a single interface (i.e., a beam width many times the wavelength of the light, which is a good approximation for unfocused or weakly focused light); see Fig. 8.2. When a light beam propagating through a transparent medium 3 of high index of refraction (e.g., glass) encounters a planar interface with medium 1 of lower index of refraction (e.g., water), it undergoes total internal reflection for incidence angles θ measured from the normal to the interface) greater than the "critical angle" θ_c given by:

$$\theta_c = \sin^{-1}\left(n_1/n_3\right) \tag{8.1}$$

where n_1 and $n_3$are the refractive indices of the liquid and the solid respectively. Ratio n_1/n_3 must be less than unity for TIR to occur. (A refractive index n_2 will refer to an optional intermediate layer to be discussed below.) For "subcritical" incidence angles $\theta < \theta_c$, most of the light propagates through the interface into the lower index material with a refraction angle (also measured from the normal) given by Snell's Law. (Some of the incident light also internally reflects back into the solid). But for "supercritical" incidence angles $\theta < \theta_c$, all of the light reflects back into the solid. Even in this case, some of the incident light energy penetrates through the interface and propagates parallel to the surface in the plane of incidence. The field in the liquid (sometimes called the evanescent "wave"), is capable of exciting fluorescent molecules that might be present near the surface.

The intensity I of the evanescent field at any position (measured as perpendicular distance z from the TIR interface) is the squared amplitude of the complex electric field vector $\mathbf{E}$ at that position:

$$I(z) = \mathbf{E}(z) \cdot \mathbf{E}^*(z) \tag{8.2}$$

For an infinitely wide beam, the intensity of the evanescent wave (measured in units of energy/area/sec) exponentially decays with z:

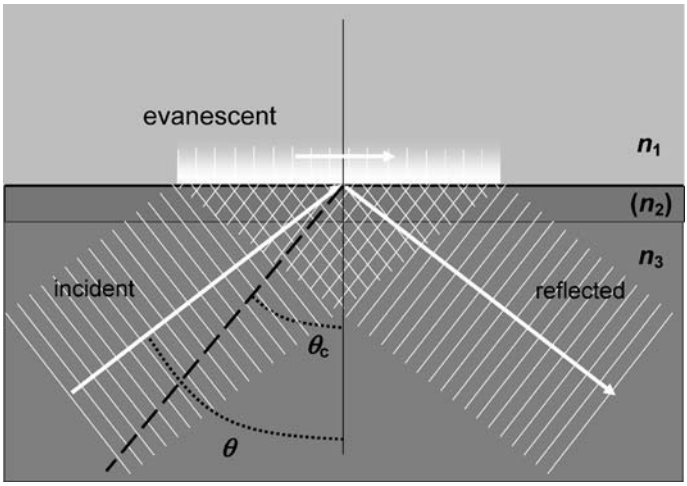


Fig. 8.2. TIR illumination scheme. Refractive index n_3 must be greater than n_1. The intermediate layer (consisting of metal or a dielectric material of refractive index n_2) is not necessary for TIR to occur, but can introduce some useful optical properties as explained in the text. Most applications of TIRF do not use an intermediate layer. The incidence angle θ must be larger than the critical angle θ_c for TIR to occur. The exponentially-decaying evanescent field in the n_1 material is used to excite fluorophores in TIRF

$$I(z) = I(0)\mathrm{e}^{-z/d} \tag{8.3}$$

where

$$\begin{aligned} d &= \frac{\lambda_o}{4\pi}\left(n_3^2 \sin^2\theta - n_1^2\right)^{-1/2} \\ &= \frac{\lambda_o}{4\pi n_3}\left(\sin^2\theta - \sin^2\theta_c\right)^{-1/2} \end{aligned} \tag{8.4}$$

Parameter λ_o is the wavelength of the incident light in vacuum. Depth d is independent of the polarization of the incident light and decreases with increasing θ. Except for supercritical $\theta \to \theta_c$ (where $d \to \infty$), d is generally the order of λ_o or smaller. A physical picture of refraction at an interface shows TIR to be part of a continuum, rather than a sudden new phenomenon appearing at $\theta = \theta_c$. For small θ, the refracted light waves in the liquid are sinusoidal, with a certain characteristic period noted as one moves normally away from the surface. As θ approaches θ_c, that period becomes longer as the refracted rays propagate increasingly parallel to the surface. At exactly $\theta = \theta_c$, that period is infinite, since the wavefronts of the refracted light are themselves normal to the surface. This situation corresponds to $d = \infty$. As θ increases beyond θ_c, the period becomes mathematically imaginary. Physically, this corresponds to the exponential decay of (8.3).

The local intensity I of the evanescent field at any point is proportional to the probability rate of energy absorption by a fluorophore that can be situated at that

point. Although the emission from a fluorophore excited by an evanescent field as viewed by a microscopic objective might be expected to follow an exponential decay with z according to (8.3), this is not precisely true. Fluorescence emission near a dielectric interface is rather anisotropic and the degree of anisotropicity is itself z-dependent [89,90]. One cause of the anisotropicity of emission is simply partial reflection from the interface and consequent interference between the direct and reflected emission beams. But another more subtle cause is the interaction of the fluorophore's "near field" with the interface and its consequent conversion into light propagating at high angles into the higher index n_3 material (the solid substrate). In general, the closer a fluorophore is to the surface, the larger the proportion of its emitted light will enter the substrate. If a sufficiently high aperture objective is used to gather the near field emission, then the effective thickness of the surface detection zone is reduced beyond the surface selectivity generated by the evanescent field excitation of TIR. Low aperture objectives will not produce this enhanced effect because they miss capturing the near field light. On its own, the near-field emission capture effect can be the basis of surface selective fluorescence detection and imaging, as discussed in a separate section below.

The polarization (i.e., the vector direction of the electric field **E**) of the evanescent wave depends on the incident light polarization, which can be either "p-pol" (polarized in the plane of incidence formed by the incident and reflected rays, denoted here as the x-z plane) or "s-pol" (polarized normal to the plane of incidence).

For p-pol incident light, the evanescent electric field vector direction remains in the plane of incidence, but it "cartwheels" along the surface with a nonzero longitudinal component (see Fig. 8.3):

$$\mathbf{E}_{\mathrm{p}}(z) = 2\cos\theta\left(\sin^4\theta_{\mathrm{c}}\cos^2\theta + \sin^2\theta - \sin^2\theta_{\mathrm{c}}\right)^{1/2} \mathrm{e}^{-i\delta_{\mathrm{p}}} e^{-z/2d} \left[-i\left(\sin^2\theta - \sin^2\theta_{\mathrm{c}}\right)^{1/2}\hat{\mathbf{x}} + \sin\theta\,\hat{\mathbf{z}}\right] \tag{8.5}$$

The evanescent wavefronts travel parallel to the surface in the x-direction. Therefore, the p-pol evanescent field is a mix of transverse (z) and longitudinal (x) components; this distinguishes the p-pol evanescent field from freely propagating subcritical refracted light, which has no component longitudinal to the direction of travel. The longitudinal component of the p-pol evanescent field diminishes to zero amplitude as the incidence angle is reduced from the supercritical range back toward the critical angle.

For s-pol incident light, the evanescent electric field vector direction remains purely normal to the plane of incidence:

$$\mathbf{E}_{\mathrm{s}}(z) = 2\frac{\cos\theta}{\cos\theta_{\mathrm{c}}}\mathrm{e}^{-i\delta_{\mathrm{s}}}\mathrm{e}^{-z/2d}\hat{\mathbf{y}} \tag{8.6}$$

In (8.5) and (8.6), the incident electric field amplitude in the substrate is normalized to unity for each polarization, and the phase lags relative to the incident light are:

$$\delta_{\mathrm{p}} = \tan^{-1}\left[\frac{(\sin^2\theta - \sin^2\theta_{\mathrm{c}})^{1/2}}{\sin^2\theta_{\mathrm{c}}\cos\theta}\right] \tag{8.7}$$

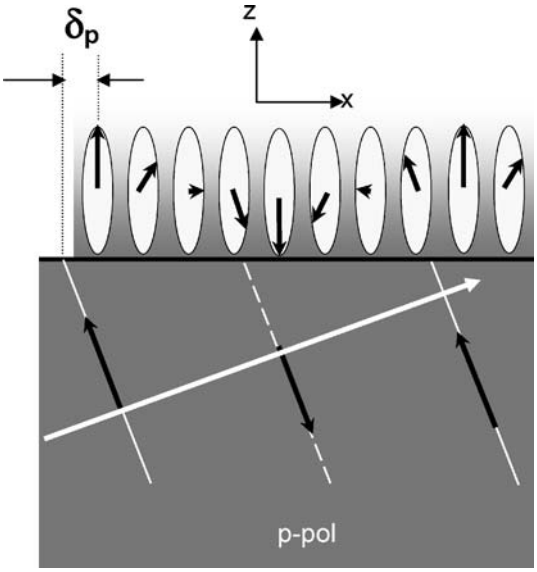


Fig. 8.3. Schematic drawing of the evanescent polarization resulting from p-pol incident light. The incident light wavefronts (with the intervals from solid to dashed wavefront lines representing one-half of a wavelength in the glass) determine the spacing of the wavefronts in the evanescent field. The wavelength in the n_3 medium is λ_o/n_3. Reflected light is not shown. The p-pol evanescent field is elliptically polarized in the x-z plane as shown (primarily polarized in the z-direction with a weaker x-component at a relative phase of $\pi/2$). For pictorial clarity, only one full period of evanescent electric field oscillation are shown; in reality, the evanescent region is much more extended and contains many more periods of oscillation in the x-direction. The exact phase relationship between the incident field and the evanescent field is a function of incidence angle and is represented symbollically by δ_p here. δ_p is a phase angle, not a distance

$$\delta_s = \tan^{-1}\left[\frac{(\sin^2\theta - \sin^2\theta_c)^{1/2}}{\cos\theta}\right] \tag{8.8}$$

The corresponding evanescent intensities in the two polarizations (assuming incident intensities normalized to unity) are:

$$I_p(z) = \frac{(4\cos^2\theta)(2\sin^2\theta - \sin^2\theta_c)}{\sin^4\theta_c\cos^2\theta + \sin^2\theta - \sin^2\theta_c}e^{-z/d} \tag{8.9}$$

$$I_s(z) = \frac{(4\cos^2\theta)}{\cos^2\theta_c}e^{-z/d} \tag{8.10}$$

Intensities $I_{p,s}(0)$ are plotted vs. θ in Fig. 8.4. The evanescent intensity approaches zero as $\theta \to 90°$. On the other hand, for super-critical angles within ten degrees of θ_c, the evanescent intensity is as great or greater than the incident light intensity. The plots can be extended without breaks to the sub-critical angle range, where the intensity is that of the freely propagating refracted light in medium 1. One might

at first expect the subcritical intensity to be slightly *less* than the incident intensity (accounting for some reflection at the interface) but certainly not *more* as shown. The discrepancy arises because the intensity in Fig. 8.3 refers to EE^* alone rather than to the actual energy flux of the light, which involves a product of EE^* with the refractive index of the medium in which the light propagates.

Regardless of polarization, the spatial period of the evanescent electric field is $\lambda_o/(n_3 \sin\theta)$ as it propagates along the surface. Unlike the case of freely propagating light, the evanescent spatial period is not at all affected by the medium 1 in which it resides. It is determined only by the spacing of the incident light wavefronts in medium 3 as they intersect the interface. This spacing can be important experimentally because it determines the spacing of interference fringes produced when two coherent TIR beams illuminate the same region of the surface from different directions.

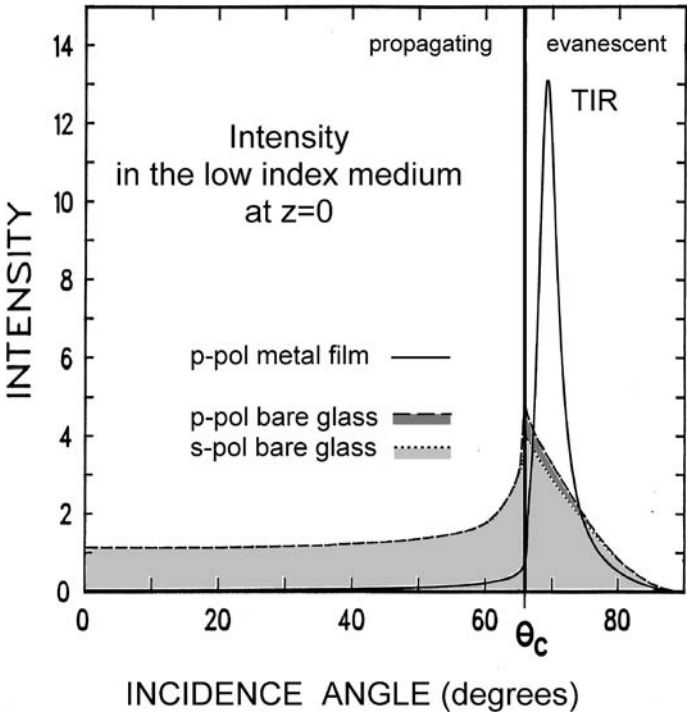


Fig. 8.4. Evanescent intensities $I_{p,s}$ at $z = 0$ vs. θ, assuming the incident intensities in the glass are set equal to unity. At angles $\theta > \theta_c$, the transmitted light is evanescent; at angles $\theta > \theta_c$, it is propagating. Both s- and p-polarizations are shown. Refractive indices $n_3 = 1.46$ (fused silica) and $n_1 = 1.33$ are assumed here, corresponding to $\theta_c = 65.7°$. Also shown is the evanescent intensity that would be obtained with a thin (20 nm) aluminum film coating, as discussed in the Intermediate Films section

8.2.2 Finite Width Incident Beams

For a finite-width beam, the evanescent wave can be pictured as the beam's partial emergence from the solid into the liquid, travel for some finite distance along the surface, and then re-entrance into the solid. The distance of propagation along the surface is measurable for a finite-width beam and is called the Goos-Hanchen shift. The Goos-Hanchen shift ranges from a fraction of a wavelength at $\theta = 90°$ to infinite at $\theta = \theta_c$, which of course corresponds to the refracted beam skimming along the interface. A finite incident beam can be expressed mathematically as an integral of infinite plane waves approaching at a range of incidence angles. In general, the intensity profile of the resulting evanescent field can be calculated from the mathematical form for the evanescent wave that arises from each infinite plane wave incidence angle, integrated over all the constituent incident plane wave angles. For a TIR Gaussian laser beam focused with a typically narrow angle of convergence, the experimentally observed evanescent illumination is approximately an elliptical Gaussian profile, and the polarization and penetration depth are approximately equal to those of a single infinite plane wave.

8.2.3 Intermediate Layers

In actual experiments in biophysics, chemistry, or cell biology, the interface may not be a simple interface between two media, but rather a stratified multilayer system. One example is the case of a biological membrane or lipid bilayer interposed between glass and an aqueous medium. Another example is a thin metal film coating, which can be used to quench fluorescence within the first ~10 nm of the surface. We discuss here the TIR evanescent wave in a three-layer system in which incident light travels from medium 3 (refractive index n_3) through the intermediate layer (n_2) toward medium 1 (n_1). Qualitatively, several features can be noted:

(a) Insertion of an intermediate layer never thwarts TIR, regardless of the intermediate layer's refractive index n_2. The only question is whether TIR takes place at the $n_3 : n_2$ interface or the $n_2 : n_1$ interface. Since the intermediate layer is likely to be very thin (no deeper than several tens of nanometers) in many applications, precisely which interface supports TIR is not important for qualitative studies.
(b) Regardless of n_2 and the thickness of the intermediate layer, the evanescent wave's profile in medium 1 will be exponentially decaying with a characteristic decay distance given by Eq. (8.3). However, the intermediate layer affects intensity at the interface with medium 1 and the overall distance of penetration of the field as measured from the surface of medium 3.
(c) Irregularities in the intermediate layer can cause scattering of incident light that then propagates in all directions in medium 1. Experimentally, scattering appears not be a problem on samples even as inhomogeneous as biological cells. Direct viewing of incident light scattered by a cell surface lying between the glass substrate and an aqueous medium confirms that scattering is many orders of magnitude dimmer than the incident or evanescent intensity, and will thereby excite a correspondingly dim contribution to the fluorescence.

A particularly interesting kind of intermediate layer is a metal film. Classical electromagnetic theory [89] shows that such a film will reduce the s-polarized evanescent intensity to nearly zero at all incidence angles. But the p-polarized behavior is quite different. At a certain sharply defined angle of incidence θ_p ("the surface plasmon angle"), the p-polarized evanescent intensity becomes an order of magnitude brighter than the incident light at the peak (see Fig. 8.4). This strongly peaked effect is due to a resonant excitation of electron oscillations at the metal/water interface. For an aluminum, gold, or silver film at a glass/water interface, θ_p is greater than the critical angle θ_c for TIR. The intensity enhancement is rather remarkable since a 20 nm thick metal film is almost opaque to the eye.

There are some potentially useful experimental consequences of TIR excitation through a thin metal film coated on glass:

(a) The metal film will almost totally quench fluorescence within the first 10 nm of the surface, and the quenching effect is virtually gone at a distance of 100 nm. Therefore, TIR with a metal-film coated glass can be used to selectively excite fluorophores in the 10 nm to 200 nm distance range.
(b) A light beam incident upon a 20 nm aluminum film from the glass side at a glass/aluminum film/water interface evidently does not have to be collimated to produce TIR. Those rays that are incident at the surface plasmon angle will create a strong evanescent wave; those rays that are too low or high in incidence angle will create a negligible field in the water. This phenomenon may ease the practical requirement for a collimated incident beam in TIR and make it easier to set up TIR with a conventional arc light source.
(c) The surface plasmon angle is strongly sensitive to the index of refraction on the low density side of the interface. With a sample of spatially varying refractive index (such as an adhered cell), some areas will support the surface plasmon resonance while neighboring areas will not. The locations supporting a resonance will show a much stronger light scattering. This effect can be used as the basis of "surface plasmon microscopy" [91].
(d) The metal film leads to a highly polarized evanescent wave, regardless of the purity of the incident polarization.

Another type of intermediate film is now coming into commercial use in bioassay surfaces based on TIRF: the thin film planar waveguide. An ordinary glass (or plastic) surface is etched or embossed with closely spaced grooves that behave like a diffraction grating so that some of the light coming through the glass toward the surface will diffract away from the normal at the surface. Coated upon the grooved glass is a thin intermediate film of a material (such as titanium dioxide) with refractive index higher than that of either the glass substrate or the water at its boundaries. The diffracted incident light passing from the glass into the intermediate film has a sufficient angle not only to totally reflect first at the thin film/aqueous interface but also to totally reflect subsequently at the thin film/glass interface, thereby trapping the light in the film. At a very particular angle (which depends on the wavelength and the film thickness), the film acts like an optical waveguide where the

reflections at the two surfaces create wavefronts that constructively add in-phase with each other. The result is a very strong field in the high index film and a corresponding very strong evanescent field in the aqueous medium that can be used to excite fluorescence in biosensors devices [92]. The thin film waveguide surface preparation should also be applicable to observation of living cells in a microscope.

8.2.4 Combination of TIR with Other Fluorescence Techniques

TIR illumination can be combined with other standard fluorescence spectroscopy techniques. Brief comments here point out certain unique properties of the combinations.

(a) *TIR-FRET* (fluorescence resonance energy transfer) provides an opportunity to observe real-time changes in the conformation of single molecules attached to the substrate due to the much reduced background provided by TIR.
(b) *TIR-fluorescence lifetime measurements* should give somewhat different results than results of the same fluorophores in solution. Proximity to the surface may directly perturb the molecular state structure, thereby affecting lifetimes. In addition, the near field capture effect generally increases the rate of transfer of energy out of the fluorophore and thereby decreases lifetimes. The rotational mobility of surface-proximal molecules can be studied by TIR combined with fluorescence anisotropy decay [93,94].
(c) In *polarized TIR*, a p-polarized evanescent field can be uniquely utilized to highlight submicroscopic irregularities in the plasma membrane of living cells [68], as shown schematically in Fig. 8.5. The effect depends on the incorporation of fluorophore into the membrane with a high degree of orientational order. In the case of the phospholipid-like amphipathic carbocyanine dye 3-3' dioctadecylindocarbocyanine (diI), the direction of absorption dipole of the fluorophores is known to be parallel to the local plane of the membrane and free to rotate in it [95].

    If one views singly-labeled single molecules rather than membranes with highly oriented heavy labeling, the ensemble averaging problem disappears and the requirement for orientational order can be eliminated. With this approach, polarized TIR can successfully determine the orientation of single molecules of sparsely-labeled F-actin [22].

    An unusual kind of single molecule polarized TIR has been demonstrated that depends on optical aberrations. Molecules with emission dipoles oriented orthogonally to the substrate emit very little of their light along the dipole axis. Most their light is emitted at more skimming angles to the substrate surface; this light is captured by the periphery of high aperture objectives. If a source of aberration is deliberately introduced into the detection system (e.g., a layer of water), those extreme rays will not focus well into a spot but instead form a small donut of illumination at the image plane. This donut can be readily distinguished from a dipole oriented parallel to the substrate, which focuses well to a spot [96].

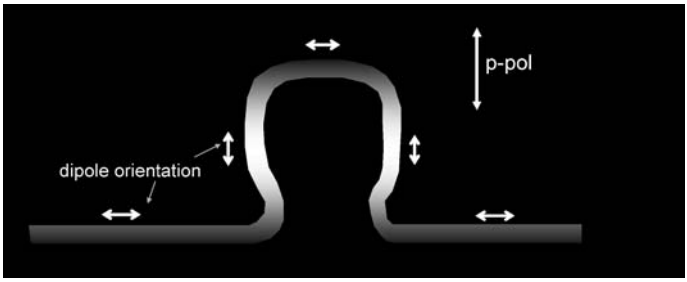


Fig. 8.5. Schematic drawing of the excitation probability of oriented carbocyanine fluorophores embedded in a membrane in a *z*-polarized evanescent field (the dominant direction of a p-polarized evanescent field). The membrane is depicted in cross section with a curved region corresponding to a bleb or an invagination. The direction of absorption dipole of the fluorophores is shown with bidirectional arrows. Higher excitation probability is depicted by lighter shades. The *z*-component of the electric field selectively excites the regions of oblique membrane orientation

The asymmetry of the emission from oriented dipoles can manifest itself as an anisotropy of intensity at the back focal plane of the objective, since the back focal plane essentially forms a spatial map of the angular distribution of rays emanating from the sample molecule. This technique has been demonstrated on single molecules illuminated with standard epi optics [97], but it should also work well with TIR illumination.

It may sometimes be desirable to suppress the effect of molecular orientation or rotation on the observed fluorescence, which might otherwise be mistaken for fluctuations in the local concentration or chemical environment. Two orthogonally directed TIR beams with appropriate polarizations can be made to produce isotropic excitation in two or three dimensions [98].

(d) In *TIR-FRAP* (fluorescence recovery after photobleaching), a bright flash of the evanescent field bleaches the surface-proximal fluorophores, and the subsequent rate of recovery of fluorescence excited by dimmer evanescent illumination provides a quantitative measure of the desorption kinetic rate k_{off} [53,54]. A concentration jump in the bulk, rather than a bleaching flash, provides a different kind of perturbation and leads to a relaxation rate in the surface proximal concentration that depends on a linear combination of both k_{off} and the adsorption kinetic rate k_{on}. Therefore, by combining concentration jump experiments with FRAP experiments on the same TIR-illuminated sample, both k_{off} and k_{on} can be deduced separately [99].

(e) In *TIR-FCS* (fluorescence correlation spectroscopy), the thin evanescent field is combined with an image plane diaphragm that passes light only from a small (typically sub-micron) lateral region. This defines an extremely small observation volume ($\sim 0.02\,\mu m^3$) in the sample. Random fluctuations in the number of fluorophores in this volume provide information about the diffusive rate in the bulk solution near the surface, the kinetic desorption rate, and the absolute concentration of surface-proximal fluorophores [24,53,100].

(f) In *multiphoton TIRF*, excitation by an intense flash of evanescent infra-red light excites a fluorescent response that is proportional to the square (or higher order) of the incident light intensity [101–103]. This might seem at first to have some advantages for further decreasing the effective evanescent wave depth d. Although d for two photon TIR excitation should be half that of the single photon case described by (8.3), note that the infrared wavelength is double what would have been used in "single photon" TIRF, so the effect on the depth d would be minimal (apart from chromatic dispersion effects in the substrate and water). On the other hand, scattering of the evanescent field by inhomogeneities in the sample would likely not be bright enough to excite much multiphoton fluorescence, so multiphoton TIRF may produce a cleaner image, less affected by scattering artifacts.

One theoretical study proposes a three-way combination of TIR with two-photon excitation and with stripe-pattern FRAP. The advantage there would be a higher-contrast imprinted bleach pattern [104].

(g) *Optical trapping with TIRF* should be possible since the gradient of the evanescent intensity can be made comparable to the gradient of intensity in standard focused spot optical traps. The evanescent optical gradient force will always be directed toward the interface. The analog of a "photon pressure" force may be directed parallel to the substrate in the plane of incidence; it might be canceled out by the use of oppositely directed TIR beams.

(h) *TIR and optical interference techniques*. Specific ligand binding to receptors immobilized on a surface can be viewed simultaneously with nonspecific mass adsorption by combining TIRF with reflectance interferometry, thus far in non-microscopic mode [105]. By combining TIRF with interference reflection contrast (IRC), Ca^{2+} microdomains can be viewed simultaneously with exocytosis events in synapses [106]. Both TIR and IRC were used on the same sample to study the approach of neuronal growth cones to the substrate [107].

(i) *TIR with atomic force microscopy.* By combining TIRF with atomic force microscopy, micromechanical properties can be directly measured on living cells [108,109] and labeled molecules can be located more easily [110].

8.2.5 Surface Near Field Emission Imaging

Since TIR selectively excites only surface-proximal fluorophores, the special emission behavior of such fluorophores should be considered. This emission behavior also suggests designs for potentially useful microscope modifications.

The classical model for an excited fluorophore in electromagnetic theory is an oscillating dipole. The radiation emission pattern of an oscillating dipole can be expressed (by spatial Fourier transforming) as a continuous superposition of plane waves traveling in all directions from the fluorophore. Some (but not all) of these plane waves have wavelengths given by $\lambda = c/(n\nu)$ as expected for propagating light, where ν is the frequency (color) of the light, n is the refractive index of the

liquid in which the fluorophore resides, and c is the speed of light in vacuum. However, this restricted set of plane waves is not sufficient to describe the actual radiation emission pattern; other plane waves with shorter wavelengths must be included. Since the frequency of the light is invariant as determined by the color, the only way to obtain shorter wavelengths is for the plane waves to be exponentially decaying in one of the three spatial directions.

In a slightly more mathematical view [89], a plane wave has a wave vector $\boldsymbol{k}$ given by:

$$\boldsymbol{k} = k_x\hat{\boldsymbol{x}} + k_y\hat{\boldsymbol{y}} + k_z\hat{\boldsymbol{z}} \tag{8.11}$$

where the z-direction is chosen as the normal to the interface, and the square scalar amplitude of $\boldsymbol{k}$ is fixed at a constant value by the frequency ν according to

$$k^2 = (2\pi n\nu/c)^2 = k_x^2 + k_y^2 + k_z^2 \tag{8.12}$$

Short wavelengths are obtained by using plane waves with $\left(k_x^2 + k_y^2\right) > k^2$, which forces k_z^2 to be negative. This forces k_z to be imaginary, which corresponds to plane waves that exponentially decay in the z direction. The exponential decay "starts" at the z position of the fluorophore. The fluorophore's "near field" is defined as this set of exponentially decaying plane waves with $\left(k_x^2 + k_y^2\right) > k^2$. Clearly, this set is not a single exponential but superposition (actually a continuous integral) of exponentially decaying waves of a range of characteristic decay lengths each given by $2\pi/|k_z|$. These near field waves have wavefronts more closely spaced than would be expected for propagating light in the liquid medium surrounding the fluorophore. Where the "tails" of the exponentially decaying wavefronts touch the surface of the glass, refraction converts some of the near field energy into propagating light in the glass.

Because the periodicity of the wavefronts must match at both sides of the boundary, the near field-derived propagating light in the glass is directed in a hollow cone *only* into angles greater than some critical angle ("supercritical"). (That critical angle is the very same as would be the TIR critical angle for the same frequency of light passing from the solid toward the liquid.). *None* of the far-field propagating energy, for which $\left(k_x^2 + k_y^2\right) \leq k^2$, is cast into that high angle hollow cone; it is all cast into "subcritical" angles. Therefore, collection and imaging of just the supercritical high angle hollow cone light should select only those fluorophores that are sufficiently close to the surface for the surface to capture their near fields. An implementation of this principle is discussed in the Optical Configurations section below.

Note that the light-collecting advantage of very high aperture objectives (> 1.4) resides purely in their ability to capture supercritical near-field light. The "extra" numerical aperture does not help in gathering far-field emission light because none of it propagates in the glass at the supercritical angles (and thence into the corresponding high apertures).

The capture of the fluorophore near-field and its conversion into light propagating into the substrate at supercritical angles causes deviations from the expected

exponential decay of the TIR evanescent field implied by (8.3). It is true that the *excitation* rate follows the exponential decay in z, but the *emission* pattern depends on the distance the fluorophore resides from the surface in a complicated manner. Therefore, since every objective gathers only part of this emission pattern, the dependence of gathered emission upon z will be nonexponential and will depend upon the numerical aperture of the objective (as well as upon the orientational distribution of the fluorophore dipoles, the refractive indices bordering the interface, and the polarization of the excitation). For objectives of moderate aperture (NA < 1.33) through which no near field light is gathered, the deviation from the simple exponentiality is generally only 10 − 20%. For very high aperture objectives that do gather some near-field emission, the deviation is larger and generally leads to a steeper z-dependence. This corrected z-dependence can be approximated as an exponential with a shorter decay distance, but the exact shape is not truly exponential.

A metal coated surface affects the emission, both the intensity and the angular pattern. Far-field light for which $(k_x^2 + k_y^2) \leq k^2$ reflects off the metal surface and very little penetrates through. But for a near-field component with a very particular $(k_x^2 + k_y^2) > k^2$, a surface plasmon will form in the metal and then produce a thin-walled hollow cone of emission light propagating into the glass substrate [111]. This hollow cone of light can be captured by a high aperture objective [112] as can be seen in Fig. 8.6, and it could potentially be used for imaging. The rapid conversion of near-field light into either a surface plasmon or heat (via the quenching effect mentioned earlier) leads to a much reduced excited state lifetime for fluorophores near a surface with a consequent great reduction in photobleaching rate. These effects could form the basis of a new sensitive means of single fluorophore detection with minimal photobleaching.

8.2.6 Measurement of Distances from a Surface

In principal, the distance z of a fluorophore from the surface can be calculated from the fluorescence intensity Fwhich as an approximation might be considered to be proportional to the evanescent intensity $I(z)$ as given by Eq. (8.2). In practice, the situation is more complicated for several reasons.

(a) The proportionality factor between F and I depends on efficiencies of absorption, emission, and detection, all of which can be (at least slightly) z-dependent in a complicated and generally not well known manner.
(b) $F(0)$ may not be known if a fluorophore positioned exactly at $z = 0$ (the substrate surface) cannot be identified. In this circumstance, (8.3) can be used only to calculate relative distances. That is, if a fluorophore moves from some (unknown) distance z_1 with observed intensity I_1 to another distance z_2 with intensity I_2, then assuming (as an approximation) an exponentially decaying $F(z)$,

$$\Delta z = z_1 - z_2 = \mathrm{d} \ln(I_2/I_1) \tag{8.13}$$

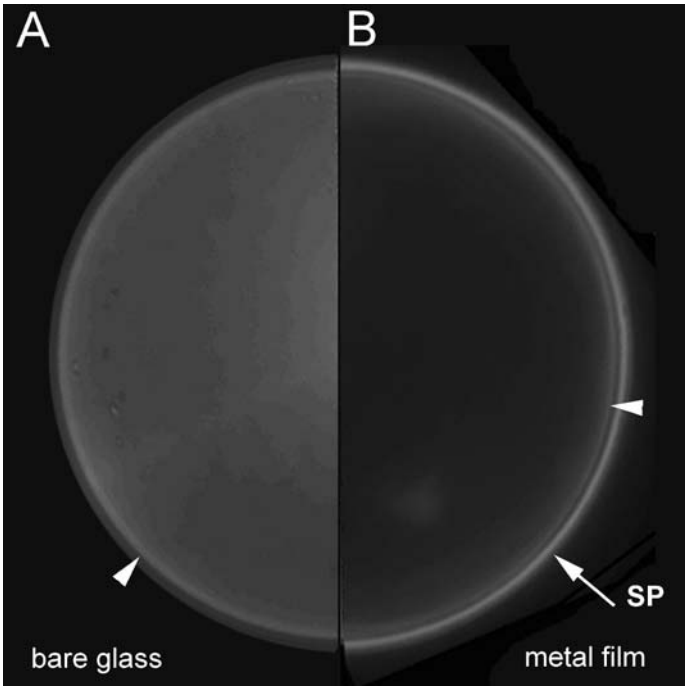


Fig. 8.6. Back focal plane (BFP) intensity distribution produced by a fluorescent sample. The view is split down the middle between that for (**A**) bare glass coverslip and (**B**) coverslip coated with 20 nm of aluminum. Note that angles of emission map into radii at the BFP. On bare glass, the intensity peaks at the critical angle for the emission wavelength, as expected from theory [89]; an arrowhead indicates this intensity peak. On metal film coated glass, this critical intensity peak still appears at the same radius, again shown with an arrowhead. But a new strong peak also appears at a larger radius (and thereby larger emission angle) due to the surface plasmon emission light (SP). The objective was a 63X 1.4NA. The sample was a silica bead (refractive index 1.42), surface coated with carbocyanine dye in an index-matched glycerol/water medium, although the nature of the sample is not detectable from the BFP intensity pattern shown here

This relationship is (approximately) valid even if the fluorescent structure consists of multiple fluorophores and is large or irregularly shaped (e.g., a secretory granule). Again assuming an exponentially decaying $F(z)$, a motion of Δz for the whole structure causes each fluorophore in the structure to change its fluorescence by the same multiplicative factor, and the whole structure therefore changes its fluorescence by that same factor.

For small frame-to-frame changes in I, Eq. (8.13) can deduce very small motions of subcellular organelles, as small as 2 nm. But this deduction should be interpreted cautiously. Shot noise and CCD camera dark noise and readout noise can cause apparent fluctuations in I where none exist. A theoretical method of

separating out the actual Δz due to organelle motion from the artifactually overestimated Δz that includes the effect of instrumental noise is provided in [44].

(c) The z-dependence of the emission pattern (because of near-field capture by the substrate as discussed above) causes $F(z)$ to deviate from exponential decay behavior, as discussed above. Since the deviation depends on numerical aperture in a theoretically predictable manner, comparison of the emission intensity gathered by moderate and very high aperture objectives on the same sample should allow calculation of absolute z distances.

(d) For TIRF studies on biological cells, the distance of a fluorophore or organelle from the plasma membrane is likely to be of more interest than its distance from the substrate surface. The absolute distance between the plasma membrane and the substrate can be deduced by adding a membrane impermeant fluorescent dye to the medium of the cell culture under TIRF illumination. (For viewing labeled organelles at the same time as this impermeant dye, fluorophores of distinctly different spectra should be selected.) In off-cell regions, the fluorescence F_{offcell} will appear uniformly bright, arising from the full depth of the evanescent field. In cell-substrate contacts, the dye will be confined to a layer thinner than the evanescent field. The fluorescence F in contact regions will be darker by a factor that can be converted to separation distance h between the substrate and the plasma membrane gap according to:

$$h = -\mathrm{d} \ln\left[1 - F/F_{\text{offcell}}\right] \tag{8.14}$$

This formula is an approximation because it assumes an exponential decay of gathered fluorescence intensity vs z and thereby neglects the near-field effects discussed earlier.

8.2.7 Variable Incidence Angle TIR: Concentration Profiles

It may be of interest to deduce what is the concentration profile $C(z)$ of fluorophores within the evanescent field, to a resolution much shorter than the evanescent field depth. As can be seen from Eqs. (8.4), (8.9) and (8.10), the exponential decay depth d and intensity+I (at $z = 0$) of the evanescent field both vary with polar incidence angle θ in a well-predicted manner. The observed fluorescence $F(\theta)$ is

$$F(\theta) = \int_{z=0}^{\infty} \beta(z)\, I_{z=0}(\theta)\, C(z) \mathrm{e}^{-z/\mathrm{d}(\theta)} \mathrm{d}z \tag{8.15}$$

where $\beta(z)$ combines all the previously discussed effects of z-dependent emission collection efficiency through the chosen numerical aperture objective. In principal, it should be possible to measure F at several different values of θ and use Eq. (8.15) to rule in or rule out various hypothesized forms for $C(z)$. This approach has been employed in practice to measure cell surface-to-substrate separation distances [113] and the z-position of subcellular organelles [114].

8.2.8 Image Deconvolution

Deconvolution of an image with the actual imaging point spread function (PSF) (or blind deconvolution that does not require an input of the PSF) is a form of filtering used frequently to sharpen images. With standard epi-illumination, a stack of images stepped through z must be obtained because features in any one z-plane appear out-of-focus in adjacent planes. The deconvolution procedure must then be iterative, involving at least neighboring planes. This procedure is fairly computer-intensive and time-consuming. In the case of TIR, deconvolution is extremely fast because no iteration is necessary: adjacent out-of-focus planes make no contribution to the image.

8.3 Optical Configurations

A wide range of optical arrangements for TIRF can been employed. Some configurations use a high numerical aperture (NA > 1.4) microscope objective for both TIR illumination and emission observation, and others use a prism to direct the light toward the TIR interface with a separate objective for emission observation. This section gives examples of these arrangements. We assume for explanatory convenience that the sample consists of fluorescence-labeled cells in culture adhered to a glass coverslip.

8.3.1 High Aperture Objective-Based TIR

By using an objective with a sufficiently high NA, supercritical angle incident light can be cast upon the sample by illumination through the objective [115,116]. Although an arc lamp can be used as the light source, the general features are best explained with reference to a laser source. The optical system has the following general features:

(a) The laser beam used for excitation is focused (by an external focusing lens) to a point at the back focal plane of the objective so that the light emerges from the objective in a collimated form (i.e., the "rays" are parallel to each other). This insures that all the rays are incident upon the sample at the same angle θ with respect to the optical axis.
(b) The point of focus in the back focal plane is adjusted to be off-axis. There is a one-to-one correspondence between the off-axis radial distance ρ and the angle θ. At a sufficiently high ρ, the critical angle for TIR can be exceeded. Further increases in ρ serve to reduce the characteristic evanescent field depth d in a smooth and reproducible manner.
The beam can emerge into the immersion oil (refractive index n_{oil}) at a maximum possible angle θ_{m} measured from the optical axis) given by:

$$\mathrm{NA} = n_{\mathrm{oil}} \sin \theta_{\mathrm{m}} \tag{8.16}$$

Since $n \sin\theta$ is conserved (by Snell's Law) as the beam traverses through planar interfaces from one material to the next, the right side of (8.16) is equal to $n_3 \sin\theta_3$ (where subscript 3 refers to coverslip substrate upon which the cells grow). For total internal reflection to occur at the interface with an aqueous medium of refractive index n_1, θ_3 must be greater than the critical angle θ_c as calculated from

$$n_1 = n_3 \sin\theta_c \tag{8.17}$$

From (8.16) and (8.17), it is evident that the NA must be greater than n_1, preferably by a substantial margin. This is no problem for an interface with water with $n_1 = 1.33$ and a NA = 1.4 objective. But for viewing the inside of a cell at $n_1 = 1.38$, a NA = 1.4 objective will produce TIR at just barely above the critical angle. The evanescent field in this case will be quite deep, and dense heterogeneities in the sample (such as cellular organelles) will convert some of the evanescent field into scattered propagating light.

Fortunately, objectives are now available with NA > 1.4. The highest aperture available is an Olympus 100X 1.65NA; this works very well for objective-based TIRF on living cells. However, that objective requires the use of expensive 1.78 refractive index coverslips made of either LAFN21 glass (available from Olympus) or SF11 glass (custom cut by VA Optical Co, San Anselmo, CA). SF11 glass is the less expensive of the two but it has a chromatic dispersion not perfectly suited to the objective, thereby requiring slight refocusing for different fluorophores. The 1.65 objective also requires special $n = 1.78$ oil (Cargille) which is volatile and leaves a crystalline residue. Several other objectives that are now available circumvent these problems: Olympus 60X 1.45NA, Olympus 60X 1.49NA, Zeiss 100X 1.45NA, and four Nikon objectives: Nikon 60X 1.45NA w/correction collar, Nikon 100X 1.45NA, Nikon 60X 1.49NA w/correction collar, and Nikon 100X 1.49NA w/correction collar. The 1.45 – 1.49 objectives all use standard glass (1.52 refractive index) coverslips and standard immersion oil and yet still have an aperture adequate for TIRF on cells. The 1.49NA objectives are probably the method of choice for TIR except for cells which have particularly dense organelles. Dense organelles tend to scatter the evanescent field, and this effect is less prominent with the higher angles of incidence accessible through higher aperture objectives.

(c) The angle of convergence/divergence of the laser beam cone at the back focal plane is proportional to the diameter of the illuminated region subsequently created at the sample plane. Large angles (and consequent large illumination regions) can be produced by use of a beam expander placed just upbeam from the focusing lens.

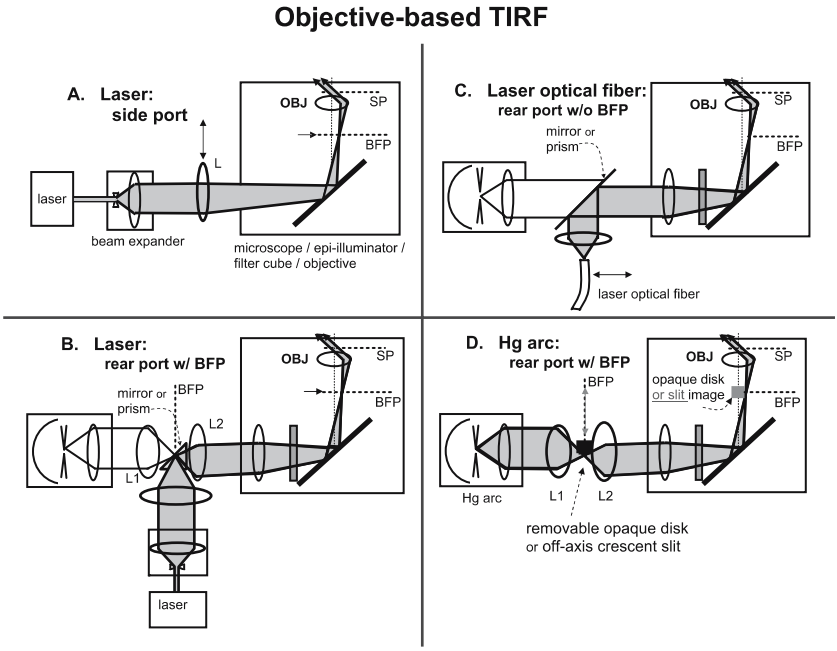
(d) The orientation of the central axis of the laser beam cone at the back focal plane determines whether the TIR-illuminated portion of the field of view is centered on the microscope objective's optical axis.

A microscope can be configured in several variations for objective-based TIRF excited by a laser beam by use of either commercial accessories or fairly simple

Fig. 8.7. Four arrangements for objective-based TIRF in an inverted microscope. In all these configuration, OBJ refers to the objective, SP refers to sample plane, and BFP refers to the objective's back focal plane or its equivalent planes (also called "aperture planes"). Components drawn with heavier lines need to be installed; components in lighter lines are possibly pre-existing in the standard microscope. (**A**) Laser illumination through a side port (requires a special dichroic mirror cube facing the side, available for the Olympus IX-70 microscope). The beam is focused at the back focal plane at a radial position sufficient to lead to supercritical angle propagation into the coverslip. Moving the "focusing lens L transversely changes the angle of incidence at the sample and allows switching between subcritical (EPI) and supercritical (TIR) illumination. This is how Fig. 8.1 was produced. (**B**) Laser illumination in microscope systems containing an equivalent BFP in the rear path normally used by an arc lamp. The laser beam is focused at the BFP where the arc lamp would normally be imaged. The Zeiss Axiovert 200 provides this BFP, marked as an "aperture plane". If (as in the Olympus IX-70) an aperture plane does not exist in the indicated position, it can be created with the pair of lens L1 and L2. (**C**) Laser illumination introduced by an optical fiber through the rear port normally used by the arc lamp. This scheme is employed by the commercial Olympus TIRF device. (**D**) Arc lamp TIR illumination with no laser at all. The goal is to produce a sharp-edged shadow image of an opaque circular disk at the objective back focal plane such that only supercritical light passes through the objective. The actual physical opaque disk (ideally made of aluminized coating on glass) must be positioned at an equivalent upbeam BFP which, in Kohler illumination, also contains a real image of the arc. The Zeiss Axiovert 200 provides this BFP, marked as an "aperture plane". If (as in the Olympus IX-70) an aperture plane does not exist in the indicated position, it can be created with the pair of lens L1 and L2. The illumination at the back focal plane is a circular annulus; it is shown as a point on one side of the optical axis for pictorial clarity only. The through-the-lens arc-lamp TIRF configuration D can be switched easily to laser TIRF configuration C by insertion of the reflecting prism in the arc lamp light path

custom-built add-on modifications (Fig. 8.7A-C). An arc-lamp illumination system, rather than a laser, can also be configured for TIRF illumination by use of an opaque disk of the correct diameter inserted in a plane equivalent (but upbeam) from the objective back focal plane (Fig. 8.7D). This allows only rays at significantly off-axis radii in the back focal plane to propagate through to the TIR sample plane, upon which they are incident at supercritical angles. By placing or removing the opaque disk as shown, illumination can be switched easily between EPI and TIR. Arc illumination has the advantages of easy selection of excitation colors with filters and freedom from coherent light interference fringes, but it is somewhat dimmer because much of the arc lamp power directed toward the sample at subcritical angles is necessarily blocked. Somewhat more of the incident light can be utilized rather than wasted by employing a conical prism in the arc lamp light path [116]. This kind of prism converts the illumination profile, normally brightest in the center, into a dark-centered ring.

Commercial arc-lamp TIRF systems use a curved slit cutout at an equivalent back focal plane. The slit is positioned off-axis to transmit only a thin curved band of light at a radius that will produce supercritical angles. To optimize the throughput,



the brightest part of the Hg arc is focused off-axis right upon the curved slit cutout, with a focusing aberration that curves the arc image to match well with the curved slit.

The evanescent illumination is not "pure" with objective-based TIRF: a small fraction of the illumination of the sample results from excitation light scattered within the objective (discussed in more detail later), and a small fraction of the observed fluorescence arises from luminescence of the objective's internal elements.

Commercial objective-based TIRF systems are available both with laser/optical fiber sources (often multiple ones for quick color switching) and arc lamp sources. They work well, but they are somewhat expensive. A less expensive practical protocol follows for setting up "homemade" objective-based TIRF with free laser beam illumination through an inverted microscope, and verifying that the setup works correctly. Such a homemade setup is also more open and amenable to special modification. The description is most relevant to the setup in Fig. 8.7A, but modifications to the other Fig. 8.7 panels (in which the microscope already has lenses installed in the illumination beam path) are straightforward.

(a) Prepare a test sample consisting of a film of fluorescent material adsorbed to a coverslip of the appropriate type of glass (n = 1.52 for NA = 1.45 or 1.49 objectives; n = 1.78 for NA=1.65 objectives). A convenient and durable uniform film is made from 3,3' dioctadecylindocarbocyanine (also known as "diI", available from Molecular Probes, Eugene OR). Dissolve the diI at about 0.5 mg/ml in ethanol, and place a single droplet of the solution on a glass coverslip. Then,

before it dries, rinse off the coverslip with distilled water. A monolayer of diI fluorophore will remain adhered to the glass. When excited by laser light at 488 nm and observed through the filter set appropriate for fluorescein, the diI adsorbed to the surface will appear orange. Above the diI coated surface, place a bulk layer of aqueous fluorescein solution, which will appear green in the same spectral system. Use enough aqueous fluid so that its upper surface is flat rather than a curved blob. As an alternative to the diI/fluorescein test sample, a suspension of fluorescent microbeads in water can be used. Some of the beads will adhere to the surface. Although the same color as the suspended beads, the surface adhered beads will be immobile while the bulk suspended beads will be jittering with Brownian motion. Place the sample on the microscope stage and raise the objective with the focusing knob to make optical contact through the appropriate immersion oil.

(b) Remove all obstructions between the test sample and the ceiling. Allow a collimated laser beam (the "raw" beam) to enter the side or rear epi-illumination port along the optical axis. A large area of laser illumination will be seen on the ceiling, roughly straight up.

(d) Place a straight edge opaque obstruction (such as an index card) part way in the beam before it enters the microscope. A fuzzy shadow of the card will be seen on the ceiling illumination. By moving the card longitudinally in the right direction, the shadow will become fuzzier. After moving it beyond some point of complete fuzzinesss, it will again become sharper, but now as a shadow on the opposite side of the ceiling illumination. The longitudinal position of maximum fuzziness is a "field diaphragm plane" (FDP), complimentary to the sample plane. Once the location of the FDP is determined, remove the obstruction.

(e) Place a focusing lens (plano- or double-convex,) at that FDP, mounted on a 3D translator. The illuminated region on the ceiling will now be a different size, probably smaller. The goal is to choose a focal length for the focusing lens that almost minimizes the spot size on the ceiling. A focusing lens with close to the proper focal length will require only minor longitudinal position tweaking (i.e., along the axis of the laser beam) to absolutely minimize the illuminated region on the ceiling. At this lens position, the beam focuses at the objective's back focal plane (BFP) and emerges from the objective in a nearly collimated form. This procedure also ensures that the focusing lens is also close to the FDP. Therefore, moving the focusing lens laterally will change the radial position of focus at the BFP (and hence the angle of incidence at the sample plane) but the motion will not affect the centering of the region of illumination in the sample plane.

(f) Fine tune the lateral position of the focusing lens so that the beam spot on the ceiling moves down a wall to the right or left. The inclined path of the beam through the fluorescent aqueous medium in the sample will be obvious to the eye. Continue to adjust the focusing lens lateral position until the beam traverses almost horizontally through the fluorescent medium and then farther, past where

it just disappears. The beam is now totally reflecting at the substrate/aqueous interface.

(g) View the sample through the microscope eyepieces. The diI/fluorescein sample should appear orange, not green, since only the surface is being excited; the microbead sample should show only immobilized dots. Back off the focusing lens to where the beam reappears in the sample (i.e., subcritical incidence). When viewed through the eyepieces, the diI/fluorescein sample should now appear very bright green and the microbead sample very bright with most beads jittering laterally and longitudinally in and out of focus.

(h) If the illuminated region in the field of view is not centered, adjust the lateral position of the "raw" laser beam before it enters the focusing lens.

(i) If the illuminated region is too small, increase the width of the beam before it enters the focusing lens with a beam expander or long focal length diverging lens.

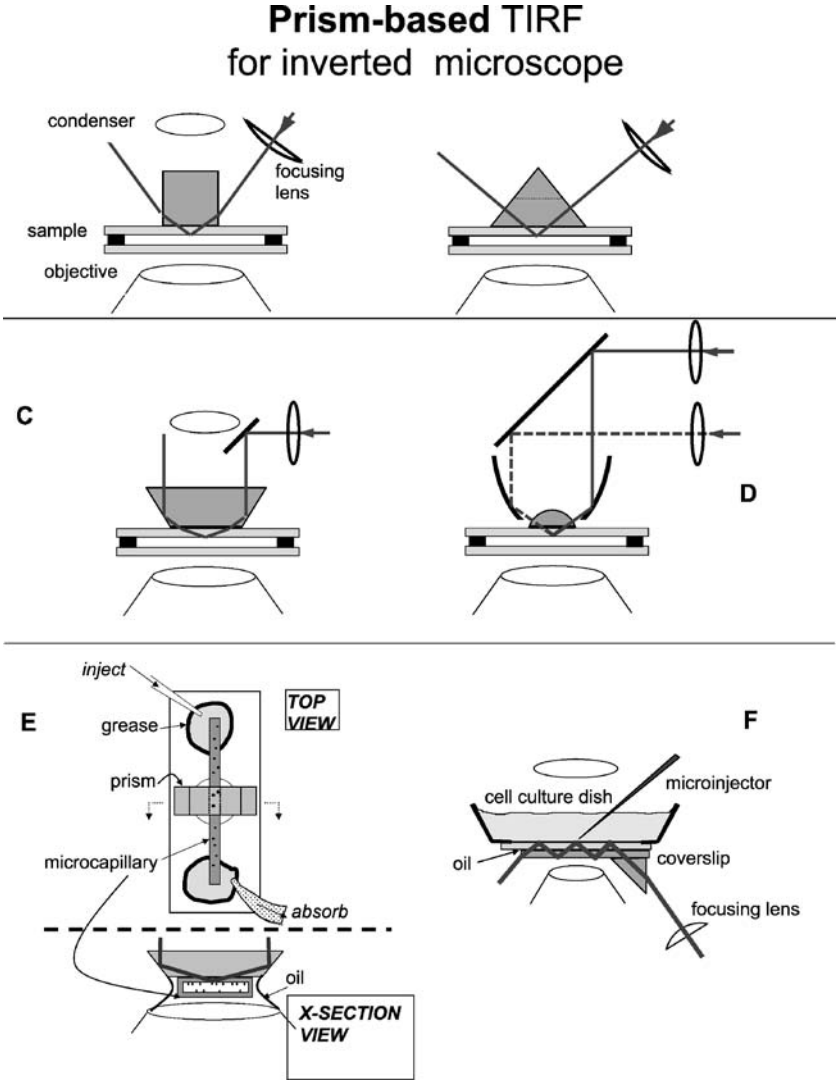
(j) Replace the test sample with the sample of interest and focus the microscope. As the lateral position of the external focusing lens is adjusted through the critical angle position, the onset of TIRF should be obvious as a sudden darkening of the background and a flat two-dimensional look to the features near the surface such that the entire field of view has only one plane of focus.

(k) If interference fringe/stripe illumination TIR is desirable, use a beam splitter and arrange the second beam to enter the focusing lens off-center, parallel to the optic axis, but at a different azimuthal angle around it. A relative azimuthal angle of 180° will give the closest spaced fringes. Be sure that any difference in path length of the two beams from the beam splitter to the focusing lens is less than the coherence length of the laser (a few mm or cm); otherwise, no interference will occur.

8.3.2 TIRF with a Prism

Although a prism may restrict sample accessibility or choice of objectives in some cases, prism-based TIR is very inexpensive to set up and produces a "cleaner" evanescent-excited fluorescence (i.e., less excitation light scattering in the optics) than objective-based TIR. Figure 8.8 shows several schematic drawings designated A-F for setting up laser/prism-based TIR in an inverted microscope.

Figure 8.9 shows an exceptionally convenient (and low cost) prism-based TIRF setup for an upright microscope. The laser beam is introduced in the same port in the microscope base as intended for the transmitted light illuminator (which should be removed), thereby utilizing the microscope's own in-base optics to direct the beam vertically upward. The prism, in the shape of a trapezoid, is mounted on the microscope's condenser mount, in optical contact (through a layer of immersion oil) with the underside of the glass coverslip sample plane. An extra lens just upbeam from the microscope base allows adjustment of the incident beam to totally reflect one of the sloping sides of the prism, from which the beam continues up at an angle toward the sample plane where it is weakly focused. This system gives particularly high-quality images if a water immersion objective is employed and submerged directly



into the buffer solution in an uncovered cell chamber. Sample with cells adhering directly on the bottom of tissue culture plastic dishes rather than on coverslips can also be used; the plastic/cell interface is then the site of TIR. If the objective has a long enough working distance, reasonable accessibility to micropipettes is possible.

In this configuration with the trapezoidal prism, flexibility in incidence angle (to obtain a range of evanescent field depths) is sacrificed in exchange for convenience. However, a set of various-angled trapezoids will allow one to employ various discrete incidence angles. The most inexpensive approach is to start with a commercially available equilateral triangle prism (say 1" x 1" x 1" sides, and 1" length),

Fig. 8.8. Schematic drawings for prism-based TIR in an inverted microscope, all using a laser as a light source. The vertical distances are exaggerated for clarity. The first five configurations (**A**–**E**) use a TIR prism above the sample. In configurations **A**–**D**, the buffer-filled sample chamber sandwich consists of a lower bare glass coverslip, a spacer ring (made of 60 μm thick Teflon or double-stick cellophane tape) and the cell coverslip inverted so the cells face down. The upper surface of the cell coverslip is put in optical contact with the prism lowered from above by a layer of immersion oil or glycerol. The lateral position of the prism is fixed but the sample can be translated while still maintaining optical contact. The lower coverslip can be oversized and the spacer can be cut with gaps so that solutions can be changed by capillary action with entrance and exit ports. In configuration D, two incident beams split from the same laser intersect at the TIR surface, thereby setting up a striped interference pattern on the sample which is useful in studying surface diffusion. Configuration E shows a rectangular cross-section microcapillary tube (Wilmad Glass Co.) instead of a coverslip sandwich. With the ends of the microcapillary tube immersed in droplet-sized buffer baths delimited by silicon grease rings drawn on a support (one for supply and one for draining by absorption into a filter paper tab), very rapid and low volume solution changes during TIRF observation can be accomplished. If an oil immmersion objective is used here, the entire region outside the microcapillary tube between the objective and the prism can be filled with immersion oil. Configuration F places the prism below the sample and depends on multiple internal reflections in the substrate. This configuration thereby allows complete access to the sample from above for solutions changing and/or electrophysiology studies. However, only air or water immersion objectives may be used because oil at the substrate's lower surface will thwart the internal reflections

cleave off and polish one of the vertices to form a trapezoid, and slice the length of the prism to make several copies. Note however, this prism will provide an incidence angle of only 60° at the top surface of the prism. If the prism is made from ordinary $n = 1.52$ glass, that angle is insufficient to achieve TIR at an interface with water. (Recall that we need $(n \sin 60°)$ to exceed 1.33, the refractive index of water). However, equilateral prisms made from flint glass ($n = 1.648$) are commercially available (Rolyn Optics) and these will provide a sufficiently high $(n \sin 60°)$ for TIR to occur at the aqueous interface. In an alternative approach for varying incidence angles over a continuous range, a hemispherical or hemicylindrical prism can be substituted for the trapezoidal prism [112,117,118]. The incident laser beam is directed along a radius line at an angle set by external optical elements.

A commercially available variation of prism-based TIR uses a high aperture condenser fed with laser light from an optical fiber tip instead of a custom-installed prism fed by a free laser beam (Fig. 8.10). This system has the advantage of ready-made convenience and easy variation of both polar and azimuthal incidence angle adjustment, but is limited in maximum incidence angle.

Choice of optical materials for the prism-based methods is somewhat flexible, as follows.

(a) The prism used to couple the light into the system and the (usually disposable) slide or coverslip in which TIR takes place need not be matched exactly in refractive index.

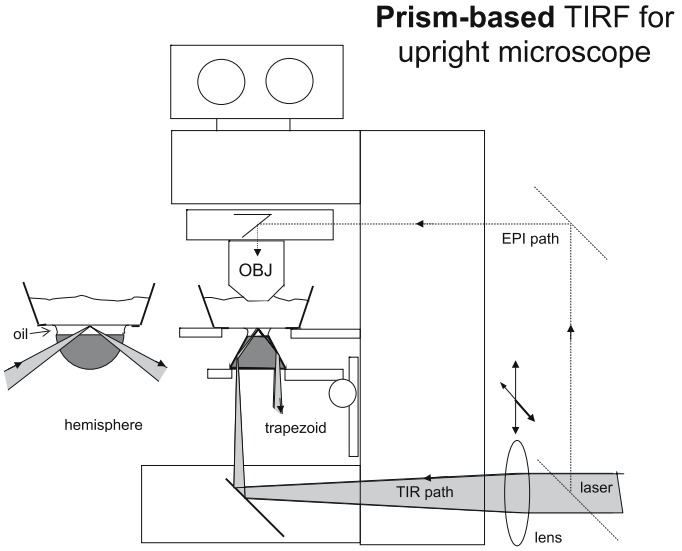


Fig. 8.9. TIRF for an upright microscope utilizing the integral optics in the microscope base and a trapezoidal prism on the vertically movable condenser mount. The position of the beam is adjustable by moving the external lens. An alternative hemispherical prism configuration for variable incidence angle is also indicated to the left. Vertical distances are exaggerated for clarity. An extra set of mirrors can be installed to deflect the beam into an epi-illumination light path (shown with *dashed lines*)

(b) The prism and slide may be optically coupled with glycerol, cyclohexanol, or microscope immersion oil, among other liquids. Immersion oil has a higher refractive index (thereby avoiding possible TIR at the prism/coupling liquid interface at low incidence angles) but it tends to be more autofluorescent (even the "extremely low" fluorescence types).

(c) The prism and slide can both be made of ordinary optical glass for many applications (except as noted above for the configuration in Fig. 8.9) unless shorter penetration depths arising from higher refractive indices are desired. Optical glass does not transmit light below about 310 nm and also has a dim autoluminescence with a long (several hundred microsecond) decay time, which can be a problem in some fluorescence recovery after photobleaching (FRAP) experiments. The autoluminescence of high quality fused silica (often called "quartz") is much lower. Tissue culture dish plastic (particularly convenient as a substrate in the upright microscope setup) is also suitable, but tends to have a significant autofluorescence compared to ordinary glass. More exotic high n_3 materials such as sapphire, titanium dioxide, and strontium titanate (with n as high as 2.4) can yield exponential decay depths d as low as $\lambda_o/20$, as can be calculated from (8.4).

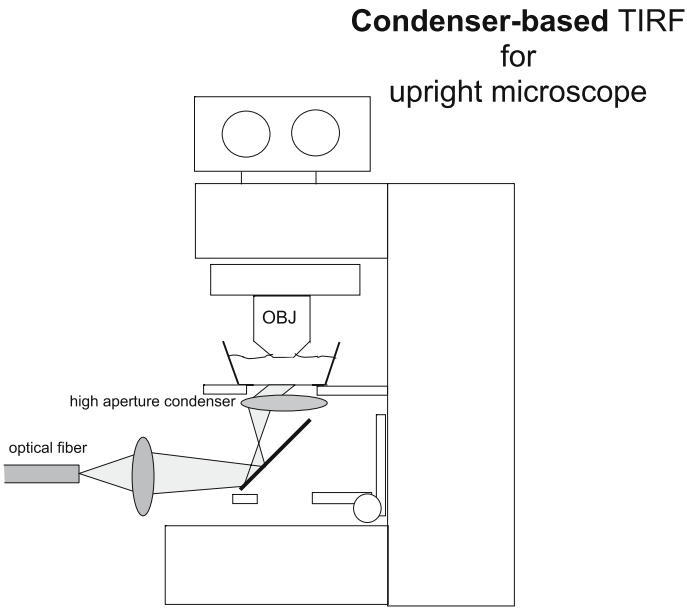


Fig. 8.10. TIRF for an upright microscope in which the prism is replaced by a high-aperture condenser, used in conjunction with a fiber optic bundle. This system is commercially available

In all the prism-based methods, the TIRF spot should be focused to a width no larger than the field of view; the larger the spot, the more that spurious scattering and out-of-focus fluorescence from the immersion oil layer between the prism and coverslip will increase the generally very low fluorescence background attainable by TIRF. Also, the incidence angle should exceed the critical angle by at least a couple of degrees. At incidence angles very near the critical angle, the cells cast a noticeable "shadow" along the surface.

Here is a practical protocol for setting up "homemade" prism-based TIRF, by far the least expensive approach to TIR.

(a) Mount the prism on the condenser mount carrier if possible.
(b) Depending on the configuration, a system of mirrors with adjustable angle mounts fixed to the table must be used to direct the beam toward the prism. One of these mirrors (or a system of shutters) should be movable and placed near the microscope so switching between standard epi-illumination and TIR is possible without interrupting viewing.
(c) Place a test sample (e.g., a diI-coated coverslip, see the description in the objective-based TIR section) in the same kind of sample holder as to be used for cell experiments. An aqueous medium, possibly containing some fluorescein, can be used to fill the sample chamber. Alternatively, a test sample of fluorescent microbeads can be used as previously described.

(d) With the test sample on the stage, focus on the fluorescent surface with transmitted (usually tungsten) illumination. Usually, dust and defects can be seen well enough to assay the focus. Fluorescent epi-illumination can also be used to find the focus because only at the focal position are laser interference fringes seen sharply.
(e) Place a small droplet of immersion oil on the non-diI surface of the sample coverslip or directly on the prism (depending on which one faces upward in the chosen configuration) and carefully translate the prism vertically so it touches and spreads the oil but does not squeeze it so tightly that lateral sliding motion is inhibited. Too much oil will bead up around the edges of the prism and possibly interfere with the illumination path.
(f) Without any focusing lens in place, adjust the unfocused ("raw") collimated laser beam position with the mirrors so that TIR occurs directly in line with the objective's optical axis. This can usually be seen by observing the scattering of the laser light as it traverses through the prism, oil, and TIR surface.
(g) Insert the focusing lens so that the focus is roughly at the TIR area under observation. Adjust its lateral position with translators on the focusing lens so that the TIR region occurs directly in line with the objective. To guide this adjustment, look for three closely aligned spots of scattered light, corresponding to where the focused beam first crosses the immersion oil layer, where it totally reflects off the sample surface, and where it exits by recrossing the oil.
(h) The TIR region should now be positioned well enough to appear in view in the microscope when viewed as test sample fluorescence with the standard filters in place. In general, the TIR region will appear as a yellow-orange ellipse or streak with a diI test sample and a region of discrete tiny bright dots with a microbead test sample. Make final adjustments with the focusing lens to center this area. The TIR area can be distinguished from two out-of-focus blurs past either end of the ellipse or streak (arising from autofluorescence of the immersion oil) because the TIR spot contains sharply focused images of defects in the diI coating or sharp dots of adsorbed microbeads. The focusing lens can be moved forward or backward along the laser optical path to achieve the desired size of the TIR area. If fluorescein solution was used to fill the sample chamber, the characteristic green color of fluorescein should *not* be seen with successful TIR. If the alignment is not correct and propagating light reaches the solution, then a bright featureless blur of green will be seen.
(i) With the optics now correctly aligned for TIR, translate the prism vertically to remove the diI sample, and replace it with the actual cell sample. Relower the prism to make optical contact. Although the TIR region will not be exactly in the same spot (because of irreproducibility in the prism height), it will be close enough to make final adjustments with the focusing lens while observing fluorescence from the cell sample.

8.3.3 TIR from Multiple Directions

Configurations involving a single laser beam that produces TIR from one incidence direction (i.e., from one azimuthal angle around the optical axis) are the simplest ones to set up, but configurations involving multiple or a continuous range of azimuthal angles incident on the sample plane have some unique advantages. A single illumination direction tends to produce shadows on the "downstream" side of cells because of evanescent field scattering by the cells; these shadows are less apparent at higher incidence angles even with a single beam. However, a hollow cone of illumination over all the azimuthal angles around the optical axis virtually eliminates this shadow artifact. The objective-based TIR configuration that uses an arc lamp for illumination system (Fig. 8.7D) automatically provides such a hollow cone. A hollow cone can also be produced from a laser beam by positioning a conical lens in the incident beam path exactly concentric with the laser beam, in a similar fashion as already described for arc illumination [116].

Illumination by two mutually coherent TIR laser beams produces a striped interference fringe pattern in the evanescent field intensity in the sample plane. For a relative azimuthal angle of ϕ between the two beams, both with an incidence angle of θ, the node-to-node spacing s of the fringes is given by:

$$s = \lambda_o / [2n_3 \sin\theta \sin(\phi/2)] \tag{8.18}$$

The spacing s is not dependent upon the refractive index n_1 of the medium (or cell). For beams coming from opposite azimuthal directions ($\phi = \pi$), $s = \lambda_o/2$ at the critical angle of incidence, and $s = \lambda_o/2n_3$ at glancing incidence ($\theta = \pi/2$). For a typical glass substrate with $n_3 = 1.5$, this latter spacing is smaller than the Raleigh resolution limit of the microscope and can barely be discerned by the imaging system although the fringes do exist physically at full contrast.

These interference stripes can be used in combination with TIR-FRAP (see above) to bleach anti-node regions but not node regions. The subsequent fluorescence recovery will then provide information about surface diffusion of fluorophores [53,57,103,119].

The spatial phase of the node/antinode intensity stripe pattern can be controlled easily by retarding the phase of one of the interfering beams with an extra glass plate inserted into its light path. By illuminating with four discrete coherent TIR beams at relative azimuthal angles of $\phi = 0, \pi/2, \pi$ and $3\pi/2$ a checkerboard evanescent pattern in both the x and y lateral directions (i.e., the sample plane) can be produced. Then by imaging a sample at lateral node/antinode spatial phase steps of 0°, 120°, and 240°, images with super-resolution well below the Raleigh limit in the sample plane can be computed [120,121]. Such "structured illumination" by stripes or checkerboard can be produced in standard epi-illumination with a grid at the field diaphragm plane. But with interfering-beam TIRF, the spacing of the stripes can be made smaller and thereby produce unprecedented lateral resolution, down to about 0.13 μm for 514 nm wavelength illumination.

More than four uniformly separated simultaneous azimuthal angles will produce a bright central spot with regular spaced off-center bands. In the limit of an infinite

number of azimuthal angles (i.e., a coherent ring), a small very bright central spot of sub-wavelength dimension and with increasingly dim rings at large radii can be produced. Coherent ring illumination can be produced by illuminating a broad circular band [102] at the objective back focal plane (BFP) which gives rise to a range of supercritical incidence angles, or by focusing the illumination into a narrow single radius at the BFP. In the latter case, coherent ring TIR illumination provides a way of illuminating a very small lateral area on the TIR surface while maintaining a single supercritical polar incidence angle. The pattern and size of the central spot will depend strongly on the polarization of the thin incident azimuthal ring. The smallest attainable spot on the TIR surface requires radial polarization at the BFP: it has a radius from the central maximum to the first minimum of only $0.26\lambda_o$, which is considerably less than the Raleigh resolution limit for a 1.49NA objective ($0.41\lambda_o$). Figure 8.11 shows the patterns that should appear on the sample when the BFP is illuminated in a thin ring at a single radius with the indicated polarization, based on theoretical calculations of circular evanescent wavefronts of isotropic amplitude converging to a central point [123]

In general, incident light approaching the interface from a single angle cannot produce a small illumination region. For this goal, converging illumination from a large range of angles is needed. However, the angular range must not include subcritical polar angles in order to preserve TIR. The ring illumination discussed above creates an appropriate range azimuthally. A variation of ring illumination, a system called "confocal TIR", utilizes a range of polar incidence angles [124,125]. To ensure these polar angles are confined to the supercritical range, an opaque disk installed into a custom-made parabolic glass reflection objective blocks subcritical rays. In that system, only high polar angle emitted light can pass that opaque disk, but the fluorescence rays that do so are focused onto a small pinhole, behind which resides a non-imaging detector (as in standard confocal systems). An image is constructed by scanning the sample. The advantage of this system is that only a very small volume is illuminated at one time, and the effective lateral resolution appears at least as good as standard epi imaging. It is likely that this sort of ring illumination could also be produced through a commercially available high aperture objective by placing the opaque disk farther "upstream" at an aperture plane [115] and thereby avoid blocking any emitted rays.

8.3.4 Rapid Chopping between TIR and EPI

Regardless of the method chosen to produce TIR in a microscope, it is often desirable to switch rapidly between illumination of the surface (by TIR) and deeper illumination of the bulk (by standard epi-fluorescence). For example, a transient process may involve simultaneous but somewhat different fluorescence changes in both the submembrane and the cytoplasm, and both must be followed on the same cell in response to some stimulus [72]. For spatial resolved images, the switching rate is often limited by the readout speed of the digital camera or by photon shot noise. Single channel integrated intensity readings over a region defined by an image plane

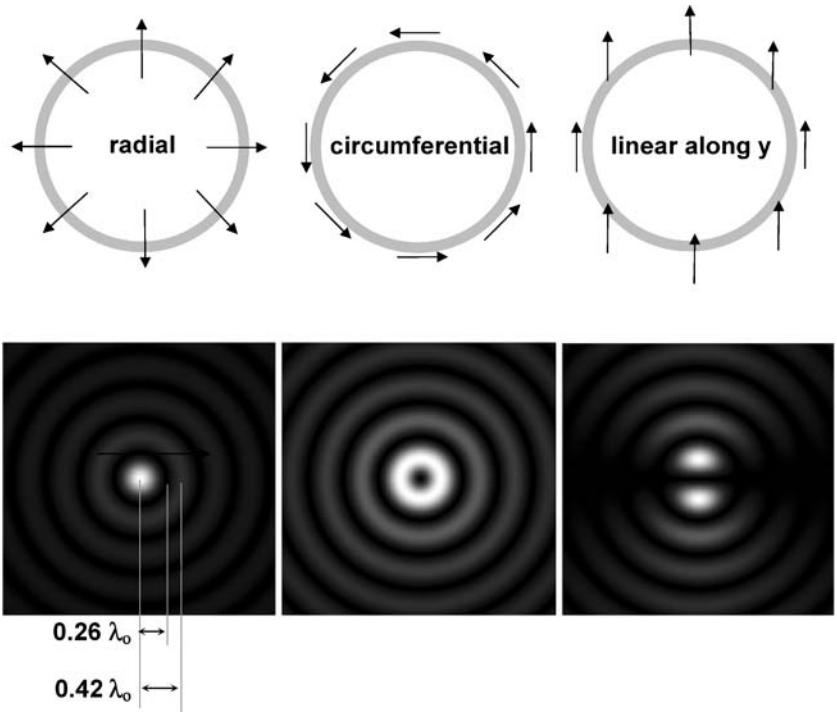


Fig. 8.11. Theoretical intensity patterns that should result at the TIR surface from coherent ring illumination at a single radius at the BFP. The pattern depends strongly on the polarization at the back focal plane; three experimentally attainable possibilities are shown. The resulting polarization at the TIR surface is mixed. In the case of radial BFP polarization, the TIR polarization is about l90% (depending on polar incidence angle) in z-direction (i.e., along optic axis, normal to this page); the *z*-polarized intensity shown. In the case of circumferential BFP polarization, the TIR polarization is purely circumferential; that is the polarization shown. In the case of linear (along *y*) BFP polarization, there is significant intensity in *x*-,*y*-, and *z*-polarizations. Of the three, the *z*-polarization has the most confined illumination area and that is the one shown here

diaphragm can be performed much more rapidly. Figure 8.12 shows a method using computer driven acousto-optic modulators by which very rapid chopping (with a switching time as small as ten microseconds) can be done.

8.3.5 Surface Near-Field Imaging

As discussed in the Theoretical Principles section, the technique for surface selective emission imaging does not depend on TIR excitation at all, and in fact it works with standard epi-illumination and, in principle, should even work with non-optical excitation such as chemiluminescence. Near-field light from surface-proximal fluorophores propagates into the glass exclusively at supercritical hollow cone angles.

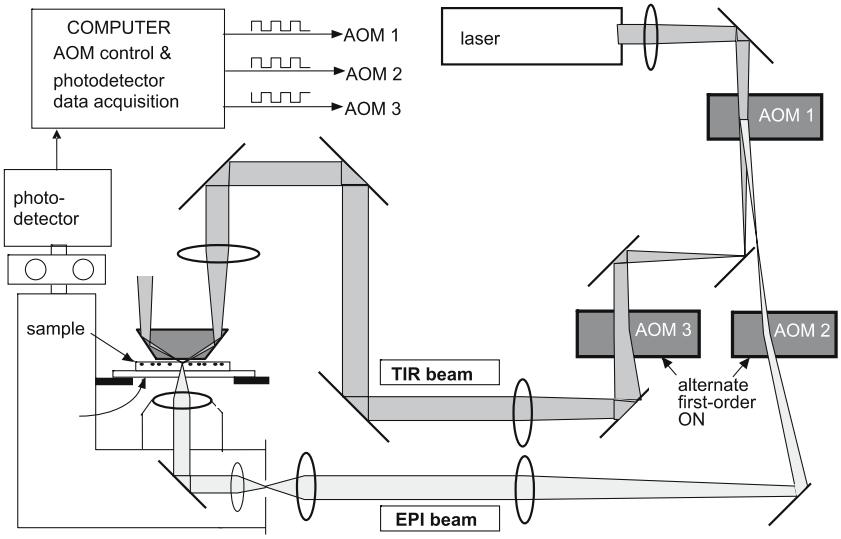


Fig. 8.12. Rapid alternation between TIR and EPI illumination. Three acousto-optic modulators (AOM) are used, with driving voltage phases as shown. The resulting chopped output of the single-channel photodetector can be separated into two traces vs.time by software

Figure 8.13 shows a method for blocking all subcritical light emanating from the fluorophore far-field while passing a range of angles of supercritical light from the near-field by use of an opaque disk at a back focal plane in the emission path. Any one particular supercritical hollow cone angle of emission corresponds to a single exponential near-field decay, but typically a wider range of supercritical angles are collected by the objective and passed outside the periphery of the opaque disk. The resulting surface selectivity corresponds to a weighted integral of exponentials with a range of decay lengths. An actual implementation of this modification has been demonstrated to work on both artificial systems of surface-bound fluorophores and on living biological cells, albeit with a loss of some lateral resolution [115].

8.4 General Experimental Considerations

Laser source. A laser with a total visible output in the 100 mW or greater range is more than adequate for most TIRF applications. Air cooled argon or diode lasers in the 10 mW to 100 mW are also usually adequate, but possibly marginal for dim samples or for samples where a weaker laser line (e.g., the 457 nm line of argon) may be desired to excite a shorter wavelength fluorescent marker (such as cyan fluorescent protein, CFP).

Laser interference fringes. Laser illumination can produce annoying interference fringes because optical imperfections in the beam path can shift the local phase of coherent light. For critical applications, several approaches can alleviate

the problem. One approach employs an optical fiber bundle in which individual fiber-to-fiber length differences exceed the laser light's coherence length. This produces an extended source at the output tip with point-to-point relative phases that are temporally randomized and defocused on the sample [126]. This elegant system, with no moving parts, produces a speckle pattern that changes so rapidly that it appears uniform down to the nanosecond time scale. Commercially available mode scramblers and rapid flexing of optical fibers may also reduce some fringing.

Another set of methods uses a free laser beam rather than fibers. For example, a finely frosted glass surface or a plastic plate (such as tissue culture dish bottom), spinning laterally to the laser beam, temporally randomizes the phase and produces a fringe pattern that fluctuates and can be averaged over the duration of a camera exposure [8].

Interference fringes can be effectively suppressed by spinning the azimuthal angle of TIR incidence with electro-optical or mechano-optical devices such that the incident beam traces a circle where it focuses at the objective's BFP. Since only one azimuthal direction illuminates the sample at any instant but the spinning period is much shorter than the camera's exposure time or retinal image retention time, the interference fringes are superimposed in their intensities (rather than in their electric fields), giving the illusion of a much more uniform illumination field [127]. An example of the fringe suppression effect is shown in Fig. 8.14.

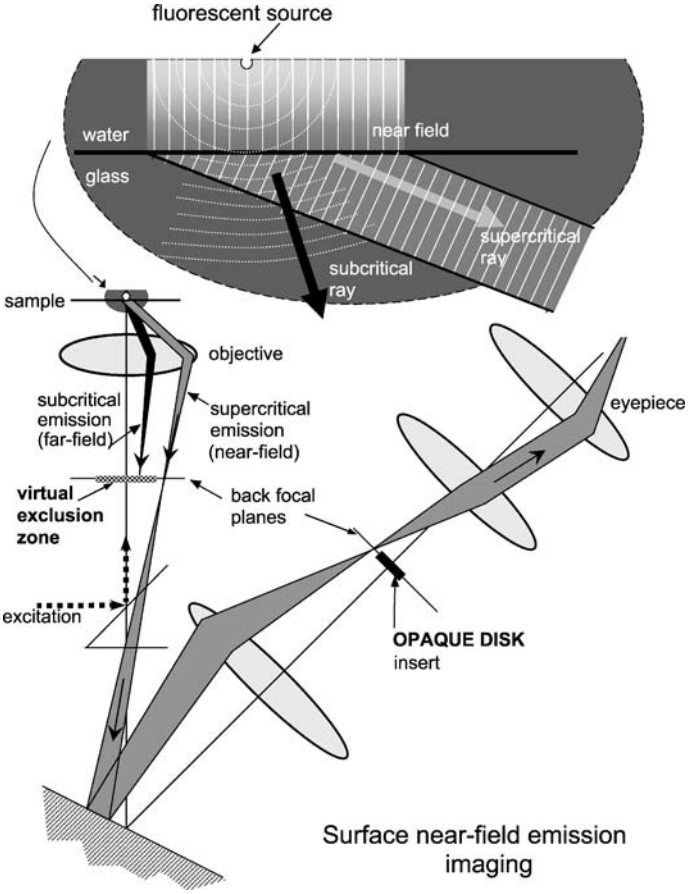
An alternative non-optical approach to suppressing fringes computes a normalization of sample digital images against a control digital image of a uniform concentration of fluorophores. This works well only if the fringes were not induced by the sample itself.

Functionalized substrates. TIRF experiments often involve specially coated substrates. A glass surface can be chemically derivatized to yield special physi- or chemi-absorptive properties. Covalent attachment of certain specific chemicals are particularly useful in cell biology and biophysics, including: poly-l-lysine for enhanced adherence of cells; hydrocarbon chains for hydrophobicizing the surface in preparation for lipid monolayer adsorption; and antibodies, antigens, or lectins for producing specific reactivities. Derivatization generally involves pretreatment of the glass by an organosilane [57].

A planar phospholipid coating (possibly with incorporated proteins) on glass can be used as a model of a biological membrane. Methods for preparing such model membranes on planar surfaces suitable for TIR have been reviewed [128].

Aluminum coating (for surface fluorescence quenching) can be accomplished in a standard vacuum evaporator; the amount of deposition can be made reproducible by completely evaporating a premeasured constant amount of aluminum. After deposition, the upper surface of the aluminum film spontaneously oxidizes in air very rapidly. This aluminum oxide layer appears to have some similar chemical properties to the silicon dioxide of a glass surface; it can be derivatized by organosilanes in much the same manner [57].

Photochemistry at the surface. Illumination of surface adsorbed proteins can lead to apparent photochemically-induced crosslinking. This effect is observed as



a slow, continual, illumination-dependent increase in the observed fluorescence. It can be inhibited by deoxygenation (aided by the use of an O_2-consuming enzyme/substrate system such as protocatachuic deoxygenase/protocatachuic acid or a glucose/glucose oxidase system), or by 0.05 M cysteamine [57].

Actual evanescent field depth and profile. The evanescent field characteristic depth in an actual sample may deviate from that predicted by Eq. (8.4), even if the incidence angle θ is well measured, for several reasons. Depth d depends on the refractive index of the liquid sample above the TIR surface, and this is not well known for samples such as cell cytoplasm. So far, no good way of accurately measuring the local evanescent field depth and profile in an actual heterogeneous sample has been proposed.

Measurements of evanescent depth in a homogeneous artificial sample with a refractive index that approximated the average in a heterogeneous sample can be done. In such a sample, the characteristic time for diffusion of fluorescent beads through the evanescent field (possibly by TIR/FRAP or TIR/FCS) can be measured

Fig. 8.13. Schematic diagram of the surface-selective near-field emission in an Olympus IX-70 microscope. At the top, a magnified view of the sample region shows the electric field of the near-field corresponding to one particular k_z as an exponential decay (shown with graded shading) extending in the z-direction away from a single fluorophore (shown as a *white dot*). The lateral spacing (left-to-right here) of the wavefronts for that one contribution to the near-field is shown in a series of *vertical stripes*. If the fluorophore is close enough to the glass, the tails of the decaying near-field wavefronts are captured by the glass and converted into light propagating away at a particular supercritical angle. (A complete picture of the near-field would show a continuous spectrum of decay lengths, each corresponding to different wavefront spacings, with shorter decay lengths associated with closer wavefront spacings. The total near field is a superposition of all such different k_z contributions, with their relative wavefronts all aligned exactly in phase at the lateral position of the fluorophore so that the region laterally nearest the fluorophore is the brightest with a rapid decay to either side. Each different k_z contribution gives rise to light in the glass propagating at a different supercritical angle.) Far-field light is shown as a series of circular wavefronts emanating from the fluorophore and refracting into the glass at subcritical angles only. The *lower part* of the figure shows the fate of the super- and subcritical light rays in the microscope. An opaque disk is inserted in an accessible region at a plane complimentary to the objective back focal plane, concentric with the optical axis and with a radius just sufficient to block out all the subcritical light. The correct radius of the opaque disk can be determined by observing this plane directly (with the eyepiece removed) with a diI/fluorescein preparation (as described in the objective-based TIR section) on the sample stage; the subcritical light will appear green, surrounded by an orange annulus. When placed at the correct longitudinal position along the optical axis, the opaque disk will show no parallax relative to the orange annulus. The emission light forms a hollow cylindrically symmetric cone around the optical axis after the opaque disk; emission light is shown traversing one off-axis location for pictorial clarity only. Subcritical light is actually blocked only at the real opaque disk; the diagram shows the blockage at the image of the disk at the objective's back focal plane as a "virtual exclusion zone" to illustrate the principle

and (given a known diffusion coefficient) converted to a corresponding characteristic distance. However, this method gives only a single number for effective depth but does not provide unambiguous information about the possibly non-exponential intensity profile as a function of z.

In samples with heterogeneous refractive index, regions with higher index may not support TIR, whereas lower index regions do. Clearly, it is best to illuminate with as high an incidence angle as possible to be sure that all regions support TIR. Even if the incidence angle is sufficiently high to support TIR everywhere, irregularities and heterogeneities in the sample give rise to scattering of the evanescent field. This scattered light can excite fluorescence as it penetrates much more deeply into the liquid than does the evanescent light. Figure 8.15 shows an example of this effect, with glass beads in an aqueous fluorescein solution, under TIR illumination. The beads capture evanescent light and focus it into bright propagating light "comets" on the downbeam side of the beads' equatorial planes, as can be visualized by the higher fluorescence in those regions. The effect is much less pronounced for higher TIR incidence angles.

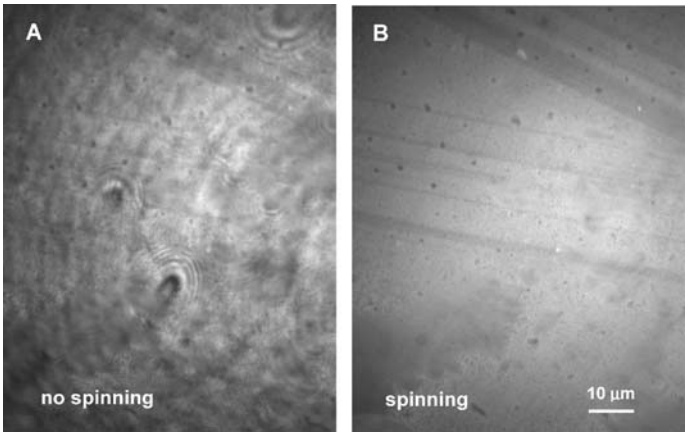


Fig. 8.14. Effect of azimuthal spinning on TIRF images of diI adsorbed to a glass coverslip. (**A**) Sample illuminated with a single azimuthal incidence angle; i.e., not spinning. Interference fringes are very evident. (**B**) Same field of view with wedge spinning. The laser interference fringes are no longer visible. The remaining non-uniformities are due to features of the sample and also the fading of illumination intensity toward the corners

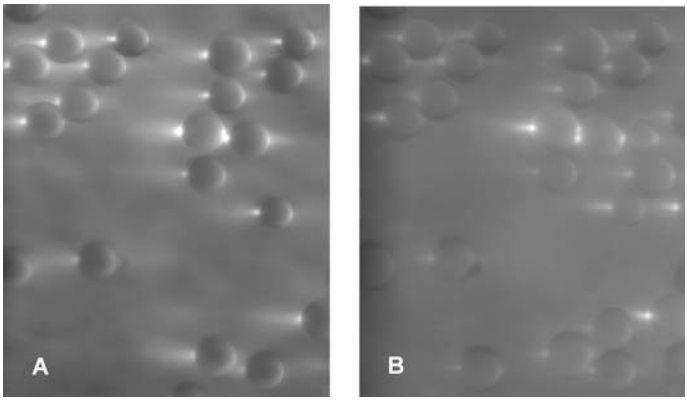


Fig. 8.15. Effect of increasing incidence angle on evanescent field scattering. High n (= 1.5) beads of 10 μm diameter are settled onto the TIR surface in an aqueous medium containing fluorescein. (**A**) The incidence angle is about 1° greater than the critical angle. The beads clearly convert some of the local evanescent field into propagating light, creating beacons and shadows immediately "downstream". (**B**) These effects are reduced with a thinner evanescent field created by an incidence angles several degrees higher. In both images, the microscopic focus is at the bead equatorial plane, not at the TIR surface

The TIR optics itself, particularly the high aperture objectives used in objective-based illumination, can produce scattered light. Actual intensity profiles arising from objective-based TIR can be obtained by at least a couple of methods. In one method, a quantum dot can be attached to the tip of an atomic force microscopy

(AFM) cantilever probe mounted onto a z-translator. In this way, intensity can be measured directly as a function of the quantum dot z-position [129]. One problem with this approach (aside from the necessity of using AFM equipment) is that the probe and quantum dot themselves can disrupt the evanescent field and produce scattering.

Another approach is to observe a large (~8 μm diameter) spherical bead with a refractive index matched to the surrounding liquid to avoid disruption and scattering. The bead can be fluorescence-labeled on its surface. At the image of the point of contact between the bead and the TIR substrate, the fluorescence is the brightest. It becomes dimmer with increasing distance in the TIR plane from the contact point. Simple geometry then allows calculation of the fluorescence vs. z. In an actual test of this method in 1.45NA and 1.65NA objective-based TIRF [130], the evanescent field was found to account for about 90% of the intensity at $z = 0$, and the exponential decay depth agreed well with expectations based on angle and refractive indicies. However, a much slower decaying component, presumably from scattering within the objective, was also evident, and this component became dominant in the distant $z > 2d$ zone.

Apart from this dim scattering component, bright collimated shafts of scattered light, emanating from the edge of the top lens of the objective, can be seen in objective-based TIR. Fortunately, however, these shafts do not cross the field of view so they do not excite spurious fluorescence in the image.

8.5 TIRF vs. other Optical Section Microscopies

Confocal microscopy achieves axial selectivity by exclusion of out-of-focus emitted light with a set of image plane pinholes. Confocal microscopy has the clear advantage in versatility; its method of optical sectioning works at any plane of the sample, not just at an interface between dissimilar refractive indices. However, other differences exist which, in some special applications, can favor the use of TIRF.

(a) The depth of the optical section in TIRF is typically ≤ 0.1 μm whereas in confocal microscopy it is a relatively thick ~0.6 μm.
(b) In some applications (e.g., FRAP, FCS, or on cells whose viability is damaged by light), illumination and not just detected emission is best restricted to a thin section; this is possible only with TIRF.
(c) Since TIRF can be adapted to and made interchangeable with existing standard microscope optics, even with "homemade" components, it is much less expensive than confocal microscopy. Laser-based and arc-based TIRF microscopy kits are now available from most of the major microscope manufacturers and third-party suppliers.
(d) Unlike confocal microscopy, TIRF is suitable not only for microscopic samples but also for macroscopic applications; in fact those were the first TIRF studies. A previous review covers much of that work [131].

Two-photon (or more generally, multiphoton) microcopy has many desirable features, including true optical sectioning, whereby the plane of interest is the only one that is actually excited, as in TIRF. Multiphoton microscopy is not restricted to the proximity of an interface, but its optical section depth is still several times thicker than that of TIRF. The setup expense of multiphoton microscopy for an infrared laser with sufficient pulse peak intensity can also be a consideration. Both multiphoton and confocal microscopy are necessarily scanning techniques; TIRF microscopy is a "wide field" technique and is thereby not limited in speed by the scanning system hardware or image reconstruction software.

Cell-substrate contacts can be located by a nonfluorescence technique completely distinct from TIRF, known as "interference reflection contrast" (IRC) microscopy. Using conventional illumination sources, IRM visualizes cell-substrate contacts as dark regions. Internal reflection microscopy has the advantage that it doesn't require the cells to be labeled, but the disadvantages that it contains no information of biochemical specificities in the contact regions and that it is less sensitive to changes in contact distance (relative to TIRF) within the critical first 100 nm of the surface.

Although TIRF microscopy became widely used only in the 1990's and increasingly so thereafter, well after the introduction and popularity of confocal microscopy, it actually had much earlier origins in surface spectroscopy. TIR in a nonmicroscopic nonfluorescent form had long been used for visible and infrared absorption spectroscopy [132]. Nonmicroscopic TIR combined with visible fluorescence spectroscopy at various kinds of interfaces also has a long history, beginning in the early 1960's [133–138]. The first application of TIR to microscopy was nonfluorescent, describing evanescent light scattering from cells [139] in 1961. The combination of all three elements – TIR and fluorescence and microscopy – was presented first in 1981 [2].

Acknowledgements

This work was supported by NIH grant 5 R01 NS38129. The author wishes to thank Drs. Ron Holz and Edwin Levitan for the chromaffin cells and the GFP construct, respectively, depicted in Fig. 8.1, and Dr. Geneva Omann for reviewing a draft of this chapter. The author also thanks all the coworkers who have contributed to aspects of the TIRF work described here: Ron Holz, Thomas Burghardt, Nancy Thompson, Edward Hellen, Ariane Mc Kiernan, Andrea Stout, Michelle Dong Wang, Robert Fulbright, Laura Johns, Susan Sund, Miriam Allersma, Mary Bittner, Alexa Mattheyses, Keith Shaw, and Geneva Omann.

References

1. H. Schneckenburger: Curr. Opin. Biotech. 16,13 (2005)
2. D. Axelrod: J. Cell Biol. 89, 141 (1981)
3. R.M. Weis, K. Balakrishnan, B. Smith, H.M. McConnell: J Biol Chem. 257, 6440 (1982)
4. M. Cohen, Z. Kam, L. Addad1, B. Geiger: EMBO J. 25, 302 (2006)
5. D. Gingell, O.S. Heaven, J.S. Mellor: J. Cell Sci. 87, 677 (1987).
6. I. Todd, J.S. Mellor, D. Gingell: J. Cell Sci. 89, 107 (1988)
7. T. Lang, I. Wacker, I. Wunderlich, A. Rohrbach, G. Giese, T. Soldati, W. Almers: Biophys. J. 78, 2863 (2000)
8. J.R. Kuhn, T.D. Pollard: Biophys. J. 88, 1387 (2005)
9. M.C. Adams, A. Matov, D. Yarar, S.L. Gupton, G. Danuser, C.M. Waterman-Storer: J. Microsc-Oxford 216, 138 (2004)
10. R.D. Vale, T. Funatsu, D.W. Pierce, L. Romberg, Y. Harada, T. Yanagida: Nature 380, 451 (1996)
11. S. Khan, D. Pierce, R.D. Vale: Curr. Biol. 10, 927 (2000)
12. R.M. Dickson, D.J. Norris, Y-L Tzeng, W.E. Moerner: Science 274, 966 (1996)
13. R.M. Dickson, D.J. Norris, W.E. Moerner: Phys. Rev. Lett. 81, 5322 (1998)
14. Y. Sako, S. Miniguchi, T. Yanagida: Nature Cell Biol. 2, 168 (2000)
15. A.E. Knight, J.E. Molloy: in "Molecular Motors, Essays in Biochemistry, Vol 35" (S.J. Higgins and G. Banting, eds.), p. 200. Portland Press, London, 2000.
16. T.J. Ha, A.Y. Ting, J. Liang, W.B. Caldwell, A.A. Deniz, D.S. Chemla, P.G. Schultz, S. Weiss: Proc Natl Acad Sci USA 96, 893 (1999)
17. A. Graneli, C.C. Yeykal, T.K. Prasad, E.C. Greene: Langmuir 22, 292 (2006)
18. T. Wazawa, M. Ueda: Microsc. Techniques 95, 77 (2005)
19. H. Yokota, K. Kaseda, H. Matsuura, Y. Arai, A. Iwane, Y. Ishii, T. Kodama, T. Yanagida: J. Nanosci. Nanotech. 4, 616 (2004)
20. J. Ichinose, Y. Sako: Trends Anal. Chem. 23, 587 (2004)
21. M.A. Osborne: J. Phys. Chem. B 109, 18153 (2005)
22. J.N. Forkey, M.E. Quinlan, Y.E. Goldman: Biophys. J. 89,1261 (2005)
23. C. Wang, L. Liu, G.Y. Wang, Z.Z. Xu: Chinese Phys. Lett 21, 843 (2004)
24. T.E. Starr, N.L. Thompson: Biophys. J. 80, 1575 (2001)
25. K. Hassler, M. Leutenegger, P. Rigler, R. Rao, R. Rigler, M. Gosch, T. Lasser: Optics Express 13, 7415 (2005)
26. A.Y. Kobitski, C.D. Heyes, G.U. Nienhaus: Appl. Surf. Sci. 234, 86 (2004)
27. T. Lang, I. Wacker, J. Steyer, C. Kaether, I. Wunderlich, T. Soldati, H-H. Gerdes, W. Almers: Neuron 18, 857 (1997)
28. J.A. Steyer, W. Almers: Biophys. J. 76, 2262 (1999)
29. D. Toomre, J.A. Steyer, P. Keller, W. Almers, K. Simons: J. Cell Biol. 149, 33 (2000)
30. D.P. Zenisek, J.A. Steyer, W. Almers: Biophys. J 78, 1538 (2000)
31. J.A. Steyer, W. Almers: Nature Rev. Mol. Cell Biol. 2, 268 (2001)
32. T. Lang, D. Bruns, D. Wenzel, D. Riedel, P. Holroyd, C. Thiel, R. Jahn, EMBO J. 20, 2202 (2001)
33. M. Oheim, D. Loerke, W. Stuhmer, R.H. Chow: Eur. Biophys. J. 27, 83 (1998)
34. M. Oheim, D. Loerke, W. Stuhmer, R.H. Chow: Eur. Biophys. J. 28, 91 (1999)
35. M. Oheim, W. Stuhmer: J. Memb. Biol. 178, 163 (2000)
36. W. Han,Y-K. Ng, D. Axelrod, E.S. Levitan: Proc Natl Acad Sci USA 96, 14577 (1999)
37. J. Schmoranzer, M. Goulian, D. Axelrod, and S.M. Simon: J. Cell Biol. 149, 23 (2000)

38. L.M. Johns, E.S. Levitan, E.A. Shelden, R.W. Holz, D. Axelrod: J. Cell Biol. 153, 177 (2001)
39. T. Tsuboi, C. Zhao, S. Terakawa, G.A. Rutter: Curr. Biol. 10, 1307 (2000)
40. T. Tsuboi, T. Kikuta, A. Warashina, S. Terakawa: Biochem. Biophys. Res. Comm. 282, 621 (2001)
41. A. Rohrbach: Biophys. J. 78, 2641 (2000)
42. D. Toomre, D.J. Manstein: Trends Cell Biol 11, 298 (2001)
43. M.W. Allersma, L. Wang, D. Axelrod, R.W. Holz: Molec. Biol. Cell 15, 4658 (2004)
44. M.W. Allersma, M.A. Bittner, D. Axelrod, R.W. Holz: Molec. Biol. Cell, *in press* (2006)
45. D. Perrais, I.C. Kleppe, J.W. Taraska, W. Almers: J. Physiol 560, 413 (2004)
46. J.K. Jaiswal, S. Chakrabarti, N.W. Andrews, S.M. Simon: Plos. Biol. 2, 1224 (2004)
47. J.W. Taraska, W. Almers: Proc. Natl. Acad. Sci. USA 101, 8780 (2004)
48. M. Fix, T.J. Melia, J.K. Jaiswal, J.Z. Rappoport, D.Q. You, T.H. Sollner, J.E. Rothman, S.M. Simon: Proc. Natl. Acad. Sci. USA 101, 7311 (2004)
49. J. Schmoranzer, G. Kreitzer, S.M. Simon: J. Cell Sci. 116, 4513 (2003)
50. M.A. Silverman, S. Johnson, D. Gurkins, M. Farmer, J.E. Lochner, P. Rosa, B.A. Scalettar: J. Neurosci 25, 3095 (2005)
51. B.A. Scalettar, P. Rosa, E. Taverna, M. Francolini, T. Tsuboi, S. Terekawa, S. Koizumi, J. Roder, A. Jeromin: J. Cell Sci. 115, 2399 (2002)
52. T. Tsuboi, S. Terekawa, B.A. Scalettar, C. Fantus, J. Roder, A. Jeromin: J. Biol. Chem. 277, 15957 (2002)
53. N. L Thompson, T.P. Burghardt, D. Axelrod: Biophys. J. 33, 435 (1981)
54. T.P. Burghard, D. Axelrod: Biophys. J. 33, 455 (1981)
55. N.L. Thompson, D. Axelrod: Biophys. J. 43, 103 (1983)
56. E. Hellen, D. Axelrod: J. Fluor. 1, 113 (1991)
57. R.M. Fulbright, D. Axelrod: J. Fluor 3, 1 (1993)
58. A.L. Stout, D. Axelrod: Biophys. J. 67, 1324 (1994).
59. A.M. McKiernan, R.C. MacDonald, R.I. MacDonald, D. Axelrod: Biophys. J. 73, 1987 (1997)
60. S.E. Sund, D. Axelrod: Biophys. J. 79, 1655 (2000)
61. E. Kalb, J. Engel, L.K. Tamm: Biochemistry 29, 1607 (1990)
62. R. Gilmanshin, C.E. Creutz, L.K. Tamm: Biochemistry 33, 8225 (1994)
63. P. Hinterdorfer, G. Baber, L.K. Tamm: J. Biol. Chem. 269, 20360 (1994)
64. B.C. Lagerholm, T.E. Starr, Z.N. Volovyk, N.L. Thompson: Biochemistry 39, 2042 (2000)
65. R.D. Tilton, A.P. Gast, C.R. Robertson: Biophys. J. 58, 1321 (1990)
66. J.M. Moran-Mirabal, J.B. Edel, G.D. Meyer, D. Throckmorton, A.K. Singh, H.G. Craighead: Biophys. J. 89, 296 (2005)
67. B.S. Khakh, J.A. Fisher, R. Nashmi, D.N. Bowser, H.A. Lester: J. Neurosci. 25, 6911 (2005)
68. S.E. Sund, J.A. Swanson, D. Axelrod: Biophys. J. 77, 2266 (1999)
69. S. Hadjidemetriou, D. Toomre, J.S. Duncan: Medical Image Computing and Computer-Assisted Intervention, Lecture Notes in Computer Science 3749, 761 (2005)
70. M.D. Wang, D. Axelrod: Devel. Dynam. 201, 29 (1994)
71. A. Demuro, I. Parker: Biol. Res. 37, 675 (2004)
72. G.M. Omann, D. Axelrod: Biophys. J. 71, 2885 (1996)
73. F.M. Boldt, J. Heinze, M. Diez, J. Petersen, M. Borsch: Anal. Chem. 76, 3473 (2004)
74. A. Hoshino, K. Fujioka, N. Manabe, S. Yamaya, Y. Goto, M. Yasuhara, K. Yamamoto: Microbiol Immunol 49, 461 (2005)

75. E. Matveeva, Z. Gryczynski, J. Malicka, I. Gryczynski, J.R. Lakowicz: Anal Biochem. 334, 303 (2004)
76. H.P. Jennissen, T. Zumbrink: Biosensors Biolectron. 19, 987 (2004)
77. D. Willard, G. Proll, S. Reder, G. Gauglitz: Envir. Sci. Pollution Res. 10, 188 (2003)
78. D. Pant, H.H. Girault: Phys. Chem. Chem. Phys. 7, 3457 (2005)
79. S. Ishizaka, Y. Ueda, N. Kitamura: Anal. Chem. 76, 5075 (2004)
80. F. Hashimoto, S. Tsukahara, H. Watarai: Langmuir 19, 4197 (2003)
81. A. Banerjee, K.D. Kihm: Phys. E Rev. 72, 042101 (2005)
82. S. Blumberg, A. Gajraj, M.W. Pennington, J.C. Meiners: Biophys. J. 89, 1272 (2005)
83. Y. He, H.W. Li, E.S. Yeung: J. Phys. Chem. B 109,8820 (2005)
84. E. Mei, A. Sharonov, J.H. Ferris, R.M. Hochstrasser: Appl. Phys. Lett. 86, 043102 (2005)
85. K.D. Kihm, A. Banerjee, C.K. Choi, T. Takagi: Exper. Fluids 37, 811 (2004)
86. K.S. McCain, J.M. Harris: Anal. Chem. 75,3616 (2003)
87. K.S. McCain, P. Schluesche, J.M. Harris: Anal. Chem. 76, 930 (2004)
88. S. Bike: Curr. Opin. Colloid. Interf. Sci. 5, 144 (2000)
89. E.H. Hellen, D. Axelrod: J. Opt. Soc. Am. B 4, 337 (1987)
90. J. Mertz: J. Opt. Soc. Am. B 17, 1906 (2000)
91. M.G. Somekh, S.G. Liu, T.S., Velinov, C.W. See: Appl. Opt. 39, 6279 (2000)
92. G.L. Duveneck, A.P. Abel: Proc. SPIE 3858, 59 (1999)
93. H. Schneckenburger, K. Stock, W.S.L. Strauss, J. Eickholz, R. Sailer: J. Microsc.-Oxford 211, 30 (2003)
94. C. Czeslik, C. Royer, T. Hazlett, W. Mantulin: Biophys. J. 84, 2533 (2003)
95. D. Axelrod: Biophys. J. 26, 557 (1979).
96. R.M. Dickson, D.J. Norris, W.E. Moerner: Phys. Rev. Lett. 81,5322 (1998)
97. M.A. Lieb, J.M. Zavislan, L. Novotny: J. Opt. Soc. Am. B 21, 1210 (2004)
98. S. Wakelin, C.R. Bagshaw: J. Microsc.-Oxford 209,143 (2003)
99. P. Chang, D. Axelrod, G.M. Omann, J.J. Linderman: Cell. Signalling 17, 605 (2005)
100. K. Hassler, T. Anhut, R. Rigler, M. Gosch, T. Lasser: Biophys. J. 88, L1 (2005)
101. M. Oheim, F. Schapper: J. Phys. D-Appl. Phys 38, R185 (2005)
102. J.W.M. Chon, M. Gu: Appl. Optics 43, 1063 (2004)
103. F. Schapper, J.T. Goncalves, M. Oheim: Eur. Biophys. J. 32, 635 (2003)
104. Z. Huang, N.L. Thompson: Biophys. Chem 47, 241 (1993)
105. M. Gavutis, S. Lata, P. Lamken, P. Muller, J. Piehler: Biophys. J. 88, 4289 (2005)
106. V. Beaumont, A. Llobet, L. Lagnado: Proc. Natl. Acad. Sci. USA 102, 10700 (2005)
107. H. Tatsumi, Y Katayama, M Sokabe: Neurosci. Res. 35, 197 (1999)
108. A.B. Mathur, G.A. Truskey, W.M. Reichert: Biophys. J. 78, 1725 (2000)
109. A. Trache, G.A. Meininger: J. Biomed. Optics 10, 064023 (2005)
110. T. Yamada, R. Afrin, H. Arakawa, A. Ikai: FEBS Lett. 569,59 (2004)
111. W.H. Weber, C.F. Eagen: Opt. Lett. 4, 236 (1979)
112. A.M. Mattheyses, D. Axelrod: J. Biomed. Optics 10, 054007 (2005)
113. K. Stock, R. Sailer, W.S.L. Strauss, M. Lyttek, R. Steiner, H. Schneckenburger: J. Microsc.-Oxford 211, 19 (2003)
114. D. Loerke, B. Preitz, W. Stuhmer, M. Oheim: J. Biomed. Optics 5, 23 (2000)
115. D. Axelrod: J. Biomed. Opt. 6, 6 (2001)
116. A.L. Stout, D. Axelrod: Appl. Opt. 28, 5237 (1989)
117. D. Loerke, B. Preitz, W. Stuhmer, M. Oheim: J. Biomed. Opt. 5, 23 (2000)
118. D. Loerke, W. Stuhmer, M. Oheim: J. Neurosci. Meth. 119, 65 (2002)
119. J.R. Abney, B.A. Scalettar, N.L. Thompson: Biophys. J. 61, 542 (1992)

120. F. Lanni and T. Wilson, in "Imaging Neurons – A Laboratory Manual" (R. Yuste, F., Lanni, A. Konnerth, eds.), Chap. 8. Cold Spring Harbor Laboratory Press, New York, 1999
121. G.E. Cragg, P.T.C. So: Opt. Lett. 25, 46 (2000)
122. M. Gustafsson, D.A. Agard, J.W. Sedat: Proc. SPIE 3919, 141 (2000)
123. D. Axelrod, manuscript on ring polarization, *in preparation*
124. T. Ruckstahl, S. Seeger: Optics Lett. 29, 569 (2004)
125. T. Ruckstahl, S. Seeger: Appl. Optics 42, 3277 (2003)
126. S. Inoue, R.A. Knudson, M. Goda, K. Suzuki, C. Nagano, N. Okada, H. Takahashi, K. Ichie, M. Iida, K. Yamanak: K. J. Microsc. 201, 341 (2001)
127. A. Mattheyeses, D. Axelrod: Microsc Res. Techniques, *in press* (2006)
128. N.L. Thompson, K.H. Pearce, H.V. Hsieh: Eur. Biophys. J. 22, 367 (1993)
129. A. Sarkar, R.B. Robertson, J. M Fernandez: Proc. Natl. Acad. Sci. USA 101, 12882 (2004)
130. A. Mattheyses, D. Axelrod, D: J. Biomed Optics 11, 014006/1 (2005)
131. D. Axelrod, E.H. Hellen, R.M. Fulbright: in "Fluorescence spectroscopy: principles and applications, V. 3: Biochemical applications", (J. Lakowicz, ed.) p. 289. Plenum Press, New York, 1992
132. N.J. Harrick: "Internal Reflection Spectroscopy", Wiley Intersci., N.Y. (1967)
133. A.G. Tweet, G.L. Gaines, W.D. Bellamy: J. Chem. Phys. 40, 2596 (1964)
134. T. Hirschfeld: Can. Spectrosc. 10, 128 (1965)
135. C.K. Carniglia, L. Mandel, K.H. Drexhage: J. Opt. Soc. Am. 62, 479 (1972)
136. N.J. Harrick, G.I. Loeb: Anal. Chem. 45, 687 (1973)
137. M.N. Kronick, W.A. Little: J. Immunol. Methods 8, 235 (1975)
138. R.W. Watkins, C.R. Robertson: J. Biomed. Mater. Res. 11, 915 (1977)
139. E.J. Ambrose: Exp. Cell Res. Suppl. 8, 54 (1961)

9 Nonlinear Optical Microscopy

François Lagugné Labarthet and Yuen Ron Shen

9.1 Introduction

The constant evolution of optical microscopy over the past century has been driven by the desire to improve the spatial resolution and image contrast with the goal to achieve a better characterization of smaller specimens. Numerous techniques such as confocal, dark-field, phase-contrast, Brewster angle and polarization microscopies have emerged as improvement of conventional optical microscopy. Being a pure imaging tool, conventional optical microscopy suffers from its low physical and chemical specificity. This can be remedied by combining it with spectroscopic technique like fluorescence, infrared or Raman spectroscopy. Such microscopes have been successfully applied to the study of a wide range of materials with good spectral resolution. However their spatial resolution is restricted by the diffraction limit imposed by the wavelength of the probe light. In infrared microscopy, for instance, the lateral resolution is a few microns which is insufficient to resolve sub-micron structures. Conventional microscopy also does not provide microscopic information about the real surface structure of a specimen. Even in reflection geometry, they can only probe the structure of a surface layer averaged over a thickness of a reduced wavelength. Furthermore, they are insensitive to the polar organization of molecules in the layer although this could be important. In biophysics, for example, it is interesting to know the polar orientation of molecules adsorbed on a membrane and its influence on the membrane physiology.

In this context, nonlinear optical measurements used in conjunction with microscopy observation have created new opportunities [1,2]. Second-order nonlinear processes such as second harmonic (SH) or sum frequency generation (SFG) are highly surface-specific in media with inversion symmetry and uniquely suited for *in-situ* real-time investigation of surfaces and interfaces [3–5]. With their submonolayer surface sensitivity, SHG and SFG microscopies can be used to characterize inhomogeneities, impurities, formation of domains on surfaces or buried interfaces by mapping out the spatial variation of nonlinear susceptibilities at the interfaces. They are also sensitive to the orientation and distribution of molecules useful for evaluation of the structure and reactivity of a surface [6,7]. Third-order processes such as third harmonic generation (THG), coherent anti-Stokes Raman scattering (CARS) and two-photon excited fluorescence (TPEF) microscopies are of interest for the study of buried structures [8–10]. Because the output of second- and third- order processes scales, respectively, with the square and the cube of the

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

excitation intensity, the focal excitation volume is greatly reduced, enhancing the depth resolution and reducing the out-of-focus background noise. This simple idea has stimulated the development of THG, CARS and TPEF microscopies in an effort to image buried structures in transparent materials or in biological tissues with a three dimensional sectioning capability. Using ultra-short pulses (10^{-12}–10^{-15} s) CARS has also been demonstrated to be a possible technique to probe the interior of living cells in real time [11].

Vibrational spectra are known as fingerprints of molecules. Nonlinear optical microspectrometry finds more interest in the mapping of localized domains via vibrational resonances. Microscopy with infrared (IR)–visible (Vis) sum frequency vibrational spectroscopy allows imaging of molecule-specific domains at a surface. Using a conventional far-field microscope, it can have a spatial resolution one order of magnitude better than Fourier-transform infrared microspectrometry. CARS microscopy also provides vibrational identification of chemical species and is an alternative way to conventional Raman confocal microscopy. Microspectrometry in the electronic transition region is also useful. Multiphoton fluorescence microscopy is becoming a standard technique for biological research [12]. In contrast to one-photon-excited fluorescence, multiphoton-excited fluorescence allows the use of input radiation in the transparent region of a sample and is capable of probing the interior structure of the sample with little laser-induced damages.

This brief introduction to nonlinear optical microscopy would be incomplete without mentioning the recent development in combining nonlinear optical measurements with near-field scanning optical microscopy (NSOM) techniques. In the past several years a growing number of studies of NSOM-SHG/SFG, NSOM-THG and NSOM-TPEF have been reported [13–16]. The spatial resolution of such optical microscopy is limited by the tip apex radius of the optic fiber; it varies from 20 to 200 nm depending on the techniques. In such an experiment, the sample-tip distance is maintained constant by using a feedback mechanism, which then also yields the surface topographical information. The latter can be correlated with the nonlinear NSOM result on specific molecular properties of the sample with a high spatial resolution. With short laser pulses, transient properties of nanostructures can also be probed.

In this chapter, we review a number of nonlinear optical techniques combined with microscope measurements that have been developed in recent years. We will describe separately, SHG, SFG, THG, CARS and multiphoton fluorescence microscopies. For each technique, we will first recall the underlying principle and describe the typical optical setup. We will then focus our interest on several chosen examples taken from the literature. Finally, the resolution limit as well as possible improvements of the various techniques will be discussed.

9.2 Second Harmonic Nonlinear Microscopy

9.2.1 Basic Principle of SHG

We describe here briefly the basic principle of SHG and the extension to the case of a strongly focused fundamental input beam. Details can be found elsewhere [17,18]. SHG originates from a nonlinear polarization $\boldsymbol{P}^{(2)}(2\omega)$ induced in a medium by an incoming field $\boldsymbol{E}(\omega)$:

$$\boldsymbol{P}^{(2)}(2\omega) = \varepsilon_0 \overleftrightarrow{\chi}^{(2)} : \boldsymbol{E}(\omega)\boldsymbol{E}(\omega) \,, \tag{9.1}$$

where ε_0 is the vacuum permittivity and $\overleftrightarrow{\chi}^{(2)}$ denotes the nonlinear susceptibility of the medium. If the medium has inversion symmetry, then $\overleftrightarrow{\chi}^{(2)}$ for the bulk vanishes under the electric dipole approximation. At the surface or interface, however, the inversion symmetry is necessarily broken, and the corresponding $\overleftrightarrow{\chi}^{(2)}_{\mathrm{S}}$ is non-zero. This indicates that SHG can be highly surface-specific . As a result, SHG has been developed into a useful tool for many surface studies [19,20].

For SHG in reflected direction from a surface or interface, it has been shown that the output signal is given by

$$S(2\omega) = \frac{(2\omega)^2}{8\varepsilon_0 c^3 \cos^2\beta} \left|\chi^{(2)}_{\mathrm{eff}}\right|^2 [I(\omega)]^2 AT, \tag{9.2}$$

where β is the exit angle of the SH output, $I(\omega)$ is the beam intensity at ω, A is the beam area at the surface, T is the input laser pulse-width, and

$$\chi^{(2)}_{\mathrm{eff}} = \left(\overleftrightarrow{L}_{2\omega} \cdot \hat{\boldsymbol{e}}_{2\omega}\right) \cdot \overleftrightarrow{\chi}^{(2)}_{\mathrm{S}} : \left(\overleftrightarrow{L}_{\omega} \cdot \hat{\boldsymbol{e}}_{\omega}\right)\left(\overleftrightarrow{L}_{\omega} \cdot \hat{\boldsymbol{e}}_{\omega}\right) \,, \tag{9.3}$$

with $\hat{\boldsymbol{e}}_{\omega}$ being the unit polarization vector and $\overleftrightarrow{L}_{\omega}$ the Fresnel factor at frequency ω. The surface nonlinear susceptibility $\overleftrightarrow{\chi}^{(2)}_{\mathrm{S}}$ is related to the molecular hyperpolarizability $\overleftrightarrow{\alpha}^{(2)}$ by a coordinate transformation

$$\left(\chi^{(2)}_{S}\right)_{ijk} = N_{\mathrm{S}} \sum_{i'j'k'} \left[\alpha^{(2)}_{i'j'k'} \left\langle \left(\hat{\boldsymbol{i}} \cdot \hat{\boldsymbol{i}}'\right)\left(\hat{\boldsymbol{j}} \cdot \hat{\boldsymbol{j}}'\right)\left(\hat{\boldsymbol{k}} \cdot \hat{\boldsymbol{k}}'\right)\right\rangle\right] \,, \tag{9.4}$$

where i, j, k and i', j', k' stand for the laboratory and molecular coordinate axes, the angular brackets refer to an average over the molecular orientational distribution, and N_{S} is the surface density of molecules. At a surface of high symmetry, only a few $(\chi^{(2)}_{\mathrm{S}})_{ijk}$ elements are independent and nonvanishing. They can be determined from SHG measurements with different input/output polarizations. From $(\chi^{(2)}_{\mathrm{S}})_{ijk}$ together with knowledge of $\alpha^{(2)}_{i'j'k'}$, an approximate orientational distribution of the molecules could be deduced.

From (9.4), it is seen that spatial variation of molecular species, molecular density, and molecular orientation and arrangement can lead to spatial variation of $\overleftrightarrow{\chi}^{(2)}_{\mathrm{S}}$. Such variation on the micrometer or nanometer scale could be probed by SHG microscopy.

The above theoretical description of SHG assumes the usual plane-wave approximation. However, SHG microscopy often requires a strongly focused input laser beam. Obviously, (9.2) for SHG output has to be modified by a geometric factor to take into account the focusing geometry. Fortunately, this geometric factor should remain unchanged as the focal spot scans over a flat surface, and therefore the spatial variation of the SHG output should still reflect the spatial variation of $|\chi^{(2)}_{\text{eff}}|^2$ over the surface, yielding a surface microscope image.

Geometries for Second Harmonic Far-Field Microscopy The first experiments of second harmonic microscopy were performed in the middle of the seventies by Hellwarth and Christensen [21]. In these experiments, grain structures and defects localized at the surface of thin films of nonlinear materials (ZnSe, GaAs, CdTe) were imaged by the second harmonic signal ($\lambda = 532\,\text{nm}$) generated in the transmitted direction using a nanosecond pulsed Nd:YAG laser. Boyd and coworkers [22] first demonstrated the possibility of SH microscopy imaging of a surface monolayer. Even with a crude focusing and scanning setup, they were able to observe SHG in reflection, with a spatial resolution of several microns, from a hole burned in a Rhodamine 6G monolayer deposited on a silica substrate. Later, Schultz et al. [23,24] used SH microscopy to study surface diffusion of Sb adsorbates on Ge. They adopted the parallel imaging scheme with a photodiode array. The spatial resolution limited by the pixel size of the 2-D detector was estimated to be ~5 μm. With the use of a 5 ns and 10 Hz pulsed Nd:YAG laser beam as the pump source, an acquisition time of 5–10 minutes was required.

In recent years, tremendous progress on laser technology, imaging systems, photodetectors and scanning devices has been made. This facilitates the development of SH microscopy. On the other hand, growing interest in surfaces, surface monolayers and thin films has also fostered the interest in SH microscopy [25]. The technique has been applied to the study of a wide range of surfaces and interfaces including Langmuir monolayers at the air-water interface [26], multilayers assemblies [27], self assembled monolayers [28], metallic [29] and semiconductors surfaces [30–32], nonlinear optical materials [33–35], and biomembranes [36–39].

Using a nsec Q-switched Nd:YAG laser and a CCD detector array, Flörsheimer and coworkers has developed SH microscopy in various geometries (Fig. 9.1a–d) and applied it to Langmuir monolayers [6,40–43]. Variation of the polarizations of the input and output beams allows the determination of polar orientation of molecules in microdomains. As examples, the SH microscopy images of a Langmuir monolayer of a 2-docosylamino-5-nitropyridine (DCANP) on water are displayed in Fig. 9.1e and f. They were obtained with the beam geometry described in Fig. 9.1d and the input/output polarization combinations depicted in Fig. 9.1g and h, respectively. A domain size of a few μm can be easily resolved.

In simple cases like 2-docosylamino-5-nitropyridine (DCANP), the nonlinear polarizability of the molecules can be approximated by a single tensor element $\alpha^{(2)}_{zzz}$ along the molecular axis $\hat{z}$. If the molecules are well aligned in a monolayer domain, then the nonlinear susceptibility of the domain is also dominated by a single

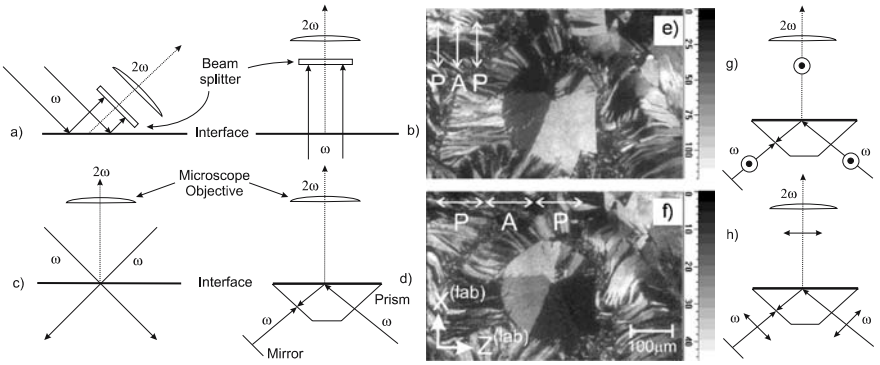


Fig. 9.1. (**a**,**b**,**c**,**d**) Schemes of the second harmonic microscopes using various geometries. (**e**,**f**) SH micrographs of a DCANP Langmuir film at the surface of water. Image (**e**) is obtained using the *ss* polarization configuration (**g**) while image (**f**) is obtained using the *pp* polarization configuration (**h**) [6]. (Copyright (1999) Wiley-VCH)

tensor element $\chi^{(2)}_{Z'Z'Z'}$ along the alignment direction $\hat{\boldsymbol{Z}}'$. In this case, the *s*-in/*s*-out polarization combination of SHG probes an effective nonlinear susceptibility

$$\left|\chi^{(2)}_{\text{eff}}\right|_{ss} = \left|F_{ZZ}\chi^{(2)}_{ZZZ}\sin^3\phi\right| = \left|F_{ZZ}\chi^{(2)}_{Z'Z'Z'}\cos^3\theta\sin^3\phi\right| \ , \tag{9.5}$$

where θ is the tilt angle of $\hat{\boldsymbol{Z}}'$ from the surface plane, ϕ is the angle between the incidence plane and $\hat{\boldsymbol{Z}}$, which is the projection of $\hat{\boldsymbol{Z}}'$ on the surface, and F_{ZZ} is the product of the three relevant Fresnel factors (see (9.3)). The beam geometry of Fig. 9.1g corresponds to such a case.

For the beam geometry of Fig. 9.1h, the effective nonlinear susceptibility deduced is

$$\begin{aligned}\left|\chi^{(2)}_{\text{eff}}\right|_{pp} &= \left|F_{ZZ}\cos^2\beta_1\cos^3\phi\chi^{(2)}_{ZZZ} + F_{ZY}\sin^2\beta_1\cos\phi\chi^{(2)}_{ZYY}\right| \\ &= \left|\left(F_{ZZ}\cos^2\beta_1\cos^3\theta\cos^3\phi + F_{ZY}\sin^2\beta_1\cos\theta\sin^2\theta\cos\phi\right)\chi^{(2)}_{Z'Z'Z'}\right| \ ,\end{aligned} \tag{9.6}$$

where β_1 is the incidence angle of the fundamental input, F_{ZZ} and F_{ZY} are the respective products of Fresnel factors, and $z - y$ defines the incidence plane. It is seen from (9.5) that the Z direction can be determined from the variation of SHG with ϕ with the *s*-in/*s*-out polarization. Then from (9.5) and (9.6), $\chi^{(2)}_{Z'Z'Z'}$ and θ can be deduced if β_1 is known and $(\chi^{(2)}_{\text{eff}})_{ss}$ and $(\chi^{(2)}_{\text{eff}})_{pp}$ are measured. Thus the various domains in Fig. 9.1e and f clearly correspond to domains of different molecular alignments.

SH microscopy has also been used to probe three dimensional structures of biological tissues and defects in materials. In this case, SHG originates from the focal region in the bulk. Using a reflection geometry and femtosecond pulses from a mode-locked dye laser, Guo et al. [48] have obtained SH images of skin tissues.

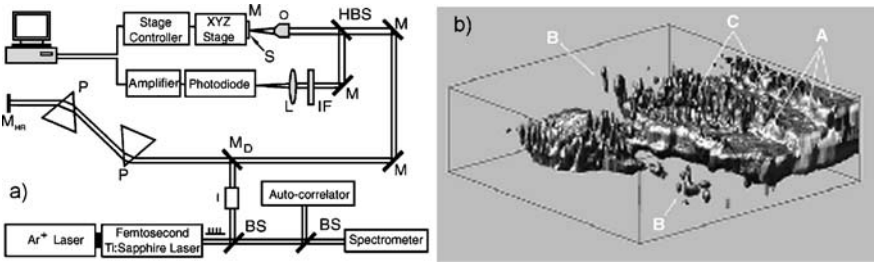


Fig. 9.2. (**a**) Optical setup used for SH tomographic imaging in the reflectance mode. BS: beam splitter, HBS: harmonic dichroic filter, L: lens, M: mirrors, O: objective, P: prisms, S: sample. (**b**) 3-D second harmonic generation(70 μm × 70 μm × 30 μm) of lithium triborate crystal fragments. Terraces (A), isolated crystallites (B), columnar stacking (C), can be identified [49]. (Copyright (1998) Optical Society of America)

In their setup, a 27× microscope objective was used for both excitation and collection of SHG, and the sample was mounted on a *XYZ* stage for lateral and axial movement, allowing 3-D imaging of a fascia membrane and muscle attached to the biological tissue. Interpretation of the images in term of molecular organization could be improved by analyzing the polarization dependence of the collected signal. In the case of the muscles tissues, *in-vivo* investigations could also be of interest to monitor the physiological change of the fibrils membranes under stress.

Gauderon et al. [49] have used femtosecond pulses from a Ti:sapphire laser to obtain by reflection 3-D SH microscopy images of lithium triborate (LBO) crystal fragments embedded in agar. The experimental arrangement in shown in Fig. 9.2a. The sample on the scanning stage was scanned by the focused laser beam both laterally and axially. A typical image is presented in Fig. 9.2b describing a scanned volume of 70 μm × 70 μm × 30 μm. Terraces, isolated microcrystallites and columnar arrangements of the crystal fragments are clearly observed. The spatial resolution of the image depends on the dimensions of the focal point. Since SHG is proportional to the square of the local fundamental input intensity, the effective point-spread function of the SH output is smaller than that of the input by a $\sqrt{2}$ factor allowing a resolution of $0.61\lambda/(\sqrt{2}\mathrm{NA})$, where λ is the input wavelength and NA denotes the numerical aperture of the focusing lens. With λ=790 nm and NA = 0.8 this lead to a spatial resolution of 0.4 μm, that is comparable to the maximum resolution obtained in confocal linear optical microscopy. It could be further improved using a confocal pinhole in the detection optics.

9.2.2 Coherence Effects in SH Microscopy

In SH microscopy, interferences from various features on the surface and/or background can lead to severe distortion of the SH image although, in most studies, such effects have not been taken into account. Parallel (non-scanning) SHG imaging using a 2-D detector array, for example, could suffer from interferences between contributions originating from different sites of the sampled area. On the other hand,

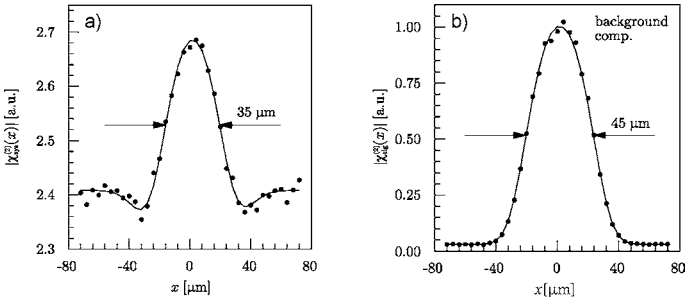


Fig. 9.3. (**a**) SH signal profile of a damaged area before correction of the background contribution. The step size is 4 μm and the dwell time at each sample spot was 5 s. (**b**) Damaged area after background compensation [52]

despite longer acquisition time, the local nature of a scanning focused fundamental beam reduces interference effects and the morphology interpretation of images is generally more straightforward [50].

However, even if the fundamental beam is tightly focused, the coherent nature of SHG implies that the sample signal can still be distorted by background contribution from sources other than the sample [51]. To cancel a uniform background contribution, one can insert a NLO crystal in the optical pathway [50,52]. By adjusting the crystal orientation properly, it is possible to generate a SH field from the crystal with proper phase and amplitude to cancel the background SH field. Figure 9.3 shows an example. The original SH microscopy image of a damaged area of Si–SiO_2 interface has two minima and a central peak. After the background suppression, only the central peak with a slightly larger width is observed. The same background correction scheme may not work with the parallel imaging method if the background contribution is non-uniform over the probed area.

One can utilize the interference effect to obtain phase-contrast microscope images. In a recent experiment, Flörsheimer et al. [53] coherently mix the SH wave generated from a sample (interface) with the reference SH wave from a crystalline quartz plate. The resultant SH wave has a spatial intensity variation depending on $\cos[(\phi_S(\boldsymbol{r}) - \phi_R]$ where ϕ_S and ϕ_R are the phases of the SH waves from the sample and the reference, respectively. The microscope image then reflects the spatial variation of $\phi_S(\boldsymbol{r})$, which can provide additional information concerning the sample.

9.2.3 Scanning Near-Field Nonlinear Second Harmonic Generation

Far-field SH microscopy still has a spatial resolution limited to μm^2. Incorporation of near-field optics in SH microscopy overcomes the diffraction limit and allows studies of nanosructures. Near-field SHG measurements on rough metallic surfaces [54], ferroelectric domains [55], piezoceramics [56], and LB films [13] have been reported.

The first near-field SHG images were obtained by Bozhevolnyi et al. [13,57]. Their experimental setup (Fig. 9.4a) is a NSOM with a shear-force-based feedback

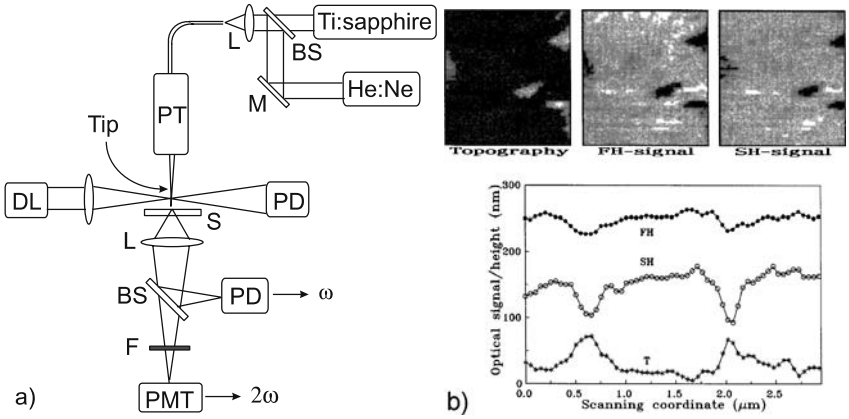


Fig. 9.4. (**a**) Experimental setup for local probing of second-order susceptibilities with a near-field optical microscope. BS: beam splitters, M: mirror, L: micro-objectives, S: sample, PT: piezo translators, DL: diode laser, PD: photodiodes, F: filter and PMT: photomultiplier tube. The detected second harmonic power versus the pump power squared is shown in the inset. (**b**) Topography, fundamental and SH near-field images ($2.5\times3\,\mu m^2$) of a $LiNbO_3$ crystal with corresponding cross-sections of these images. The images were recorded simultaneously with a total acquisition time of 9 minutes. Reprinted from [57], Copyright (1998), with permission from Elsevier Science

system. The femtosecond fundamental input pulses from a mode-locked Ti:sapphire laser irradiate the sample through a single-mode fiber with an uncoated sharp tip, and both the fundamental and the SH output in the transmitted direction are collected by photodetectors. The fiber tip has its axial position controlled by the tuning fork sensor technique [60]. The scheme then allows the mapping of surface topography with a few nm resolution like that in atomic force microscopy when the sample surface is scanned. The topography can be correlated with the observed spatial variations of the fundamental and SH outputs from the sample. As an example, Fig. 9.4b presents the three images of a $LiNbO_3$ crystal with three small scatterers on it. All the three images display the presence of the unidentified particles at the surface. In the fundamental and SH images, the particles appear as dark spots. The decrease of signals over the particles is presumably due to scattering loss as the tip moves over the particles. The observed lateral resolution of SH-NSOM was around 100–150 nm in the experiments that map the domain edges of a Langmuir–Blodgett film [13,57] and a quasi-phase matching crystal [61]. Similar resolution was obtained by Smolyaninov et al. [56].

SH-NSOM images were also obtained from an electrically poled grating inscribed on a polymer thin film containing azobenzene molecules [58]. Using an amplified femtosecond laser pumping a optical parametric amplifier, the fundamental input wavelength was set at 1.28 μm with an energy of 250 nJ/pulse for 80 fs pulses (Fig. 9.5). A uncoated chemically etched optical fiber was used as the near-field probe in the collection geometry. For a grating of sufficient modulation

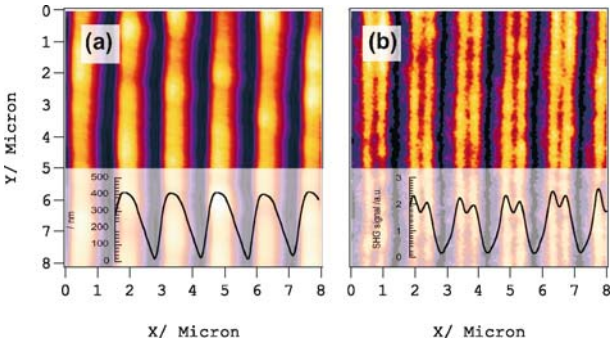


Fig. 9.5. (**a**) Topographical image of a grating with a periodicity of 1.37 μm and a modulation depth of 400 nm inscribed on p(DR1M)thin film by holography. (**b**) Corresponding SHG-NSOM image obtained with 1.28 μm femtosecond input pulses. The SH signal was measured at 640 nm. Reprinted from [58].Copyright (2003), Optical Society of America

depth, the $\overleftrightarrow{\chi}^{(2)}$ periodic structure appears quite different from the topography revealed by linear NSOM or AFM (Fig. 9.5a,b). Photoisomerization of the azobenzene side chains used to form the grating is known to result in mass transport and to make the mass density higher at the ridges than the valleys. Less-effective poling at regions with higher density could lead to the observed complex $\overleftrightarrow{\chi}^{(2)}$ structure (Fig. 9.5b). Polarized measurements of the SH-NSOM images were conducted to give a quantitative information of the orientational distribution of the azobenzene molecules via a determination of the odd order parameters $\langle P_1 \rangle = \langle \cos\theta \rangle$ and $\langle P_3 \rangle = 0.5(5\langle \cos^3\theta \rangle - 3\langle \cos\theta \rangle)$ with θ being the orientation angle of the rod-like moieties [59]. In this work, the SH signal collected from the optical fibre probe was measured with a lateral resolution better than 500 nm as shown in (Fig. 9.5b).

There are many variations of beam geometry in SHG-NSOM. With the fundamental input incident on a sample through a fiber tip, the SH output can be detected either in the transmitted and reflected direction. It is also possible to use a fiber tip to collect near-field SHG from a sample which is broadly illuminated by the fundamental input. One can also have SHG in an evanescent mode and use a metal tip close to the sample surface to induce a SH leak from the surface.

As is true for all NSOM, the theoretical description of SHG-NSOM is difficult. The attempts to model SHG-NSOM have been reported recently [62–64]. The SH signal depends in a complicated way on the polarization of the fundamental field, and multiple scattering of both the fundamental and the SH waves in the medium must be taken into account. Thus, unlike in far-field SHG, little information can be deduced from the polarization dependence of SHG-NSOM. Near-field SHG from mesoscopic structures using either the tip illumination mode [65] or the tip collection mode [66] has been described by a self-consistent integral equation formalism. The calculation shows a strong localization of the SH field around the mesostructures in both lateral and axial dimensions. This strong 3-D localization of SHG is of particular interest to imaging of nanostructures with a NSOM.

9.3 Sum Frequency Generation Microscopy

9.3.1 Basic Principle of Sum Frequency Generation

As a second-order nonlinear optical process, sum frequency generation (SFG) is also highly surface-specific in media with inversion symmetry, and can be used as a surface probe. The IR-Vis SFG has already been developed into a powerful surface vibrational spectroscopic technique [67,68]. In IR-Vis SFG, a visible laser pulse (ω_{Vis}) overlaps with a infrared pulse (ω_{IR}) at a surface and induces a nonlinear polarization $\boldsymbol{P}^{(2)}$ at the sum frequency $\omega_{\mathrm{SF}} = \omega_{\mathrm{Vis}} + \omega_{\mathrm{IR}}$:

$$\boldsymbol{P}^{(2)}(\omega_{\mathrm{SF}} = \omega_{\mathrm{Vis}} + \omega_{\mathrm{IR}}) = \varepsilon_0 \overleftrightarrow{\chi}^{(2)} : \boldsymbol{E}(\omega_{\mathrm{Vis}})\boldsymbol{E}(\omega_{\mathrm{IR}}) \,, \tag{9.7}$$

which radiates and generates the SF output. As in SHG (9.2), the SF output in the reflected direction is given by:

$$S(\omega_{\mathrm{SF}}) \propto \left|\chi^{(2)}_{\mathrm{eff}}(\phi)\right|^2 I(\omega_{\mathrm{Vis}}) I(\omega_{\mathrm{IR}}) AT \,, \tag{9.8}$$

with

$$\chi^{(2)}_{\mathrm{eff}}(\phi) = \left(\overleftrightarrow{L}_{\omega_{\mathrm{SF}}} \cdot \hat{\boldsymbol{e}}_{\omega_{\mathrm{SF}}}\right) \cdot \overleftrightarrow{\chi}^{(2)}_{\mathrm{S}} : \left(\overleftrightarrow{L}_{\omega_{\mathrm{Vis}}} \cdot \hat{\boldsymbol{e}}_{\omega_{\mathrm{Vis}}}\right)\left(\overleftrightarrow{L}_{\omega_{\mathrm{IR}}} \cdot \hat{\boldsymbol{e}}_{\omega_{\mathrm{IR}}}\right) \tag{9.9}$$

and $\overleftrightarrow{\chi}^{(2)}_{\mathrm{S}}$ is related to the molecular hyperpolarizability $\overleftrightarrow{\alpha}^{(2)}$ by (9.4). If ω_{IR} is tuned over vibrational resonances, $\overleftrightarrow{\alpha}^{(2)}$ must exhibit corresponding resonance enhancement and can be expressed in the form

$$\overleftrightarrow{\alpha}^{(2)} = \overleftrightarrow{\alpha}^{(2)}_{\mathrm{NR}} + \sum_q \frac{\overleftrightarrow{a}_q}{(\omega_{\mathrm{IR}} - \omega_q) + i\Gamma_q} \,, \tag{9.10}$$

where $\overleftrightarrow{\alpha}^{(2)}_{\mathrm{NR}}$ is the non-resonant contribution and $\overleftrightarrow{a}_q$, ω_q and Γ_q denote the strength, resonant frequency and damping constant of the q^{th} vibrational mode, respectively. Correspondingly, we have

$$\overleftrightarrow{\chi}^{(2)}_{\mathrm{S}} = \overleftrightarrow{\chi}^{(2)}_{\mathrm{NR}} + \sum_q \frac{\overleftrightarrow{A}_q}{(\omega_{\mathrm{IR}} - \omega_q) + i\Gamma_q} \,. \tag{9.11}$$

If the infrared input is tunable and scans over vibrational resonances, the resonant enhancement of the SF output naturally yields a surface vibrational spectrum. Such spectra with different input/output polarization combinations can yield information about surface composition and structure in a local surface region. This technique has found many applications in many areas of surface science. However, only a few studies combining SFG vibrational spectroscopy and optical microscope techniques have been reported.

9.3.2 Far-Field SFG Microscopy

Combination of SF vibrational spectroscopy with optical microscopy allows detection or mapping of selective molecular species at a surface or interface. It also has the advantage of an improved spatial resolution in comparison to conventional FTIR microscopy because the resolution is now limited by the SF wavelength instead of the IR wavelength. This is similar to Raman microscopy, but the latter is not surface specific.

Flörsheimer has extended his development of SH microscopy to SF microscopy [6,43]. The experimental arrangement with the input beams in total reflection geometry is depicted in Fig. 9.6a. A Langmuir–Blodgett monolayer film to be imaged is adsorbed on the base surface of a prism. Two input beams enter by the two hypotenuse faces of the prism and the SF output exits from the film surface at an oblique angle.

Figure 9.6b shows an SFG image of a monolayer of arachidic acid molecules obtained with parallel imaging with the IR frequency set at $2962\,\mathrm{cm}^{-1}$ in resonance with the CH_3 stretching mode. The image exhibits dark holes and clear nonuniformity in the illumination region. Similar experiment with the IR frequency at $2850\,\mathrm{cm}^{-1}$ in resonance with the methylene stretch yielded little SF signal. These results indicate that the alkyl chains of the monolayer are in nearly all-trans configuration. The bright areas in the image in Fig. 9.6b represent regions with a densely packed molecules in trans conformation. The dark holes likely originate from a lack of molecules in the areas. They were not observed in freshly prepared LB films, but appeared and grew with time due to a dewetting or desorption.

In the experimental geometry of Fig. 9.6a, emission of SF output at an oblique angle could lead to image distortion in the parallel imaging process and deterio-

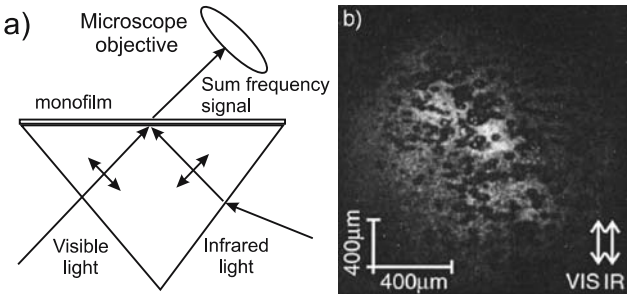


Fig. 9.6. (**a**) The thin film is deposited on the prism and is illuminated with two *p*-polarized pulsed beams, ω_{Vis} and ω_{IR}, impinging on the film in the total reflection geometry. The visible frequency, ω_{Vis}, is fixed at $18800\,\mathrm{cm}^{-1}$ and the infrared frequency, ω_{IR}, is tunable. No output analyzer is used and the SF signal is found to be mainly *p* polarized. (**b**) Sum frequency image of a LB monolayer of arachidic acid. The infrared frequency ($2962\,\mathrm{cm}^{-1}$) is tuned to the anti-symmetric stretching mode of methyl groups. Note that the scale bars are not of equal size due to the oblique incidence of the collecting objective. Figure 9.6b is reprinted with permission from [43]. Copyright (1999) American Chemical Society

ration of spatial resolution. The problem could be solved by having the SF output emitted normal to the surface. Unfortunately this may not be possible with the total-reflection geometry described in Fig. 9.6a. Matching of wave-vector components along the surface would require that the visible input beam be transmissive at the surface.

Since the first edition of the present chapter, several groups have focussed their attention on SFG microscopy. Following the pioneering work of Flörsheimer, Kuhnke and coworkers have extended SFG microscopy to opaque and reflecting samples [44]. Their microscope setup shown in Fig. 9.7 is an extension of conventional SFG spectroscopy with 35 ps input beams (λvis = 532 nm and tunable infrared) impinging on the surface at an oblique incidence angle (60 degrees from the surface normal). The imaging stage is composed of camera and microscope objectives, a blazed grating and a CCD camera yielding a lateral resolution of 4.9 μm and 3.1 μm in the lateral and axial directions, respectively.

With this setup, patterned self-assembled monolayers of octadecanethiol and ω-carboxyhexadecanethiol adsorbed on gold were studied with a selective analysis of the different chemical groups [45]. A very similar setup was later reported by Baldelli et al. with a spatial resolution of 10 μm on microcontact-printed mixed self assembled monolayers of mixtures of C16 thiols with methyl or phenyl terminal groups [46]. The main difficulties of the above-mentioned SF microscopy scheme lie in the oblique incidence geometry of the visible and infrared beams (and thus the SFG signal) resulting in a distortion of images of which only the central part

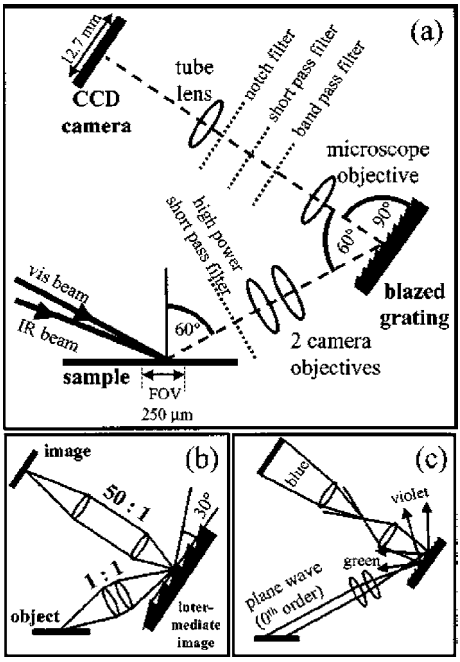


Fig. 9.7. (**a**) Schematic setup of the SFG microscope with the employed optical elements. (**b**) Path of the imaging beam. (**c**) Path of the illumination beam. FOV = field of view. Reused with permission from D.M.P. Hoffmann, K. Kuhnke, and K. Kern, Review of Scientific Instruments, 73, 3221 (2002). Copyright 2002, American Institute of Physics

is sharp. In addition, such microscopy configurations do not allow 3D sectioning imaging because the signal comes only from the surface.

SFG is however allowed in chiral media through their optical activity. Shen, Yang and coworkers have extended SFG microscopy to image systems containing chiral molecules [47]. In this optically active SFG setup (OA-SFG), the two input beams polarized in the incidence plane are focussed on the sample with the same microscope objective (60x, 1.4 NA oil immersion) and the SF beam polarized perpendicular to the incidence plane is collected in the transmission direction and detected by a UV-sensitive photomutiplier. (Fig. 9.8)

Chiral molecules, Binol (R/S-(-)- 1,1'-bi-2-naphtol), were used in their experiment. The OA-SFG signal was measured at 277 nm in a band of electronic transitions of Binol. The optical activity of Binol (R/S-(-)- 1,1'-bi-2-naphtol) was evidenced by comparing the SF signals from a racemic mixture and from a pure enantiomer. The SFG microscopic images of a thin liquid cell containing Binol decorated with silica beads of 2.4 μm diameter are shown in (Fig. 9.9). The resolution of the images is 400 nm.

Since the two beams were focussed by the same microscope objective, 3D sectional imaging was possible by raster scanning the sample around the focal region. In addition, this setup permits simultaneous measurements of chiral SHG and SFG from a sample and therefore allows both surface and bulk imaging of the same chiral system. OA-SFG microscopy has the advantage that it is free from achiral background and therefore can exhibit excellent contrast in discriminating chiral molecules from the achiral surrounding. While more research is still needed for applications to biological macromolecules, the technique could emerge as a powerful tool for microscopic imaging of biophysical processes.

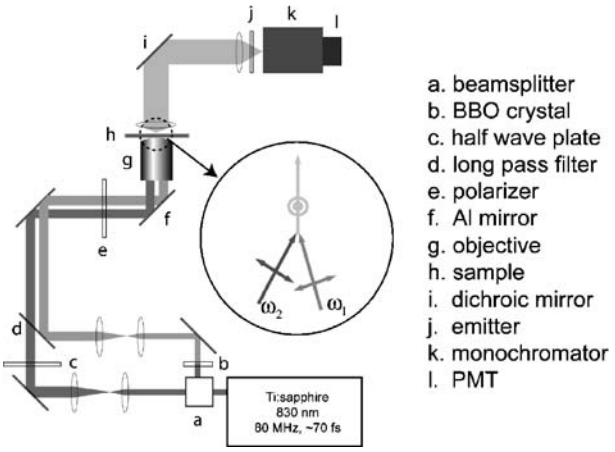


Fig. 9.8. (Scheme of experimental setup. The polarization arrangement for optically active SFG imaging is SPP (for S, P- and P-polarizations of the SF output, 1 and 2 inputs, respectively) in shown is the inset. Reprinted with permission from [47]. Copyright (2006) American Chemical Society

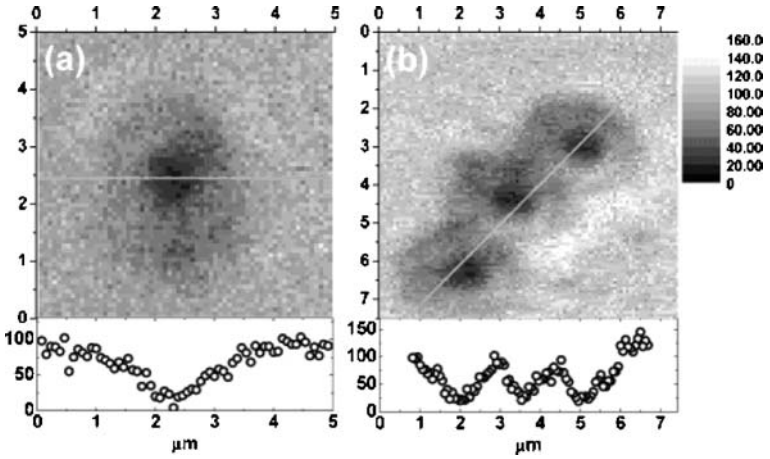


Fig. 9.9. (OA-SFG images of (**a** one and (**b**) three 2.4 μm diameter silica beads in R.Binol solutions. Bottom panels show signal variations along the lines in the images. The polarization combination used was SPP and the collection time per image was 78 ms/pixel at power levels of 0.18 mW (λ = 415 nm) and 14 mW(λ = 830 nm). Reprinted with permission from [47]. Copyright (2006) American Chemical Society

9.3.3 Near-Field SFG Imaging

Spatial resolution of SF microscopy can be improved to much below micron scale by near-field detection of the SF output, as demonstrated independently by Shen et al. [14], Schaller and Saykally [69] and Humbert et al. [70]. However, so far, only studies of SFG-allowed media have been reported.

The experimental setup of Shen et al. is described in Fig. 9.10a. It allows near-field imaging of SHG and SFG as well as topography of a sample with a lateral resolution of about 120 nm. The wavelengths of the two input beams from a nsec laser are fixed at 1.047 μm and 0.524 μm in the experiment. SHG and SFG from the sample deposited on the base surface of the prism are generated in the total reflection mode. The spatial variations of the evanescent waves above the sample surface are probed via optical tunneling by an Al-coated fiber with a 120 nm aperture attached to an *XYZ* piezoelectric scanner. With a shear force feedback to keep the probe-sample distance constant, the fiber tip also yields the surface topography as the probe scans the sample surface. Figure 9.10b shows the surface topography as well as the SH and SF microscopy images of three isolated NNP (N-(4-nitrophenyl)-(L)-prolinol) nanocrystals of ~360 nm in size that were deposited on the prism surface. The three images appear to be well correlated.

Schaller and Saykally [69] have performed near-field SFG microscopy with a tunable IR input. They adopted a commercial NSOM head for collection of the SF output (Fig. 9.11a). A femtosecond Ti:sapphire laser was used to provide the visible input and to pump an optical parametric amplifier/difference frequency generator system to generate the tunable IR input. The microscopy was applied to a thin

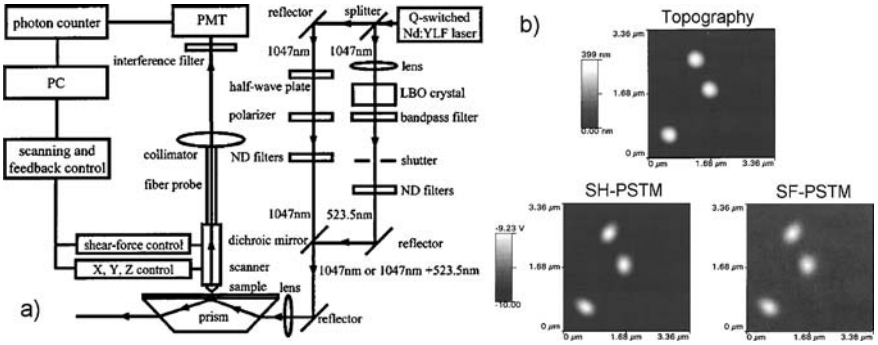


Fig. 9.10. (**a**) Optical set up used for both SHG and SFG photon scanning tunneling microscopy. (**b**) Topography, SH and SF images of NNP nanocrystals [14]

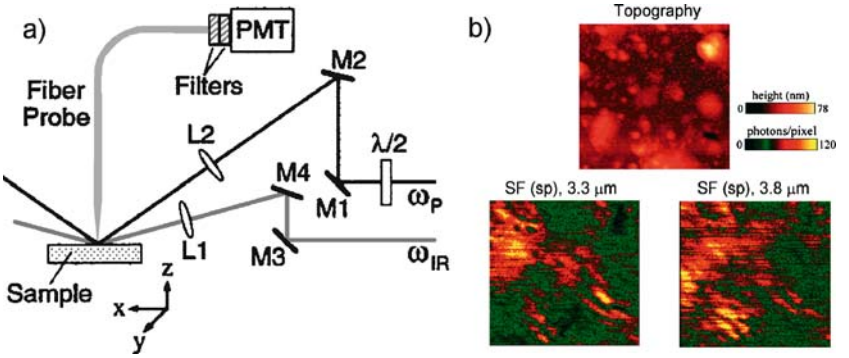


Fig. 9.11. (**a**) Setup for SFG-NSOM. The uncoated NSOM fiber probe tip (apex ~50 nm) operates in the collection mode. Infrared light is tunable from 2.8 to 10 μm. The two input beams overlap over a surface of 100 μm^2. (**b**) Topographic and SFG images (10 μm^2) of ZnSe thin film deposited on a substrate by chemical vapor deposition. Topographic and SH images are simultaneously acquired in 30 min. The spatial resolution is estimated to be 190 nm. Reprinted with permission from [69]. Copyright (2001) American Chemical Society

film of ZnSe with the IR wavelength varied from 3.1 to 4.4 μm. Figure 9.11b shows the surface topography and SF images at two IR wavelengths for a 10 μm^2 area of the sample. The SF images reveals strain patterns that are not observable in the surface topography. The lateral spatial resolution is estimated to be about 1/20 of the IR wavelength.

9.4 Third Harmonic Generation Microscopy

While second-order nonlinear optical processes require a medium without inversion symmetry, third-order processes such as third-harmonic generation are allowed in any medium. The nonlinear polarization induced in a medium responsible for

THG is

$$\boldsymbol{P}^{(3)}(3\omega) = \varepsilon_0 \overleftrightarrow{\chi}^{(3)} : \boldsymbol{E}(\omega)\boldsymbol{E}(\omega)\boldsymbol{E}(\omega) \,, \tag{9.12}$$

where $\overleftrightarrow{\chi}^{(3)}$ is the third-order nonlinear susceptibility for the process. In the plane wave approximation, the TH output is given by

$$S(3\omega) \propto \left|\chi^{(3)}_{\text{eff}}(\phi)\right|^2 I(\omega)I(\omega)I(\omega) \,, \tag{9.13}$$

with,

$$\chi^{(3)}_{\text{eff}}(\phi) = \left(\overleftrightarrow{L}_{3\omega} \cdot \hat{\boldsymbol{e}}_{3\omega}\right) \cdot \overleftrightarrow{\chi}^{(3)} : \left(\overleftrightarrow{L}_{\omega} \cdot \hat{\boldsymbol{e}}_{\omega}\right)\left(\overleftrightarrow{L}_{\omega} \cdot \hat{\boldsymbol{e}}_{\omega}\right)\left(\overleftrightarrow{L}_{\omega} \cdot \hat{\boldsymbol{e}}_{\omega}\right) \,. \tag{9.14}$$

If the fundamental input beam is strongly focused in the bulk of the medium, then the TH output becomes [18,71]

$$S_{3\omega} \propto P_{\omega}^3 \left| \int_0^l \frac{\chi^{(3)}(z') \exp(\mathrm{i}\Delta \boldsymbol{k} z') \mathrm{d}z'}{(1 + 2\mathrm{i}z'/b)^2} \right|^2 \,, \tag{9.15}$$

where the integration extends over the length of the medium, P_ω is the fundamental beam power, $\boldsymbol{k}_\omega$ and $\boldsymbol{k}_{3\omega}$ are the wave vectors at ω and 3ω, respectively, $b = \boldsymbol{k}_\omega . w_0^2$ is the confocal parameter of the fundamental beam with a waist radius of w_0 and $\Delta \boldsymbol{k} = 3\boldsymbol{k}_\omega - \boldsymbol{k}_{3\omega}$ is the phase mismatch. The integration in (9.15) nearly vanishes if $\Delta \boldsymbol{k} \leq 0$ and $l \gg 1/\Delta \boldsymbol{k}$ but is finite for $\Delta \boldsymbol{k} > 0$ [18]. However, in a medium with normal dispersion, we expect $\Delta \boldsymbol{k} < 0$ and hence a vanishing THG. Since third order nonlinearity away from resonance is generally very weak and tight focusing for THG is often required, this is obviously a shortcoming for using THG as a probe. Nevertheless, Tsang has pointed out that with focusing at an interface, the symmetry of the focusing geometry is broken and (9.15) needs to be modified [72]. Then even with $\Delta \boldsymbol{k} < 0$, THG in the forward direction can be appreciable.

In combination with a transmission optical microscope, THG can be used to image a transparent sample with a 3-D microscopy capability. Indeed, liquid crystal phase transitions [73], semiconductor microstructures [74], optical fiber structures [75,76] and biological samples [8,77] have been investigated with this technique. Recently, Schaller et al. have imaged red blood cells using THG-NSOM [15].

To illustrate the 3-D imaging capability of THG microscopy, we present in Fig. 9.12 the experimental setup and result of Squier et al. [78]. A regeneratively amplified femtosecond Ti:sapphire laser is used to pump an optical parametric amplifier, whose output is focused on the sample for TH imaging. The sample is mounted on an *XYZ* scanning stage. Figure 10.5 in Chap. 10 shows an example of such an image. The 3-D array of letters inscribed in a glass can be recognized with micron resolution.

9.5 Coherent Anti-Stokes Raman Scattering Microscopy

Similar to THG, coherent anti-Stokes Raman scattering (CARS) is a third order four-wave mixing process but is resonantly enhanced at vibrational resonances, and

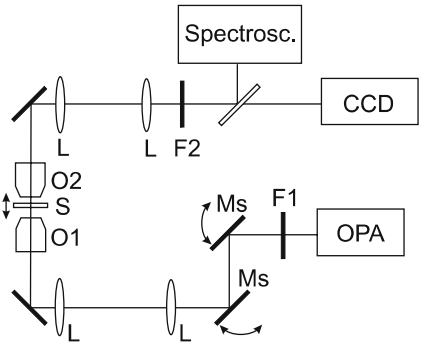


Fig. 9.12. Optical setup used by Squier et al. for both writing and reading microdamage patterns in optical glass: OPA: optical parametric amplifier, F1: long pass filter, Ms: scanning mirrors, O1 and O2: objective lenses, F2: blocking filter, M1: insertable mirror [78]. (Copyright (1999) Optical Society of America)

is a powerful vibrational spectroscopic tool [79]. The process involves two input waves at frequencies ω_{P} and ω_{S} with $\omega_{\mathrm{P}} - \omega_{\mathrm{S}}$ tuned to a vibrational resonance of the medium. The two inputs overlap in the medium and induce a third-order nonlinear polarization in the medium at the anti-Stokes frequency $\omega_{\mathrm{AS}} = 2\omega_{\mathrm{P}} - \omega_{\mathrm{S}}$:

$$\boldsymbol{P}^{(3)}_{\omega_{\mathrm{AS}}} = \epsilon_0 \overleftrightarrow{\chi}^{(3)} : \boldsymbol{E}(\omega_{\mathrm{P}})\boldsymbol{E}(\omega_{\mathrm{P}})\boldsymbol{E}^*(\omega_{\mathrm{S}}) \,, \tag{9.16}$$

which is the radiation source for CARS. Similar to SFG (9.11), the nonlinear susceptibility can be written as the sum of a non-resonant and a resonant term:

$$\overleftrightarrow{\chi}^{(3)} = \overleftrightarrow{\chi}^{(3)}_{\mathrm{NR}} + \sum_q \frac{\overleftrightarrow{A}_q}{(\omega_{\mathrm{P}} - \omega_{\mathrm{S}}) - \omega_q + i\Gamma_q} \,. \tag{9.17}$$

Under the plane-wave approximation, the CARS output intensity from a uniform sample of thickness d is given by

$$S(\omega_{\mathrm{AS}}) \propto \left|\chi^{(3)}_{\mathrm{eff}}\right|^2 I_{\mathrm{P}}^2 I_{\mathrm{S}} \sin^2(|\Delta\boldsymbol{k}|\, d/2) \,/\, |\Delta\boldsymbol{k}|^2 \,, \tag{9.18}$$

where $\Delta\boldsymbol{k} = 2\boldsymbol{k}_{\mathrm{P}} - \boldsymbol{k}_{\mathrm{S}} - \boldsymbol{k}_{\mathrm{AS}}$ and

$$\chi^{(3)}_{\mathrm{eff}} = \left(\overleftrightarrow{L}_{\omega_{\mathrm{AS}}} \cdot \hat{\boldsymbol{e}}_{\omega_{\mathrm{AS}}}\right) \cdot \overleftrightarrow{\chi}^{(3)} : \left(\overleftrightarrow{L}_{\omega_1} \cdot \hat{\boldsymbol{e}}_{\omega_1}\right)\left(\overleftrightarrow{L}_{\omega_1} \cdot \hat{\boldsymbol{e}}_{\omega_1}\right)\left(\overleftrightarrow{L}_{\omega_2} \cdot \hat{\boldsymbol{e}}_{\omega_2}\right) \,. \tag{9.19}$$

In the case of a strongly focused geometry or a heterogeneous medium, (9.18) needs to be modified, but the characteristic dependence of the CARS output on the phase matching $\Delta\boldsymbol{k}$ is still approximately true. For CARS generated in the forward direction (F-CARS), $|\Delta\boldsymbol{k}|$ is generally much smaller than $1/\lambda_{\mathrm{AS}}$. For CARS generated in the backward direction (E-CARS, where E stands for epi-detection), $|\Delta\boldsymbol{k}|$ is larger than $2\pi/\lambda_{\mathrm{AS}}$. From a thin sample, F-CARS and E-CARS, both having $|\Delta\boldsymbol{k}|d \ll \pi$, are expected to have nearly the same signal intensity. From a thick sample with $|\Delta\boldsymbol{k}|d < \pi$ for F-CARS but $|\Delta\boldsymbol{k}|d \gg \pi$ for E-CARS, F-CARS should have a much stronger output than E-CARS.

Xie and coworkers [11,80] have studied theoretically the ratio of E-CARS and F-CARS outputs in strongly focusing geometry by varying the sample thickness d.

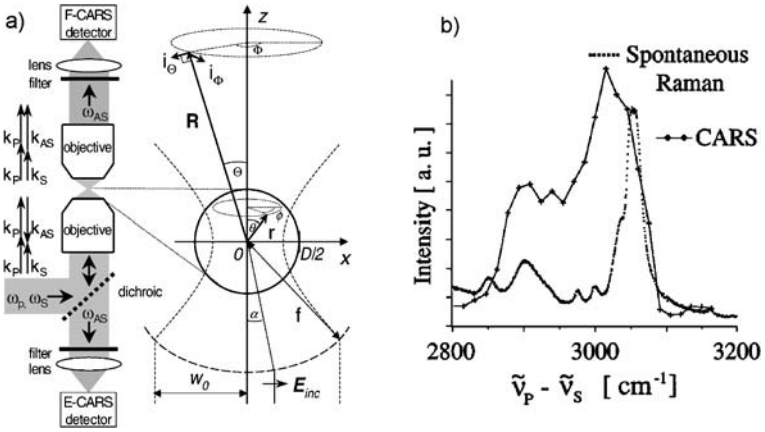


Fig. 9.13. (**a**) Optical scheme of the collinear configurations for F- and E-CARS microscopy with collinear pump and Stokes beams and confocal detection in both forward and backward directions [80]. (**b**) CARS and spontaneous Raman spectra of a 910 nm polystyrene bead. For CARS, the pump wavelength (ν_P) was fixed at 854 nm and the Stokes wavelength (ν_S) was tuned from 1.12 to 1.17 μm. The spectral resolution is estimated to 60 cm^{-1} [84]

They show that the E-CARS output reaches the relative maximum at $d = 0.65\lambda_P$. Incorporation of CARS into optical microscopy provides microspectrometric imaging of materials. Duncan and coworkers first used CARS microscopy with psec laser pulses to image biological cells soaked in deuterated water [81–83] . Asides from being able to selectively map out distributions of molecular species via their vibrational resonances, it also has a number of other advantages compared to, in particular, Raman spectroscopy. (1) Since the CARS intensity (9.18) is proportional to $I_P^2 I_S$, both the lateral and the axial resolution are $\sim \sqrt{3}$ times better than that of linear optical microscopy. For $\omega_P - \omega_S = 3053\,\text{cm}^{-1}$ (CH ring stretching mode), a lateral resolution around 300 nm and an axial resolution around 1.6 μm (even without a confocal pinhole) have been obtained [84]. (2) A moderate laser intensity and power can be used to avoid damaging of biological samples like living cells. (3) The anti-Stokes emission in CARS is spectrally separable from fluorescence as its wavelength is shorter than λ_P. (4) Transparency of the medium to the input waves allows 3-D microscopy and imaging of buried samples.

Recently, Zumbusch et al. [84] and Volkmer et al. [80] have developed F-CARS and E-CARS microscopy in the $2600 - 3300\,\text{cm}^{-1}$ region using femtosecond pulses generated by a Ti:sapphire laser and an associated optical parametric oscillator/amplifier (Fig. 9.13a). The input beams are focused through a 60× oil-immersion objective with NA=1.4. A similar objective was used to collect the signal in both forward and backward directions. The broad spectral width of femtosecond pulses limits the spectral resolution to $\sim 60\,\text{cm}^{-1}$, as seen in Fig. 9.13b. For better spectral resolution longer input pulses with a narrower spectral width are required.

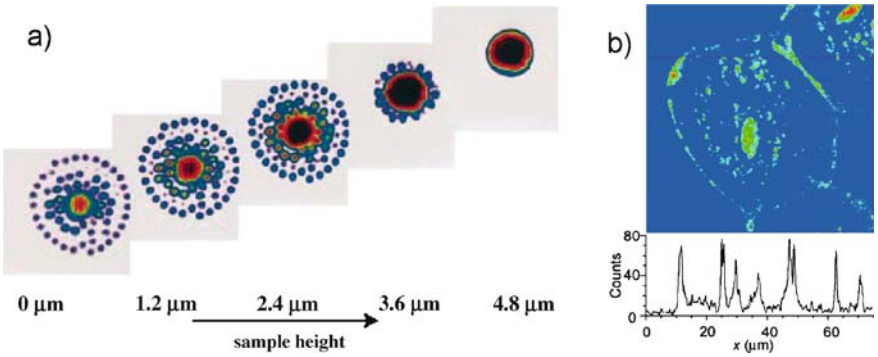


Fig. 9.14. (**a**) CARS section images ($20 \times 20\,\mu m^2$) of 910 nm polystyrene beads around a 4.3 μm bead taken at $\omega_P - \omega_S = 3038\,cm^{-1}$. Sectioning is in the z direction with 1.2 μm increments. The average powers of femtosecond pulses incident on the sample were 120 mW and 50 mW at 855 nm and 1.155 μm, respectively [84]. (**b**) E-CARS image with a Raman shift of $1570\,cm^{-1}$ of an unstained human epithelial cell in an aqueous environment (size: $75 \times 75\,\mu m^2$). The average powers of picosecond ($f = 100$ kHz) Stokes (800 nm) and pump (741 nm) pulses were 1 an 2 mW, respectively. The lateral profile of the image displays features as small as 375 nm. Figure 9.14b is reprinted with permission from [11]. Copyright (2001) American Chemical Society

The 3-D imaging capability of CARS microscopy is shown in Fig. 9.14a where images of polystyrene beads are presented. The Raman transition probed is the CH ring stretch mode of polystyrene at $3038\,cm^{-1}$. The lateral spatial resolution obtained is around 300 nm.

If the sample is immersed in a liquid, the non-resonant background contribution from the solvent can significantly distort the CARS signal [85]. In such a case, since the overall thickness of the solution with the sample is much larger than the input wavelengths, the background contribution in F-CARS is much larger than in E-CARS. In fact, because of $|\Delta \boldsymbol{k}| \gg 2\pi/\lambda_{AS}$, the background contribution in E-CARS from the solvent is largely suppressed. Figure 9.14b shows an E-CARS image taken at $\omega_P - \omega_S = 1570\,cm^{-1}$ of a human living cell ($75 \times 75\mu m^2$ in size) in an aqueous solvent. It clearly exhibits many detailed features. The total imaging time was about 8 min.

Following the idea of surface enhanced Raman scattering, Kawata et al. have demonstrated that CARS signal from molecules can be strongly enhanced if the molecules are attached to gold particles and inserted in a CARS microscope [86–88]. Combination of CARS with an AFM gold tip interacting with samples in the near field region can provide excellent spatial resolution beyond the diffraction limit of light with enhanced signal from the sample. Examples of simultaneous AFM imaging of the topography and the CARS imaging of single wall nanotubes are given in reference [86]. For maximum field enhancement, the tip has to be properly positioned with respect to the focused spot (Fig. 9.15). The enhanced electric field at the tip apex is strongly polarized along the tip axis and is effective not only to

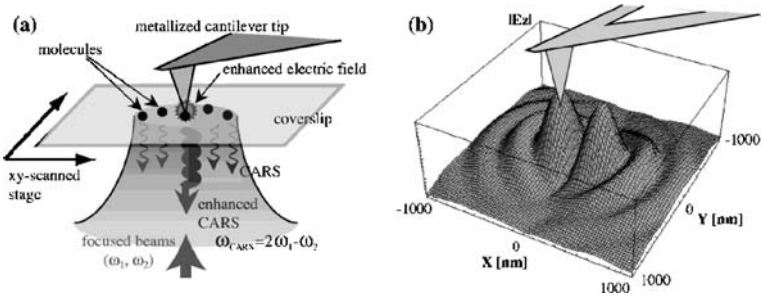


Fig. 9.15. (**a**) Schematic of tip enhanced-CARS generated from a gold coated cantilever tip adjusted onto a tightly focussed laser spot by a high NA (1.4) objective lens. (**b**) Calculated z-polarized electric field (along the tip axis) at the tightly focused spot with the excitation wavelength of 800 nm. For the electric field enhancement effect, the tip has to be adjusted onto one of the lobes of the strong z field. Reused with permission from Norihiko Hayazawa, Taro Ichimura, Mamoru Hashimoto, Yasushi Inouye, and Satoshi Kawata, Journal of Applied Physics, 95, 2676 (2004). Copyright 2004, American Institute of Physics

enhance the CARS signal but also to probe molecules oriented along the beam polarization under the tip. The strong local field enhancement makes near-field CARS microscopy possible.

9.6 Multiphoton Excited Fluorescence Microscopy

Two-photon fluorescence microscopy was first developed by Webb and coworkers for applications to biological and medical sciences [10]. Its main asset comes from the 3-D imaging possibility to probe endogenous or exogenous fluorophores [89,90] with a good lateral and axial resolution using either the scanning or the confocal scheme. The technique has brought many new research opportunities to life science and is becoming a routine microscopy tool in many laboratories. Three-photon excited fluorescence microscopy has also been demonstrated. An excellent review on the subject can be found in [91].

Multiphoton-excited fluorescence microscopy offers several advantages over confocal one-photon fluorescence microscopy [92]. First, the pump beam is now in the transparent region that allows deeper probing into the sample [12,93]. The larger difference in wavelength between the pump and the fluorescence makes spectral filtering in signal detection much easier [94,95]. The much weaker absorption of pump radiation significantly reduces thermal distortion and photobleaching hazards on a sample. Second, because of the power dependence on the excitation intensity, the fluorescence originates mainly from the tight focal region. Therefore, sectioning of planes in 3-D imaging is limited only by the focal volume, and a confocal pinhole is often not required to suppress the out-of-focus fluorescence background. Probing volumes smaller than femto-liters have been reported .

Recent advances of multiphoton-excited fluorescence microscopy have been helped by the development of efficient fluorescent dyes with large multiphoton absorption cross-sections [96]. Work has been focused on improvement of the spatial resolution and reduction of the laser power needed. We review here a few most promising techniques and their combination with NSOM.

9.6.1 Two-Photon Excited Fluorescence (TPEF) Microscopy

We consider here both far-field and near-field versions of TPEF. Far-field TPEF is already widely used for characterization of biological samples because of its 3-D imaging capability. Near-field TPEF finds its interest in surface studies of microscope systems that allow simultaneous imaging of the surface topography and the different domains tagged by specific fluorophores.

Far-field TPEF Microscopy Two-photon-excited fluorescence yield from a sample is proportional to the number of molecules excited per unit time by two-photon absorption and the fluorescence efficiency η. It can be written as [89]

$$F(t) = \eta\sigma_2 \int_V C(\boldsymbol{r}, t) I^2(\boldsymbol{r}, t) \mathrm{d}V \,, \tag{9.20}$$

where σ_2 is the two photon absorption cross-section, $C(\boldsymbol{r}, t)$ is the density of molecules in the ground state at $\boldsymbol{r}$ and t, and I is the exciting laser intensity. For common chromophores with excitation wavelength ranging from 690 nm to 1050 nm, σ_2 is about 10^{-48} to 10^{-50} cm^{4}s/photon [89]. Owing to the quadratic dependence of TPEF on laser intensity, the fluorescence in TPEF comes from a much reduced volume as compared to one-photon excited fluorescence [97]. Consider a tightly focused gaussian laser beam in a medium. If the beam attenuation in the medium is neglected, the input intensity distribution is given by

$$I(\rho, z) = \frac{2P}{\pi w_0^2(z)} \exp\left(\frac{-2\rho^2}{w_0^2(z)}\right) \,, \tag{9.21}$$

where z is in the propagation direction with $z = 0$ at the center of the focus and

$$w_0(z) = \frac{\lambda}{\pi(\mathrm{NA})} \sqrt{1 + \left(\frac{4\pi(\mathrm{NA})^2 z}{\lambda}\right)^2} \tag{9.22}$$

is the beam radius with λ being the wavelength and NA the numerical aperture. Taking the values NA = 0.45 and λ = 794 nm, the calculation shows that one-photon fluorescence comes equally from mainly from a 1.9 μm section in depth around the focal region [97]. TPEF microscopy images are often obtained with a scanning microscope. The laser beam is focused to a diffraction-limited beam waist of about 1 μm and is raster-scanned across a specimen. To improve the axial resolution,

a confocal pinhole placed in front of the detector can be used as a spatial filter to selectively detect fluorescence from a particular section plane in the focal region. However, owing to the intrinsic sectioning capability of TPEF mentioned earlier, the pinhole is not always required [98]. Because of the small two-photon absorption cross section of fluorophores, focused excitation intensity in the MW cm^{-2} to GW cm^{-2} range is often needed for detectable fluorescence. For fast scanning, CW mode-locked (femtosecond pulsed) lasers with moderate peak power but low mean power are better suited.

Typical setup can be found in references [89,97,99]. An example is given in Fig. 9.16. Femtosecond pulses from a Ti:sapphire laser with an 80 MHz repetition rate is focused on a sample with a microscope objective of high NA. TPEF is epi-collected by the same objective and sent either directly to the detector [89,99] or after being spectrally resolved using a polychromator [97]. The 3-D image can be obtained by scanning of the laser focal spot on the sample in the $x - y$ plane with a galvanometer mirror and in the z direction by translation of the microscope objective [99] or the sample mounted on a piezoelectric ceramic. The spatial resolution of TPEF microscopy is similar to that of one-photon excited fluorescence microscopy despite the difference in excitation wavelengths, with typical values of 0.3 μm laterally and 0.9 μm axially [100]. The axial resolution can be improved by up to 50% with the addition of a confocal pinhole [101].

Near-Field TPEF Microscopy Coupling of NSOM microscopy with fluorescence measurement was initially developed for single-molecule spectroscopy imaging [102] or time-resolved studies [103,104]. It has been used to probe single molecules, quantum dots and macromolecules on surfaces with a spatial resolution better than 20 nm [105–108].

Near-field TPEF microscopy has also been developed in recent years. Jenei et al. [109] used picosecond pulses from a Nd:YVO_4 laser to induce TPEF from labelling dye in a sample and an uncoated fiber tip for both excitation and collection of the fluorescent signal from the sample. Figure 9.17 shows their setup and the observed surface topography and TPEF image of a labelled mitochondria cell. The spatial res-

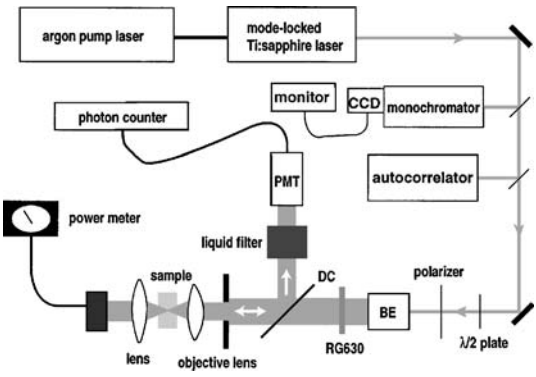


Fig. 9.16. Far-field TPEF microscopy setup. RG630, red-pass filter; BE, 5× beam expander; DC, dichroic mirror; PMT, photomultiplier tube [89]. (Copyright (1996) Optical Society of America)

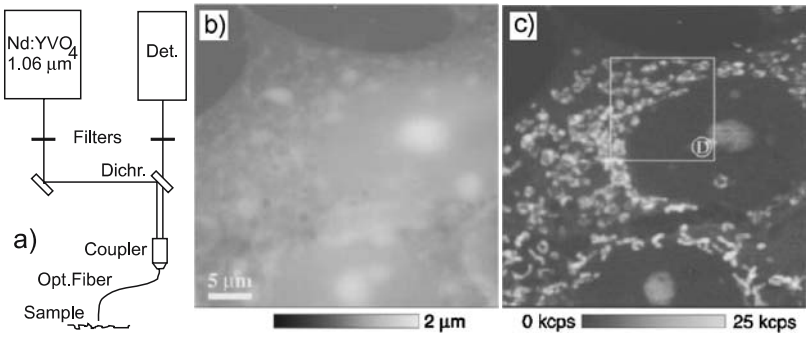


Fig. 9.17. (**a**) Optical configuration of TPEF combined with a scanning near-field optical microscope. The laser excitation of the sample is through an optical fiber tip which also collects the TPEF signal. The tip-sample distance is regulated by a shear force mechanism. (**b**) Topography of stained mitochondria cell. (**c**) TPEF image from the same area. Adapted from [109]. (Copyright (1999) Biophysical Society)

olution obtained was better than 200 nm. Similar resolution was achieved by Lewis et al. using femtosecond pulses from a Ti:sapphire laser to probe individual Rhodamine 6G molecules [110].

Hell et al. [16] and Kirsch et al. [111] showed that it is also possible to use a CW laser as the excitation source for TPEF microscopy and obtain good microscope images. An example is shown in Fig. 9.18, where both far-field and near-field TPFE images of a chromosome stained with a dye are presented. The near-field image appears to have a better spatial resolution (150 nm).

Another variation of TPEF-NSOM has been developed by Sànchez et al. [112]. They used femtosecond laser pulses to illuminate a sample and a sharp gold tip

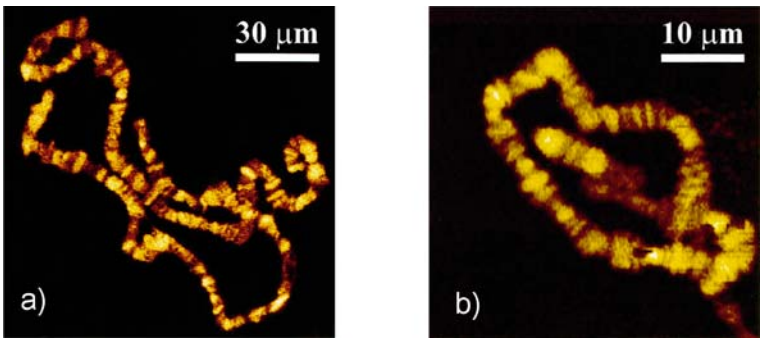


Fig. 9.18. (**a**) Far-field TPEF microscopy image of a stained chromosome using a CW ArKr laser excitation ($\lambda = 647$ nm, $P = 100$ mW, acquisition time 4 s). (**b**) Near-field TPEF microscopy image of the same stained chromosome [16]. (Copyright (1996) Optical Society of America)

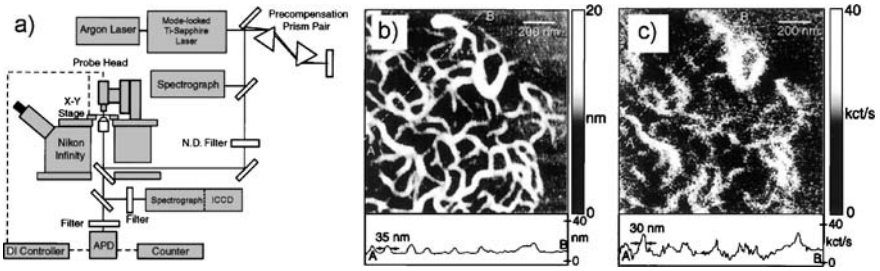


Fig. 9.19. (**a**) NSOM-TPEF set up with metallic tip excitation. The light source is a Ti:sapphire laser (λ = 830 nm, τ = 100 fs, f = 76 MHz). The beam is sent into an inverted fluorescence microscope, reflected by a dichroic beam splitter and focused by a microscope objective (NA = 1.4, 60×) on the sample surface. The metal tip is centered onto the focal spot. The TPEF is collected by the same objective lens and detected either by an avalanche photodiode or analyzed by a spectrometer in conjunction with a CCD camera. Simultaneous topography image (**b**) and near-field TPEF image (**c**) of J-aggregates of pseudoisocyanine dye embedded in polyvinyl sulfate were obtained using the apparatus [112]

near the surface to locally enhance the optical field at the sample (Fig. 9.19a). Such a local field enhancement can be very large and highly localized. In TPEF-NSOM, it strongly enhances the two-photon excitation and hence fluorescence from the sample area under the tip. (The same concept can be applied to other linear and nonlinear optical imaging techniques). Operating in both AFM and NSOM modes, the apparatus allows simultaneous imaging of surface topography and fluorescence microscopy. An example shown in Fig. 9.19b and c reveals features in fragments of photosynthetic membranes as well as J-aggregates with a resolution of 20 nm that roughly corresponds to the tip apex. As pointed out by the authors, the use of a metal tip a few nanometer away from a fluorophore may quench the fluorescence. Therefore, proper fluorophores with rapid energy transfer should be used in such a technique. Kawata et al. have studied theoretically the optical field enhancement from metallic and dielectric tips of various shapes. Laser heating of the tips does not appear to be important. The lateral resolution of such an apertureless near-field microscope is estimated to be in the nanometer range [113].

9.6.2 TPEF Far-Field Microscopy Using Multipoint Excitation

Introduced by Hell et al. [114,115] and Buist et al. [116], parallel microscopy imaging using a two-dimensional microlens array has been proved to be a powerful instrument for fast image acquisition without increasing excitation intensity. This is possible because TPEF is confined to the small focal volumes. The microlens array splits an incident laser beam into beamlets that separately focus into the sample; the number of focal points equals to the number of microlenses. Fluorescent signals from the focal regions are collected by a 2-D photodetection system. Shown in Fig. 9.20 is the apparatus developed by Bewersdorf et al. that permits real-time 3-D imaging with high efficiency and resolution [115]. The expanded beam from

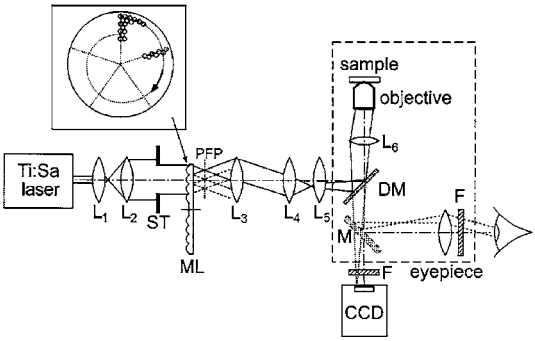


Fig. 9.20. Optical setup of a multifocal multipoint microscope for real-time direct-view TPEF microscopy. L: lenses, ML: microlens disk, M: mirror, DM: dichroic mirror. The inset shows the array of a spiral arrangement of microlenses on the disk [115]. (Copyright (1996) Optical Society of America)

a femtosecond laser illuminates a section of microlenses etched on a disk each having a diameter of 460 μm and a focal length of 6 mm. The overall arrangement of the microlenses on the disk form a pattern of spirals with 10 rows. The multiple beamlets pass through intermediate optics and form an array of independent foci on the sample with the help of an objective lens that is also used for collection and mapping of fluorescence from the multiple foci onto a CCD camera. Rotation of the microlens disk allows for scanning of foci in a sample plane. The image acquisition speed depends on the number of lenses arranged in rows and the rotation frequency of the disk. Bewersdorf et al. have used an acquisition speed 40 to 100 times faster than that of a single-beam scanning TPEF, permitting real-time TPEF imaging of living cells. The spatial resolution they obtained was similar to those of conventional TPEF microscope with an axial resolution of about 0.84 μm when an oil immersion objective lens was used.

9.6.3 4-Pi Confocal TPEF Microscopy

4-Pi confocal microscopy is 3-D TPEF microscopy technique developed to improve the axial resolution of far-field microscopy [117]. An axial resolution of about 100 nm has been demonstrated [118]. The scheme involves excitation of the sample by counter-propagating beams through two objective lenses with high numerical aperture (Fig. 9.21a). The overall excitation aperture of the system would approach 4π if the aperture extended by each lens were 2π. The acronym of the system 4-Pi was chosen as a reminder of this arrangement.

The two counter-propagating beams interfering in the focal region should yield an intensity distribution with a main peak several times narrower in the axial direction than that obtainable by a single focused beam. Furthermore, fluorescence from secondary interference peaks (axial lobes) along the axis in the focal region can be significantly reduced by a confocal pinhole (Fig. 9.21a) and by a deconvolution procedure using a restoration algorithm [119]. Figure 9.21b shows as example

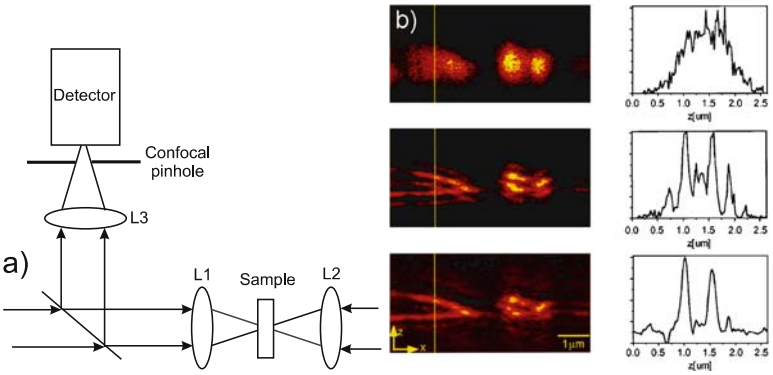


Fig. 9.21. (**a**) Setup of the TPEF 4-Pi confocal microscope. The numerical aperture of the 100× oil immersion objectives is 1.4 [119]. (**b**) Microscopy images obtained with various schemes: two photon confocal (*top*), two photon 4-Pi confocal (*center*), two photon 4-Pi confocal after side-lobe removal (*bottom*). The right column shows the intensity profiles of the fluorescence along the lines indicated in the respective images [118]. (Copyright (1999) Biophysical Society)

axial images of a fibroblast cell using various schemes (confocal, 4-Pi confocal, 4-Pi confocal with axial lobes removed). The excitation source was a mode-locked femtosecond Ti:sapphire. The 4-Pi confocal microscopy image with axial lobes suppressed obviously has the best axial resolution. It reveals details on a scale less than 200 nm.

9.6.4 Simultaneous SHG/TPEF Microscopy

SHG and TPEF can be simultaneously generated from a sample by a focused beam and detected in the same microscopy system. Both processes have an output proportional to the square of the input laser intensity, but one is coherent and the other incoherent. If they are pumped by the same laser input, then, as shown in Fig. 9.22a, both experience the same two-photon resonant excitation. However, SHG is surface-specific in media with inversion symmetry and TPEF is not. Thus the two processes can provide complementary information about the sample. Lewis and coworkers [38,39] and Moreaux et al. [37,99,120], have used combined SHG/TPEF microscopy to study biological membranes. In the setup of Moreaux et al., femtosecond pulses from a Ti:sapphire were used for excitation of a sample through a microscope objective. The SHG output in the forward direction was collected by a lens while TPEF was epi-collected with the help of a dichroic mirror. Three dimensional SHG and TPEF images were obtained by scanning the laser focal spot in the $x-y$ plane with galvanometer mirrors and in the axial direction by translation of the objective. Presented in Figs. 9.22(b) and (c) are vesicles images acquired in 1.5 s with an excitation power less than 1 mW [99]. They provide an excellent example of complementarity of SHG and TPEF: the adhesion between the two vesicles appears to be centrosymmetric as it contributes to TPEF but not to SHG.

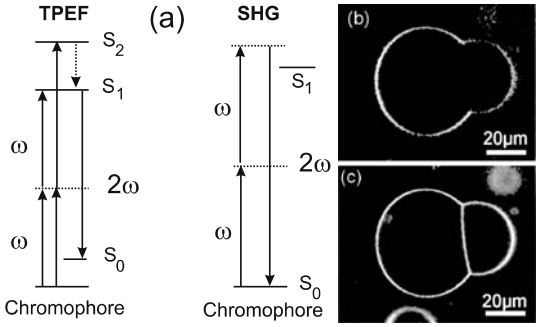


Fig. 9.22. (**a**) Level diagrams describing two-photon excited fluorescence and second harmonic generation. (**b**) SHG and (**c**) TPFE images of labelled vesicles excited by laser input at $\lambda = 880\,\text{nm}$ (τ=80 fs, f=80 MHz, $P \leq 1\,\text{mW}$) [99]. (Copyright (1996) Optical Society of America)

9.6.5 Three-Photon-Excited Fluorescence Microscopy

Three-photon-excited fluorescence microscopy has also been demonstrated. The cubic dependence of fluorescence on the local input laser intensity allows further improvement of the spatial resolution as compared to TPEF. For samples such as amino acids, proteins and neurotransmitters that are susceptible to photodamage by one-photon absorption in the ultra-violet and residual absorption in the visible-red range, the three-photon-excited fluorescence scheme with a near-infrared input beam could avoid the damage and seems to be most suitable for microscopy. Using femtosecond Ti:sapphire laser pulses for excitation, Gryczynski et al. successfully imaged the triptophan residues [121] and Maiti et al. imaged neurotransmitters in living cells [95].

A lateral and axial resolution of ~0.2 and ~0.6 μm has been achieved [90]. The main disadvantage of the technique is the narrow window of operation. Because the three-photon absorption cross-section of a sample is generally very small (10^{-75} – $10^{-84}\,\text{cm}^6\,\text{s}^2\,\text{photon}^{-2}$), high peak laser intensity is required for excitation that likely causes damage of the sample.

9.6.6 Stimulated-Emission-Depletion (STED) Fluorescence Microscopy

Recently, Klar et al. proposed and demonstrated an interesting idea that can significantly improve the spatial resolution of fluorescence microscopy [122]. It involves fluorescence quenching of excited molecules at the rim of the focal spot through stimulated emission thus significantly reducing the focal volume that emits fluorescence. This is accomplished by passing the quenching laser beam through an optical phase plate, yielding a focused wave front that produces by destructive interference a central minimum at the focal point. In the experiment (Fig. 9.23a), two synchronized pulses from a Ti:sapphire laser/optical parametric oscillator system with a 76 Mhz repetition rate were used: a visible pulse with a 0.2 ps pulsewidth for excitation followed by a near-infrared pulse with a 40 ps pulsewidth for fluorescence quenching. The visible pulse at $\lambda = 560\,\text{nm}$ originated from the optical parametric oscillator with an intracavity frequency doubler and the near-infrared pulse at $\lambda = 765\,\text{nm}$ came directly from the laser. The fluorescence was epi-collected by an

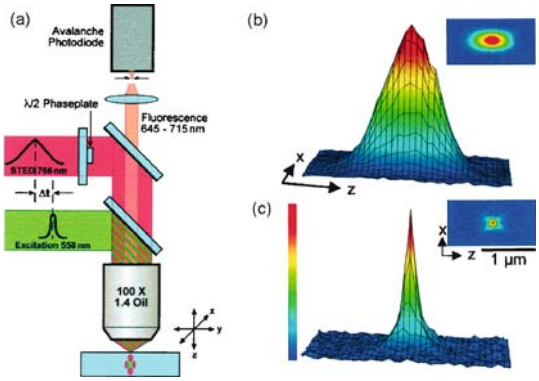


Fig. 9.23. (**a**) STED microscope setup. Excitation pulse is followed by fluorescence quenching pulse. Fluorescence is then detected after passing through dichroic filters and a confocal pinhole. (**b**) and (**c**), Intensity distributions of fluorescence along x and z (axial direction) in confocal and in confocal STED microscopy, respectively. The latter has an axial width of 97 nm which is 5 times narrower than in (**b**) [122] (Copyright (2000) National Academy of Sciences, U.S.A.)

avalanche photodiode after passing through a confocal pinhole. The lateral and axial resolutions obtained were ~100 nm, which is an improvement of a factor 5 in the axial direction and 2 in the lateral one, compared to ordinary confocal microscope (Fig. 9.23(b) and (c)).

Besides the improvement on spatial resolution which compete with most near-field optical microscopes, STED microscopy can also find other important applications in studies of ultrafast dynamics. The ultrafast control of fluorescence under microscopy conditions opens new opportunities for transient microscopy and spectroscopy of nano-objects. It should also have interesting perspectives in single molecule spectroscopy such as control of individual excited molecules with femtosecond time resolution [123].

9.7 Conclusion

In this contribution, we have briefly surveyed the field of nonlinear optical microscopy and presented a few examples of contemporary applications. The survey is by no means complete. The field is still at its infant stage. Further development of the techniques and their applications can be anticipated. As in the past, future advances would benefit from having better lasers, better optics and optical systems and better photodetectors. For example, an efficient high-speed detection and data acquisition system incorporated with an optimized pulse laser excitation scheme would allow *in-situ* probing of time-dependent surface reactions on nanostructures or *in-vivo* study of biological cells. Compared to linear-optics, nonlinear optics has the advantage that it probes a larger domain of material properties. SHG and SFG,

for example, are unique in their abilities to probe selectively surfaces and interfaces. However, nonlinear optics also has the disadvantages that it generally requires stronger input laser intensity and the effects are more complex. The latter point is important for interpretation of the observed microscopy images [124–127]. In particular, a clear understanding of the effects in near-field microscopy is essential for the future progress of nonlinear NSOM.

At the time of this updated edition, it is remarkable that a large number of biophysical processes are under investigation using nonlinear optical processes measured with a microscope. SHG or higher multiphotonic processes such as or THG and CARS are clearly revealing their assets in imaging biological materials [129–131]. Although a good understanding of nonlinear optical images has yet to be explored, new prospects in the field of chirality [128,47] or signal enhancement in the near field region [113], offer great challenges and promises.

Acknowledgements

Work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Science Division, of the US Departement of Energy under contract No.DE–AC03–765F00098. F.L.L. gratefully acknowledges the Centre National de la Recherche Scientifique (C.N.R.S., France) for his support during his stay at the University of California at Berkeley.

References

1. P.T.C. So: Opt. Express **3**, 312 (1998)
2. R.D. Schaller, J.C. Johnson, K.R. Wilson, L.F. Lee, L.H. Haber, R.J. Saykally,: J. Phys. Chem. B **106**, 5143 (2002)
3. X. Zhuang, P.B. Miranda, D. Kim, Y.R. Shen: Phys. Rev. B **59**, 12632 (1999)
4. X. Wei, S.-C. Hong, X. Zhuang, T. Goto, Y.R. Shen: Phys. Rev. E **62**, 5160 (2000)
5. V. Vogel, Y.R. Shen: Annu. Rev. Mater. Sci. **21**, 515 (1991)
6. M. Flörsheimer: Phys. Status Solidi (a) **173**, 15 (1999)
7. G.J. Simpson: Appl. Spectrosc. **55**, 16 (2001)
8. J.A. Squier, M. Müller, G.J. Brakenhoff, K.R. Wilson: Opt. Express **3**, 315 (1998)
9. E.O. Potma, W.D. de Boej, D.A. Wiersma: J. Opt. Soc. Am. B **17**, 1678 (2000)
10. W. Denk, J.H. Strickler, W.W. Webb: Science **248**, 73 (1990)
11. J.-X. Cheng, A. Volkmer, L.D. Book, X.S. Xie: J. Phys. Chem. B **105**, 1277 (2001)
12. K. König: J. Microsc. – Oxford **200**, 83 (2000)
13. S.I. Bozhevolnyi, T. Geisler: J. Opt. Soc. Am. A **15**, 2156 (1998)
14. Y. Shen, J. Swiatkiewicz, J. Winiarz, P. Markowicz, P.N. Prasad: Appl. Phys. Lett. **77**, 2946 (2000)
15. R.D. Schaller, J.C. Johnson, R.J. Saykally: Anal. Chem. **72**, 5361 (2000)
16. S.W. Hell, M. Booth, S. Wilms, C.M. Schnetter, A.K. Kirsch, D.J. Arndt-Jovin, T.M. Jovin: Opt. Lett. **23**, 1238 (1998)
17. Y.R. Shen: *The Principles of Nonlinear Optics* (Wiley, New York 1984)

18. R. Boyd: *Nonlinear Optics* (Academic, New York 1992)
19. M. Oh-e, S.C. Hong, Y.R. Shen: J. Phys. Chem. B **104**, 7455 (2000)
20. S.-C. Hong, M. Oh-e, X. Zhuang, Y.R. Shen, J.J. Ge, F.W. Harris, S.Z.D. Cheng: Phys. Rev. E **63**, 517061 (2001)
21. R. Hellwarth, P. Christensen: Appl. Optics **14**, 247 (1975)
22. G.T. Boyd, Y.R. Shen, T.W. Hänsch: Opt. Lett. **11**, 97 (1986)
23. K.A. Schultz, E.G. Seebauer: J. Chem. Phys. **97**, 6958 (1992)
24. K.A. Schultz, I.I. Suni, E.G. Seebauer: J. Opt. Soc. Am. B **10**, 546 (1993)
25. M.S. Johal, A.N. Parikh, Y. Lee, J.L. Casson, L. Foster, B.I. Swanson, D.W. McBranch, D.Q. Li, J.M. Robinson: Langmuir **15**, 1275 (1999)
26. N. Kato, K. Saito, Y. Uesu: Thin solid films **335**, 5 (1999)
27. S.B. Bakiamoh, G.J. Blanchard: Langmuir **17**, 3438 (2001)
28. L. Smilowitz, Q.X. Jia, X. Yang, D.Q. Li, D. McBranch, S.J. Buelow, J.M. Robinson: J. Appl. Phys. **81**, 2051 (1997)
29. Y. Sonoda, G. Mizutani, H. Sano, S. Ushioda, T. Sekiya, S. Kurita: Jap. J. Appl. Phys. **39**, L253 (2000)
30. H. Sano, T. Shimizu, G. Mizutani, S. Ushioda: J. Appl. Phys. **87**, 1614 (2000)
31. K. Pedersen, S.I. Bozhevolnyi, J. Arentoft, M. Kristensen, C. Laurent-Lund: J. Appl. Phys. **88**, 3872 (2000)
32. I.I. Suni, E.G. Seebauer: J. Chem. Phys. **100**, 6772 (1994)
33. S. Kurimura, Y. Uesu: J. Appl. Phys. **81**, 369 (1996)
34. F. Rojo, F. Agulló-lópez, B. del Rey, T. Torres: J. Appl. Phys. **84**, 6507 (1998)
35. S.E. Kapphan: J. Lumin. **83**, 411 (1999)
36. I. Freund, M. Deutsch: Opt. Lett. **11**, 94 (1986)
37. L. Moreaux, O. Sandre, S. Charpak, M. Blanchard-Desce, J. Mertz: Biophys. J. **80**, 1568 (2001)
38. P.J. Campagnola, M.-D. Wei, A. Lewis, L.M. Loew: Biophys. J. **77**, 3341 (1999)
39. C. Peleg, A. Lewis, M. Linial, M. Loew: P. Natl. Acad. Sci. USA **96**, 6700 (1999)
40. M. Flörsheimer, D.H. Jundt, H. Looser, K. Sutter, M. Küpfer, P. Günter: Ber. Bunsenges. Phys. Chem. **9**, 521 (1994)
41. M. Flörsheimer, M. Bösch, Ch. Brillert, M. Wierschem, H. Fuchs: J. Vac. Sci. Technol. B **15**, 1564 (1997)
42. M. Flörsheimer, M. Bösch, Ch. Brillert, M. Wierschem, H. Fuchs: Supramol. Sci. **4**, 255 (1997)
43. M. Flörsheimer, C. Brillert, H. Fuchs: Langmuir **15**, 5437 (1999)
44. D.M.P. Hoffman, K. Kuhnke, K. Kern: Rev. Sci. Instrum. **73**, 3221 (2002)
45. K. Kuhnke, D.M.P. Hoffman, X.C. Wu, A.M. Bittner, K. Kern: Appl. Phys. Lett. **73**, 3221 (2002)
46. K. Cimatu, S. Baldelli: J. Phys. Chem. B **110**, 1807 (2006)
47. N. Ji, K. Zhang, H. Yang, Y.R. Shen: J. Am. Chem. Soc. **128**, 3482 (2006)
48. Y. Guo, P.P. Ho, H. Savage, D. Harris, P. Sacks, S. Schantz, F. Liu, N. Zhadin, R.R. Alfano: Opt. Lett. **22**, 1323 (1997)
49. R. Gauderon, P.B. Lukins, C.J.R. Sheppard: Opt. Lett. **23**, 1209 (1998)
50. M. Cernusca, M. Hofer, G.A. Reider: J. Opt. Soc. Am. B **15**, 2476 (1998)
51. G. Berkovic, Y.R. Shen, G. Marowsky, R. Steinhoff: J. Opt. Soc. Am. B **6**, 205 (1989)
52. G.A. Reider, M. Cernusca, M. Hofer: Appl. Phys. B. – Lasers O. **68**, 343 (1999)
53. P. Rechsteiner, J. Hulliger, M. Flörsheimer: Chem. Mater. **12**, 3296 (2000)
54. I.I. Smolyaninov, A.V. Zayats, C.C. Davis: Phys. Rev. B **56**, 9 (2)901997
55. I.I. Smolyaninov, A.V. Zayats, C.C. Davis: Opt. Lett. **22**, 1 (5)921997

56. I.I. Smolyaninov, C.H. Lee, C.C. Davis: J. Microsc. – Oxford **194**, 4 (2)6 1999
57. S.I. Bozhevolnyi, B. Vohnsen, K. Pedersen: Opt. Commun. **150**, 49 (1998)
58. R.D. Schaller, R.J. Saykally, Y.R. Shen, F. Lagugné-Labarthet: Opt. Lett. **28**, 1296 (2003)
59. F. Lagugné-Labarthet, C. Sourisseau, R.D. Schaller, R.J. Saykally, R. Rochon: J. Phys. Chem. B **108**, 17059 (2004)
60. K. Karrai, R.D. Grobber: Appl. Phys. Lett. **66**, 1842 (1995)
61. B. Vohnsen, S.I. Bozhevolnyi: J. Microsc. – Oxford **202**, 244 (2001)
62. A. Liu, G.W. Bryant: Phys. Rev. B **59**, 2245 (1999)
63. S.I. Bozhevolnyi, V.Z. Lozovski: Phys. Rev. B **16**, 11139 (2000)
64. S.I. Bozhevolnyi, V.Z. Lozovski, K. Pedersen, J.M. Hvam: Phys. Status Solidi(a) **175**, 331 (1999)
65. A.V. Zayats, T. Kalkbrenner, V. Sandoghdar, J. Mlynek: Phys. Rev. B **61**, 4545 (2000)
66. Z.-Y. Li, B.-Y. Gu, G.Z. Yang: Phys. Rev. B **59**, 12622 (1999)
67. P.B. Miranda, Y.R. Shen: J. Phys. Chem. B **103**, 3292 (1999)
68. X. Wei, P.B. Miranda, Y.R. Shen: Phys. Rev. Lett. **86**, 1554 (2001)
69. R.D. Schaller, R.J. Saykally: Langmuir **17**, 2055 (2001)
70. B. Humbert, J. Grausem, A. Burneau, M. Spajer, A. Tadjeddine: Appl. Phys. Lett. **78**, 135 (2001)
71. Y. Barad, H. Eisenberg, M. Horowitz, Y. Silberberg: Appl. Phys. Lett. **70**, 922 (1997)
72. T.Y.F. Tsang: Phys. Rev. A **52**, 4116 (1995)
73. D. Yelin, Y. Silberberg, Y. Barad, J.S. Patel: Appl. Phys. Lett. **74**, 3107 (1999)
74. C.-K. Sun, S.-W. Chu, S.-P. Tai, S. Keller, U.K. Mishra, S.P. DenBaars: Appl. Phys. Lett. **77**, 2331 (2000)
75. M. Müller, J. Squier, C.A. De Lange, G.J. Brakenhoff: J. Microsc. – Oxford **197**, 150 (2000)
76. M. Müller, J. Squier, K.R. Wilson, G.J. Brakenhoff: J. Microsc. – Oxford **191**, 266 (1998)
77. L. Canioni, S. Rivet, L. Sarger, P. Barille, P. Vacher, P. Voisin: Opt. Lett. **26**, 515 (2001)
78. J.A. Squier, M. Müller: Appl. Optics **38**, 5789 (1999)
79. G.L. Eesley: *Coherent Raman Spectroscopy* (Pergamon, New York 1981)
80. A. Volkmer, J.-X. Cheng, X.S. Xie: Phys. Rev. Lett. **87**, 23901 (2001)
81. M.D. Duncan, J. Reintjes, T.J. Manuccia: Opt. Lett. **7**, 350 (1982)
82. M.D. Duncan, J. Reintjes, T.J. Manuccia: Opt. Commun. **50**, 307 (1984)
83. M.D. Duncan, J. Reintjes, T.J. Manuccia: Opt. Express **24**, 352 (1985)
84. A. Zumbusch, G.R. Holtom, X.S. Xie: Phys. Rev. Lett. **82**, 4142 (1999)
85. M. Hashimoto, T. Araki, S. Kawata: Opt. Lett. **25**, 1768 (2000)
86. T. Ichimura, N. Hayazawa, M. Hashimoto, Y. Inouye, S. Kawata: J. Raman. Spectrosc. **34**, 651 (2003)
87. N. Hayazawa, T. Ichimura, M. Hashimoto, Y. Inouye, S. Kawata: J. Appl. Phys. **95**, 2676 (2004)
88. S. Kawata, T. Ichimura, N. Hayazawa,Y. Inouye, M. Hashimoto : J. of Nonlinear Optical Physics and Materials **13**, 593 (2004)
89. C. Xu, W.W. Webb: J. Opt. Soc. Am. B **13**, 481 (1996)
90. C. Xu, W. Zipfel, J.B. Shear, R.M. Williams, W.W. Webb: P. Natl. Acad. Sci. USA **93**, 10763 (1996)
91. J.R. Lakowicz: *Topics in Fluorescence Spectroscopy: Nonlinear and Two-Photon Induced Fluorescence* (Plenum, New York 1997)
92. D.W. Piston, D.R. Sandison, W.W. Webb: SPIE **1640**, 379 (1992)

93. J. Ying, F. Liu, R.R. Alfano: Appl. Optics **39**, 509 (2000)
94. J.R. Lakowicz and I. Gryczynski, H. Malak, P. Schrader, P. Engelhardt, H. Kano, S.W. Hell: Biophys. J. **72**, 567 (1997)
95. S. Maiti, J.B. Shear, R.M. Williams, W.R. Zipfel, W.W. Webb: Science **275**, 530 (1997)
96. M.Barzoukas, M. Blanchard-Desce: J. Chem. Phys. **113**, 3951 (2000)
97. S. Andersson-Engels, I. Rokahr, J. Carlsson: J. Microsc. – Oxford **176**, 195 (1994)
98. R. Gauderon, P.B. Lukins, C.J.R. Sheppard: Microsc. Res. Techniq. **47**, 210 (1999)
99. L. Moreaux, O. Sandre, M. Blanchard-Desce, J. Mertz: Opt. Lett. **25**, 320 (2000)
100. W. Denk, D.W. Piston, W.W. Webb: 'Two-photon molecular excitation in laser-scanning microscopy', In: *Handbook of Biological Confocal Microscopy*, ed. by J.B. Pawley (Plenum, New York 1995) pp. 445–458
101. E. Stelzer, S. Hell, S. Lindek, R. Stricker, R. Pick, C. Storz, G. Ritter, N. Salmon: Opt. Commun. **104**, 223 (1994)
102. E. Betzig, R.J. Chichester: Science **262**, 1422 (1993)
103. X.S. Xie, R.C. Dunn: Science **265**, 361 (1994)
104. G. Parent, D. Van Labeke, F.I. Baida: J. Microsc. – Oxford **202**, 296 (2001)
105. R. Zenobi, V. Deckert: Angew. Chem. Int. Ed. **39**, 1746 (2000)
106. P.K. Yang, J.Y. Huang: Opt. Commun. **173**, 315 (2000)
107. H. Muramatsu, K. Homma, N. Yamamoto, J. Wang, K. Sakata-Sogawa, N. Shimamoto: Materials Science and Engineering C **12**, 29 (2000)
108. N. Hosaka, T. Saiki: J. Microsc. – Oxford **202**, 362 (2001)
109. A. Jenei, A.K. Kirsch, V. Subramaniam, D.J. Arndt-Jovin, T.J. Jovin: Biophys. J. **76**, 1092 (1999)
110. M.K. Lewis, P. Wolanin, A. Gafni, D.G. Steel: Opt. Lett. **23**, 1111 (1998)
111. A.K. Kirsch, V. Subramaniam, G. Striker, C. Schnetter, D.J. Arndt-Jovin, T.M. Jovin: Biophys. J. **75**, 1513 (1998)
112. E.J. Sánchez, L. Novotny, X.S. Xie: Phys. Rev. Lett. **82**, 4014 (1999)
113. Y. Kawata, C. Xu, W. Denk: J. Appl. Phys. **85**, 1294 (1999)
114. S.W. Hell, V. Andresen: J. Microsc. – Oxford **202**, 457 (2001)
115. J. Bewersdorf, R. Pick, S.W. Hell: Opt. Lett. **23**, 655 (1998)
116. A.H. Buist, M. Müller, J. Squier, G.J. Brakenhoff: J. Microsc. – Oxford **192**, 217 (1998)
117. P.E. Hänninen, S.W. Hell, J. Salo, E. Soini: J. Appl. Phys. **66**, 1698 (1995)
118. M. Schrader, K. Bahlmann, G. Giese, S.W. Hell: Biophys. J. **75**, 1659 (1998)
119. M. Schrader, S.W. Hell, H.T.M. van der Voort: J. Appl. Phys. **84**, 4033 (1998)
120. L. Moreaux, O. Sandre, J.W. Mertz: J. Opt. Soc. Am. B **17**, 1685 (2000)
121. I. Gryczynski, H. Malak, J.R. Lakowicz: Biospectroscopy **2**, 9 (1996)
122. T.A. Klar, S. Jakobs, M. Dyba, A. Egner, S.W. Hell: P. Natl. Acad. Sci. USA **97**, 8206 (2000)
123. M. Dyba, T.A. Klar, S. Jakobs, S.W. Hell: Appl. Phys. Lett. **77**, 597 (2000)
124. T. Pons, J. Mertz: Opt. Commun. **258**, 203 (2006)
125. R.M. Williams, W.R. Zipfel, W.W. Webb: Biophys. J. **88**, 1377 (2005)
126. D. Debarre, W. Supatto, E. Beaurepaire : Opt. Lett. **30**, 2134 (2005)
127. J.M. Schins, M. Müller: Appl. Optics **41**, 1995 (2002)
128. A.-M. Pena, T. Boulesteix, T. Dartigalongue, M.-C. Schanne-Klein: J. Am. Chem. Soc. **127**, 10314 (2005)
129. T. Bouleteix, E. Beaurepaire, M.-P. Sauviat, M.-C. Schanne-Klein: Opt. Lett. **29**, 2031 (2004)
130. D. Debarre, W. Supatto, E. Farge, B. Moulia, M.-C. Schanne-Klein, E. Beaurepaire: Opt. Lett. **29**, 2881 (2004)
131. E.O. Potma, X.S. Xie: J. Raman Spectrosc. **34**, 642 (2003)

10 Parametric Nonlinear Optical Techniques in Microscopy

M. Müller and G.J. Brakenhoff

10.1 Introduction

A small revolution is taking place in microscopy. Nonlinear optical spectroscopic techniques – which in bulk applications have been around almost since the invention of the laser in the sixties – are rapidly being incorporated within high resolution microscopy. Starting with the successful introduction of two-photon absorption microscopy in 1990 [1], which has found important applications in biology, a whole series of other nonlinear optical techniques have been introduced: Three-photon absorption, Second Harmonic Generation (SHG), Third Harmonic Generation (THG), Coherent Anti-Stokes Raman Scattering (CARS), the optical Kerr effect, ...

The startling progress of laser technology has enabled this new field in microscopy. Ultrashort pulses – which combine high peak power with moderate average powers – are required for efficient application of nonlinear optics at high numerical aperture (NA) conditions. Also, the lasers need to combine a low noise level of operation with high power output, "turn-key" operation and tunability over a large wavelength range. All these features are now available in current "state-of-the-art" laser technology.

A number of features make the introduction of nonlinear optical techniques into microscopy particularly interesting. The nonlinear dependence of the specimen's response to the incident laser power, provides inherent optical sectioning and thus three-dimensional imaging capability. A consequence of the fact that the interaction is limited to the focal region only, is the reduction of detrimental out-of-focus interactions, such as photobleaching and photo-induced damage. The use of nonlinear optics often enables the use of longer wavelengths, which reduces scattering and is – in some cases – less harmful to e.g. biological specimen. Another important reason for using nonlinear optics in microscopy is its possibility for spatially resolved spectroscopic measurements. Nonlinear optical spectroscopy, particularly in the form of four-wave mixing, can provide a wealth of information in both the time and frequency domain through the intricate interactions with the molecular energy levels.

Of particular interest here is the class of parametric nonlinear optical processes. Parametric processes are those in which during the interaction no net energy transfer occurs from the laser field to the sample, and *vice versa*. This absence of energy transfer makes these techniques potentially noninvasive imaging modes. In addition,

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

because of the absence of photo-induced damage to the sample, no fading of contrast takes place.

This chapter is organised as follows. We start out with a brief review of the theory of nonlinear optics in general and of parametric processes in particular. This highlights the similarities of the different techniques, while at the same time indicating the differences in response generation. All the different techniques probe different parts and orders of the nonlinear susceptibility of the material, thereby providing detailed information on molecular structure and – in some cases – intermolecular interactions. In the next chapters we consider two nonlinear optical parametric techniques in more detail: third harmonic generation and coherent anti-Stokes Raman scattering. These two techniques provide widely different information on the sample. The former can be applied to image – and analyse – changes in refractive index and/or nonlinear susceptibility in transparent materials. The latter can be used as a general "chemical microscope": the contrast is based on molecular structure and molecular interactions. We conclude with some general remarks with respect to the use nonlinear parametric processes in high numerical aperture (NA) microscopy.

10.2 Nonlinear Optics – Parametric Processes

The field of nonlinear optics is concerned with all light-matter interactions in which the response of the material depends in a nonlinear fashion on the input electromagnetic field strength. Excellent text books exist (e.g. [2–5]) that provide both a general overview of the field and a detailed theoretical analysis. Here some of the main features are reviewed which are relevant for the application of nonlinear optical techniques to high resolution microscopy. The discussion is limited to so-called parametric processes, which are processes for which the initial and final quantum-mechanical states of the system are equal. These processes include phenomena like harmonic generation and stimulated Raman scattering.

10.2.1 Introduction

The standard approach in nonlinear optics is to split the macroscopic polarisation – i.e. the result of an ensemble of oscillating dipole moments – in a linear and a nonlinear part and to expand it in successive orders of the electromagnetic field ($\boldsymbol{E}$):

$$\begin{aligned}\boldsymbol{P}(\boldsymbol{r},t) &= \boldsymbol{P}_{\mathrm{L}}(\boldsymbol{r},t) + \boldsymbol{P}_{\mathrm{NL}}(\boldsymbol{r},t) \\ &= \overleftrightarrow{\chi}^{(1)} : \boldsymbol{E} + \overleftrightarrow{\chi}^{(2)} : \boldsymbol{EE} + \overleftrightarrow{\chi}^{(3)} : \boldsymbol{EEE} + \ldots \,. \end{aligned} \tag{10.1}$$

In (10.1), $\overleftrightarrow{\chi}^{(1)}$ and $\overleftrightarrow{\chi}^{(n)}$ denote the linear and nonlinear susceptibility respectively, which relate the complex amplitude of the electromagnetic fields with the polarisation. For example, for the third-order susceptibility:

$$P_i = \sum_{jkl} \chi^{(3)}_{ijkl} E_j E_k E_l \,, \tag{10.2}$$

where i, j, k and l run over all Cartesian coordinates. In general, for a material with dispersion and/or loss, the susceptibility is a complex tensor quantity. Parametric processes however can always be described by a real susceptibility. Note that from symmetry arguments it follows immediately that even orders of the nonlinear susceptibility – $\overleftrightarrow{\chi}^{(2n)}$ – vanish for materials that possess a centre of inversion symmetry. Various approaches have been developed to relate the macroscopic susceptibility to microscopic quantities (see e.g. [6,7]).

The nonlinear optical signal intensity follows from the macroscopic polarisation that is induced by a specific order of interaction in the expansion of (10.2). The macroscopic polarisation $\boldsymbol{P}(\boldsymbol{r}, t)$ acts as a source term in the wave equation, producing an electromagnetic signal wave. Accordingly, the signal intensity is proportional to the complex square of the field. As an example consider the process of third harmonic generation (THG). This process is depicted schematically in an energy level diagram in Fig. 10.1b. Three electromagnetic waves of fundamental frequency ω_f interact non-resonantly and simultaneously with the material to produce a signal wave at the third harmonic frequency $\omega_\mathrm{th} = 3\omega_\mathrm{f}$. Thus the THG process is related to the third-order susceptibility through[1]

$$\boldsymbol{P}(\omega_\mathrm{th}) \propto \overleftrightarrow{\chi}^{(3)}(\omega_\mathrm{th} = \omega_\mathrm{f} + \omega_\mathrm{f} + \omega_\mathrm{f}) : \boldsymbol{E}(\omega_\mathrm{f})\boldsymbol{E}(\omega_\mathrm{f})\boldsymbol{E}(\omega_\mathrm{f})\,, \tag{10.3}$$

where the explicit spatial and temporal dependence of the electromagnetic fields has been dropped. It follows that the THG signal intensity is proportional to the cube of

[1] The notation used here follows [2].

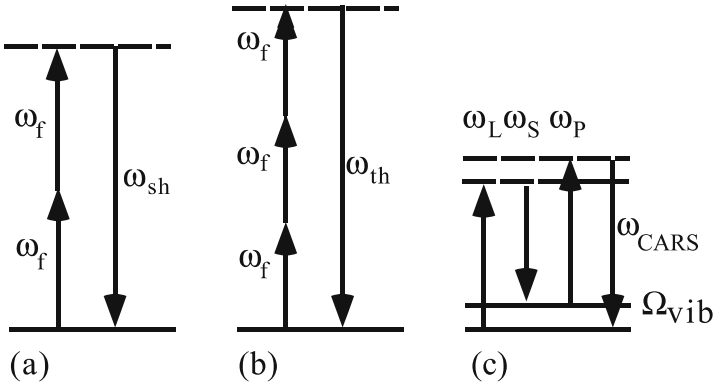


Fig. 10.1. Energy level diagrams for the parametric nonlinear optical processes of SHG (**a**), THG (**b**) and CARS (**c**). Common to all is that they are electronically nonresonant and – by definition – have no net energy transfer to the specimen. The various laser and signal frequencies represent: fundamental (ω_f), second-harmonic (ω_sh), third harmonic (ω_th), *Laser* (ω_L), *Stokes* (ω_S), and *Probe* (ω_P). $\Delta\Omega$ denotes the difference in energy between two vibrational energy levels

input power:

$$\begin{aligned} I_{\mathrm{th}} &\propto |\boldsymbol{P}(\omega_{\mathrm{th}})|^2 \\ &\propto \left|\overleftrightarrow{\chi}^{(3)}(\omega_{\mathrm{th}} = \omega_{\mathrm{f}} + \omega_{\mathrm{f}} + \omega_{\mathrm{f}}) : \boldsymbol{E}(\omega_{\mathrm{f}})\boldsymbol{E}(\omega_{\mathrm{f}})\boldsymbol{E}(\omega_{\mathrm{f}})\right|^2 = \left|\chi^{(3)}\right|^2 \, (I_{\mathrm{f}})^3 \, . \end{aligned} \tag{10.4}$$

In the second part of 10.4 it is assumed for simplicity that all fields have parallel polarisation, i.e. $\chi^{(3)} \equiv \chi^{(3)}_{1111}$. In the following the absence of explicit vector or tensor notation indicates the implicit assumption of a parallel polarisation condition for all fields involved. By definition parametric processes are coherent techniques which are electronically nonresonant with an, for all practical purposes, instantaneous interaction with the specimen. This results in the absence of a net energy transfer to the specimen, rendering these techniques potentially nondestructive.

10.2.2 Optical Sectioning Capability

Common to all nonlinear optical techniques is their capability of optical sectioning in microscopy applications. Due to the nonlinear dependence of the signal on the laser intensity – $I_{\mathrm{signal}} \propto (I_{\mathrm{input}})^n$ with $n > 1$ – the recorded signal is dominated by the in-focus contributions with negligible out-of-focus contributions.

This is in contrast to "conventional" linear microscopic techniques, where the signal response depends linearly on the input intensity – as is the case for e.g. single-photon absorption fluorescence. In this case a uniform sample layer contributes equally to the signal, whether it is in- or out-of-focus. Thus the axial position of such a layer cannot be determined with widefield fluorescence microscopy. The use of a confocal pinhole suppresses the contributions from out-of-focus planes permitting "optical sectioning" of the sample and thus a true axial resolution.

Nonlinear optics microscopy applications, on the other hand, have inherent optical sectioning capability. Even without the use of a confocal pinhole, these microscopic techniques provide three-dimensional imaging capability. This can provide significant advantages especially for imaging in turbid media.

10.2.3 Second Harmonic Generation (SHG)

Second harmonic generation is widely used to generate new frequencies, and has been used as a tool for imaging nonlinear susceptibilities in various materials (see for instance: [8–14]). SHG involves the interaction of light with the local nonlinear properties that depend on the molecular structure and hyperpolarizabilities. Therefore, the second harmonic intensity, and hence the contrast, is a function of the molecular properties of the specimen and its orientation with respect to both the direction and the propagation of the laser beam.

The induced macroscopic polarisation in second harmonic generation (SHG) is given by (see also the energy level diagram of Fig. 10.1a):

$$\boldsymbol{P}(\omega_{\mathrm{sh}}) \propto \overleftrightarrow{\chi}^{(2)}(\omega_{\mathrm{sh}} = 2\omega_{\mathrm{f}}) : \boldsymbol{E}(\omega_{\mathrm{f}})\boldsymbol{E}(\omega_{\mathrm{f}}) \, , \tag{10.5}$$

where ω_f and ω_{sh} denote the frequency of the fundamental and second harmonic field respectively. It follows that the intensity of the second harmonic signal has a square dependence on the input laser intensity:

$$I_{sh} \propto \left|\chi^{(2)}\right|^2 (I_f)^2 \ . \tag{10.6}$$

Since it is related to the second-order nonlinear susceptibility, SHG vanishes for a (locally) symmetric distribution of dipoles. The signal response is either confined to the surface of isotropic specimens or can probe local molecular asymmetries.

As characteristic for all coherent nonlinear phenomena, the SHG signal is strongly directional. The so-called phase anomaly of tightly focused laser beams – with a phase retardation of the focused wave relative to an unfocused plane wave – results in specific, in some cases non-isotropic, signal emission patterns [12]. Resonant enhancement of the SHG signal occurs when the two-photon transition approaches an electronic state of the specimen. This effect can be used effectively through the addition of a specific "SHG label". Also, SHG contrast can be used to discriminate between various molecular conformations. For instance, it has been shown that chirality of the molecule can provide a two-fold enhancement of the signal in certain molecules, relative to an achiral form [11].

10.2.4 Third Harmonic Generation (THG)

In contrast to SHG, third harmonic generation (THG) is allowed in isotropic media. Nevertheless, interference – as a result of the Gouy phase shift – of radiation generated before and after focus results in efficient THG only in case of interfaces in either the third-order nonlinear susceptibility or the dispersion. Recently it was realised that this property makes THG a powerful three-dimensional imaging technique [15,16].

For THG the induced macroscopic polarisation is given by (see also the energy level diagram of Fig. 10.1b):

$$\boldsymbol{P}(\omega_{th}) \propto \overleftrightarrow{\chi}^{(3)}(\omega_{th} = 3\omega_f) : \boldsymbol{E}(\omega_f)\boldsymbol{E}(\omega_f)\boldsymbol{E}(\omega_f)\,, \tag{10.7}$$

where ω_f and ω_{th} denote the frequency of the fundamental and second harmonic field respectively. It follows that the intensity of the third harmonic signal has a cube dependence on the input laser intensity:

$$I_{th} \propto \left|\chi^{(3)}\right|^2 (I_f)^3 \ . \tag{10.8}$$

The generation of the third harmonic under tight focusing conditions has been described in detail (see e.g.: [4,15,17]). THG is generally allowed in any material, since odd-powered nonlinear susceptibilities are nonvanishing in all materials. However, due to the Gouy phase shift of π radians that any beam experiences when passing through focus, THG is absent for $\Delta\boldsymbol{k} = 3\boldsymbol{k}_f - \boldsymbol{k}_{th} \leq 0$, i.e. for exact phase matching or in case of a negative phase mismatch. The latter is the case for media with normal dispersion, i.e. where the refractive index decreases as a function

of wavelength. Qualitatively, in a homogeneous medium the THG waves generated before and after the focal point destructively interfere, which results in the absence of a net THG production. However, in case of inhomogeneities near the focal point, efficient generation of the third harmonic is possible. This is especially the case for interfaces in the dispersion and/or third-order nonlinear susceptibility ($\overleftrightarrow{\chi}^{(3)}$). Note that THG is thus not restricted to the surface of the material only, but rather results from the bulk of the material contained within the focal volume and the presence of interfaces or inhomogeneities therein. This is confirmed also by the absence of a back-propagating THG signal [15].

These specific properties of THG are described well with Gaussian paraxial theory, even in the case of high NA focussing conditions [18]. With these approximations, an analytic solution can be derived which takes the form:

$$E_{\mathrm{th}}(\boldsymbol{r}) = \eta A_{\mathrm{th}}(z) \exp\left(\frac{-3\eta k_{\mathrm{f}} r^2}{2z_{\mathrm{R}}}\right), \tag{10.9a}$$

where

$$\begin{aligned} \eta(z) &= \frac{1}{1 + \mathrm{i}z/z_{\mathrm{R}}}, \\ A_{\mathrm{th}}(z) &= 2\pi \mathrm{i}\omega_{\mathrm{th}} \chi^{(3)} A_{\mathrm{f}}^3 \frac{S(z)}{n_{\mathrm{th}} c}, \\ S(z) &= \int^{z} \eta^2(\xi) \exp(\mathrm{i}\Delta k \xi) \mathrm{d}\xi . \end{aligned} \tag{10.9b}$$

In (10.9a), $E_{\mathrm{th}}(\boldsymbol{r})$ is the third harmonic field envelope, A denotes the peak amplitude and z_{R} is the Rayleigh length, which is equal to half the confocal parameter ($b = 2z_{\mathrm{R}}$).

10.2.5 Coherent Anti-Stokes Raman Scattering (CARS)

CARS is the nonlinear optical analogue of spontaneous Raman scattering. While amplifying the Raman signal by many orders of magnitude and enhancing the detectability of the signal against background luminescence, it retains the unique spectral specificity of Raman scattering. The Raman spectrum can provide a molecular 'fingerprint' even at room temperature and in complex environments such as living cells. In addition it is sensitive to intermolecular interactions. Recent developments in laser technology have permitted introduction of this well known non-linear spectroscopic technique into the field of high resolution microscopy [19,20].

For CARS the induced macroscopic polarisation is given by (see also the energy level diagram of Fig. 10.1c):

$$\boldsymbol{P}(\omega_{\mathrm{AS}}) \propto \overleftrightarrow{\chi}^{(3)}(\omega_{\mathrm{AS}} = \omega_{\mathrm{L}} - \omega_{\mathrm{S}} + \omega_{\mathrm{P}}) : \boldsymbol{E}(\omega_{\mathrm{L}})\boldsymbol{E}(\omega_{\mathrm{S}})\boldsymbol{E}(\omega_{\mathrm{P}}), \tag{10.10}$$

where ω_L, ω_S, ω_P and ω_{AS} denote the frequencies of the so-called *Laser*, *Stokes*, *Probe* and *Anti-Stokes* field respectively. CARS is a four-wave mixing technique in which two laser beams – *Laser* and *Stokes* – set up a grating, off which the third beam – *Probe* – undergoes a Bragg diffraction, generating an *anti-Stokes* signal beam. In a plane wave approximation and for a non-absorbing medium, the CARS signal strength is given by [25]:

$$I_{CARS} \propto \left|\chi^{(3)}\right|^2 I_L I_S I_P \,\mathrm{sinc}^2\left(\Delta \boldsymbol{k} \cdot d/2\right) . \tag{10.11}$$

where I denotes the intensity of the lasers and CARS signal, d is the thickness of the scattering volume and $\Delta \boldsymbol{k} = \boldsymbol{k}_{AS} - \boldsymbol{k}_L + \boldsymbol{k}_S - \boldsymbol{k}_P$ describes the phase mismatch. It has been shown [19,26] that for high numerical aperture focusing conditions the phase matching condition is less stringent. A collinear beam geometry ($\boldsymbol{k}_{AS} = \boldsymbol{k}_L = \boldsymbol{k}_S = \boldsymbol{k}_P$) is therefore commonly used to achieve maximum spatial resolution in microscopy applications. Also for practical reasons *Laser* and *Probe* are generally derived from the same laser source ($\omega_L = \omega_P$ and $I_L = I_P$).

For a sample consisting of one or more molecular species, the CARS signal is proportional to:

$$I_{CARS} \propto \left|\sum_k N_k \chi_k^{(3)}\right|^2 , \tag{10.12}$$

where N_k denotes the number of molecules of type k. The non-linear susceptibiltiy ($\chi^{(3)}$) for each molecular species consists of a resonant (R) and a non-resonant (NR) contribution:

$$\chi^{(3)} = \chi_R^{(3)} + \chi_{NR}^{(3)} \tag{10.13}$$

Far away from one-photon resonances, the two-photon resonant Raman contribution to the non-linear susceptibility can be written as:

$$\chi_R^{(3)} \propto \sum_j \frac{A_j}{\delta_j - i\Gamma_j} \tag{10.14}$$

where A_i is a real constant containing the vibrational scattering cross-section of the vibrational mode i, $\delta_i = \Omega_i - \omega_L + \omega_S$ denotes the detuning from the vibrational resonance Ω_i, Γ_i is the HWHM of the spontaneous Raman scattering vibrational transition and the summation runs over all vibrational resonances.

10.3 Third Harmonic Generation (THG) Microscopy

10.3.1 General Characteristics

The main properties of THG microscopy that follow from theory – most of which have now been verified in practice – are as follows:

- The contrast in THG microscopy is based on either a change in third-order susceptibility or in the dispersion properties of the material, within the focal volume of the fundamental radiation. In other words, an interface between two materials is needed for efficient THG. Denoting the two materials with 1 and 2 respectively, it is required that either $\chi_1^{(3)} \neq \chi_2^{(3)}$ or $\Delta n_1 \neq \Delta n_2$, where $\Delta n_i = n_i(\lambda_f) - n_i(\lambda_{th})$ describes the dispersion of material *i*. It follows that, in contrast to phase microscopy, THG imaging can discern the interface between two media on the basis of a difference in nonlinear susceptibility alone. An example of this is the fact that the boundary between immersion oil and a microscope coverglass – which have been matched in refractive index explicitly – is clearly imaged in THG microscopy. THG imaging is a transmission mode microscopy, similar to phase-contrast or DIC microscopy, but with inherent three-dimensional sectioning properties. Thus, whereas phase-contrast microscopy depends on accumulated phase differences along the optical path length, THG microscopy is sensitive to differences in specimen properties localised within the focal volume.
- The generation of third harmonic is restricted to the focal region. In particular, the full-width-at-half-maximum (FWHM) of the axial response of a THG microscope to an interface between two media with a difference in nonlinear susceptibility alone is equal to the confocal parameter at the fundamental wavelength [15].
- THG is a coherent phenomenon in which the third harmonic radiation is generated in the forward direction. For a linearly polarised input laser beam, the generated third harmonic is also linearly polarised in the same direction [22]. The third-order power dependence of THG on the input laser power, results in an approximately inverse square dependence on the laser pulse width. Typical conversion efficiencies from fundamental to third harmonic are in the range of 10^{-7}–10^{-9} [22], and conversion efficiencies upto 10^{-5} have been reported for specific materials [23]. The efficiency of the THG process depends critically on the orientation of the interface relative to the optical axis [16].
- The noninvasive character of THG imaging has been demonstrated in various applications of microscopic imaging of biological specimens *in-vivo* [22,24,25]. In addition, fading of contrast – equivalent to the bleaching of fluorescence – is absent in THG imaging applications.

The experimental setup for THG microscopy is shown schematically in Fig. 10.2. A near IR femtosecond laser beam – in our case the idler wave from an optical parametric amplifier – is focused by a high NA microscope objective onto the sample. Two scanning mirrors provide for *XY* beam scanning. The sample itself can be moved in the axial direction. The signal is emitted in the forward direction and collected by either a second microscope objective or a condenser lens. Note that the NA of the generated third harmonic signal is one-third of the input NA [18]. The signal passes a filter that blocks the residual laser light and is detected on a photomultiplier tube (PMT). It has also been shown [22] that the THG signal can readily be recorded directly on a video camera.

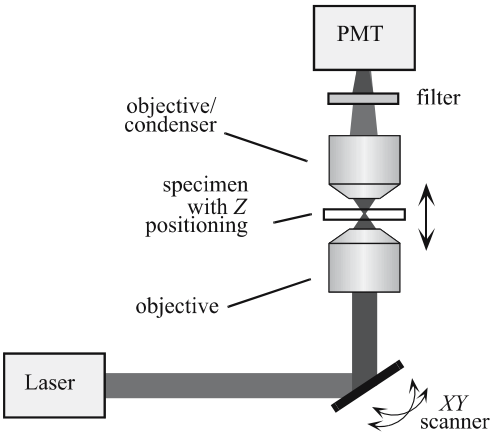


Fig. 10.2. Schematic of the experimental setup for THG microscopy. Laser light at the fundamental frequency ω_f is focussed onto the specimen by a high NA microscope objective. The third harmonic signal (at ω_{th}) is collected in the forward direction by a second microscope objective or condenser and detected on a photomultiplier tube (PMT). The specimen is raster scanned in the lateral plane by scanning of the laser beam and in the axial by movement of the specimen

10.3.2 Selected Applications

In a first demonstration of the noninvasive character of THG microscopy, the technique was applied live cells [22]. In particular, the Rhizoids from the alga *Chara* were imaged Fig. 10.3, where the strong cytoplasmic streaming was used as an internal check for the survival of these tubular single cells. These cells have been studied widely, especially with respect to their response to gravity, which is thought to be related to the so-called statoliths – which are vesicles containing $BaSO_4$ crystals, contained in the tip of the root. In *in-vivo* imaging, the statoliths show dynamic motion while remaining anchored to the actin filament network. No disruption of the cytoplasmic streaming, nor any fading of contrast, has been observed for more than an hour of continuous exposure; a first indication that the cell remains functional.

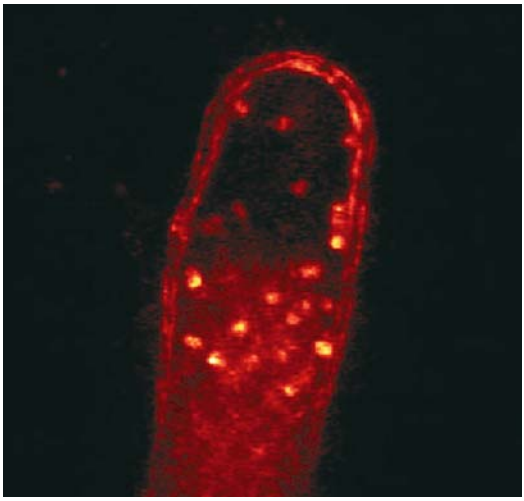


Fig. 10.3. THG image of the Rhizoid tip from the alga *Chara*. The so-called statoliths – which are vesicles containing $BaSO_4$ crystals – are clearly visible as well as the cellular membrane

Figure 10.4 shows the three-dimensional imaging capabilities of THG microscopy. A total of 36 optical sections – taken at 1 μm axial intervals – were taken of *Spyrogira*. Figure 10.4a–c shows different optical sections of top, middle and bottom part of the axial stack. Fig. 10.4d is a three-dimensional reconstruction. The diameter of the spiral is approximately 35 μm and the spacing between the individual spiral ribs ~10 μm. The shadow below the three-dimensional reconstruction is due to THG from the water-coverglass interface. We have checked – by measurement of the output wavelength and polarisation analysis –that the signal is truly from THG and not from some sort of autofluorescence from the sample. Also, we observed no fading of contrast over prolonged exposure times. Alternatively, THG microscopy can be used in the material sciences as demonstrated by the use of the technique to visualise the results of laser-induced breakdown in glass Fig. 10.5 [26]. In this data storage type application, high energy IR pulses (~0.7 μJ/pulse) were used to "write" specific patterns in different focal planes in glass. In this particular case, the letters "U", "C", "S" and "D" were written in planes with an axial separation of 19 μm. The "write" process is due to laser-induced breakdown, which is a multiphoton absorption process that causes highly localised plasma formation. The rapid expansion of the plasma causes a microexplosion which results in observable damage. The created "damage structure" is considered to be due to either a small change in refractive index or a vacuum bubble [27]. The "written" pattern – consisting of these microstructures – can subsequently be "read" with THG microscopy using low energy IR pulses (~ 40 nJ/pulse). This energy level is below the threshold of laser-induced breakdown. THG microscopy can also be used for material characterisation.

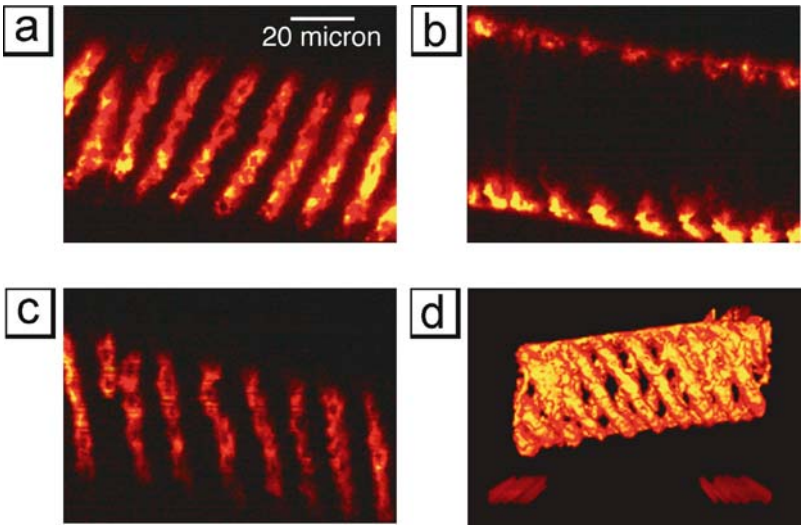


Fig. 10.4. Selection of THG images of *Spyrogira* out of a total of 36 optical sections at 1 μm axial intervals. (**a**)–(**c**) Images at the top, middle and bottom of the stack. (**d**) 3D reconstruction of the complete stack

In principle, a quantitative measurement of the dispersion, $\Delta n_i = n_i(\lambda_f) - n_i(\lambda_{th})$, and the nonlinear susceptibility, $\chi^{(3)}$, is possible [18]. An indication of this potential is shown in Fig. 10.6. In these experiments the sample consisted of a thin layer of either immersion oil or air between two microscope coverglasses. The total THG intensity was recorded as a function of the axial position for various excitation numerical apertures. To characterise the functional shape of these THG axial profiles three intensity ratios were defined: $R_{pre} = I_{THG}(-L/4)/I_{THG}(0)$, $R_{post} = I_{THG}(+L/4)/I_{THG}(0)$ and $R_{mid} = I_{THG}(-L/2)/I_{THG}(0)$, where L is the separation between the two interfaces. The THG yield is represented by I_{THG}, with the z position of the interface as argument: the ratios represent the THG intensities resulting when the focus lies exactly halfway the interfaces $I_{THG}(+L/2)$, one quarter gap thickness before the first interface $I_{THG}(-L/4)$, and likewise behind the first interface $I_{THG}(+L/4)$; all normalised by the intensity at the first interface $I_{THG}(0)$. Theoretically these ratios are calculated with Gaussian paraxial theory. With "hat"-profile beams, the agreement between experiment and theory is only qualitative. Nevertheless, the calculations reproduce the basic features of the experimental data. For the calculations we used $\Delta n_{oil} = -0.028$ and $\Delta n_{K5} = -0.032$, with $n_{oil} = 1.50276$ and $n_{K5} = 1.51146$ at $\lambda_f = 1100\,\text{nm}$). The ratio of the nonlinear susceptibilities of K5 and immersion oil was chosen arbitrarily to be $\chi^{(3)}_{oil}/\chi^{(3)}_{K5} = 3$. The numerical calculations show that the functional form of the signal is dependent almost solely on the change in dispersion, whereas the magnitude of the THG signal depends primarily on the ratio of the nonlinear susceptibilities.

In recent studies [33,34] it has been shown that quantitative measurement of the third-order non-linear susceptibility using THG is also possible as follows.

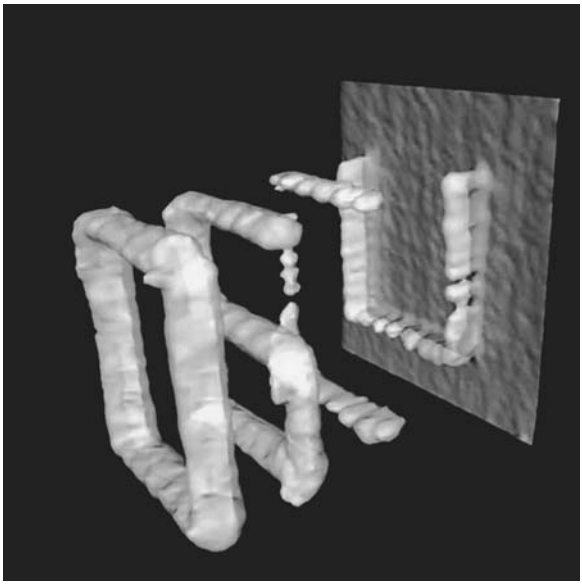


Fig. 10.5. 3D reconstruction from a series of axially sectioned THG images taken at 2 μm axial intervals. The letters are approximately 20 μm wide and were written in focal planes spaced 19 μm apart

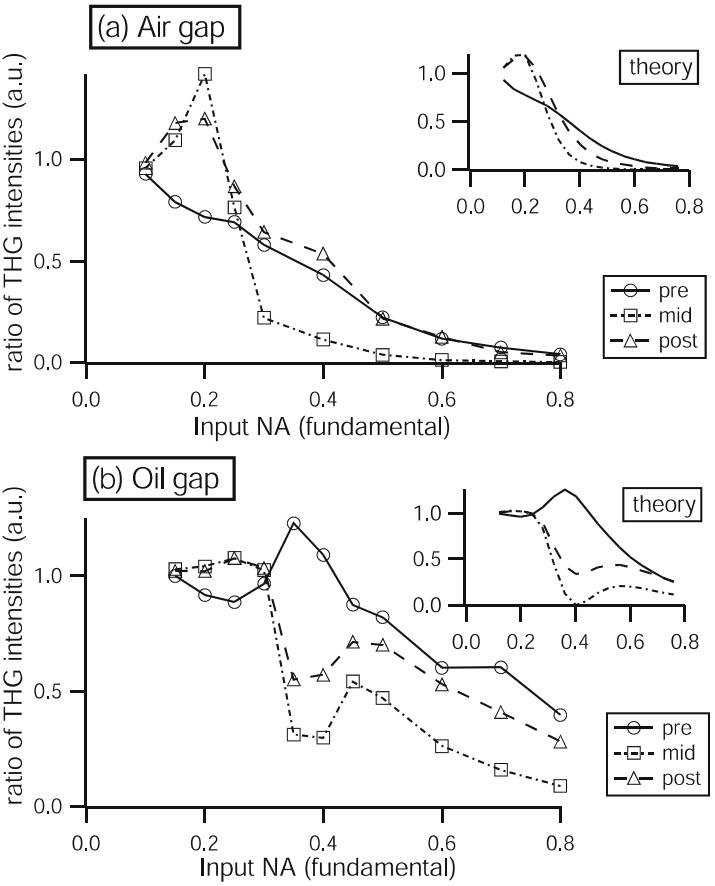


Fig. 10.6. Measured intensity ratios, R_{pre}, R_{mid}, R_{past}, for a double-interface geometry, the gap being filled with immersion air (**a**) and oil (**b**) respectively. The insets show the results from calculations based on Gaussian paraxial theory. For the calculations we used $\Delta n_{\mathrm{oil}} = -0.028$ and $\Delta n_{\mathrm{K5}} = -0.032$, with $n_{\mathrm{oil}} = 1.50276$ and $n_{\mathrm{K5}} = 1.51146$ at $\lambda_{\mathrm{f}} = 1100\,\mathrm{nm}$ and $\chi^{(3)}_{\mathrm{oil}}/\chi^{(3)}_{\mathrm{K5}} = 3$

A material for which the susceptibility is to be measured, is placed between two cover glasses. This sample is then axially scanned through a focus of moderate NA (~0.4) and the THG signal from both the air-glass and the air-material interfaces is recorded. The ratio of the peak THG intensities for the two surfaces can then be related to the non-linear susceptibility of the unknown material if the susceptibility and the dispersion for the used cover glasses is known.

10.3.3 Summary

The examples of THG microscopy applications given above show the potential of this technique. Its unique ability to observe – with three-dimensional resolution –

changes in the material properties of transparent specimen can be used effectively in both biological and material sciences applications. The lateral resolution of the technique is in the range of 400 nm, whereas the axial resolution scales with the confocal parameter of the fundamental beam, and is generally in the order of 1 μm. The fact that THG is an electronically nonresonant interaction with the specimen, renders it noninvasive, as has been shown in *in-vivo* cell biological studies. In addition this ensures that there is no fading of contrast, which is a general problem in fluorescence based techniques.

The contrast in THG microscopy is based on changes in dispersion and changes in the magnitude of the nonlinear susceptibility. The influences of both of these parameters can be largely decoupled. When measuring for instance THG axial profiles as a function of the numerical aperture, the functional shape is dependent primarily on changes in dispersion, whereas the total THG yield depends almost completely on the changes in nonlinear susceptibility alone. This provides the opportunity to access these parameters quantitatively. A prerequisite for these experiments is that the geometrical shape of the specimen is known in some detail. Being a coherent phenomenon – the magnitude of the THG signal is dependent on the specific structure of the specimen. Thus in principle, various structures may yield the same image. Note however, that due to the nonlinearity of the THG process, this sensitivity to the structure of the specimen extends only over a range of the order of the confocal parameter of the fundamental excitation beam.

10.4 Coherent Anti-Stokes Raman Scattering (CARS) Microscopy

10.4.1 General Characteristics

Raman spectroscopy is sensitive to molecular vibrations, which in turn reflect molecular structure, composition and inter-molecular interactions. As such, Raman spectroscopy is unique in that it provides detailed intra- and inter-molecular structural information and specificity at room temperature, even within extremely complex systems such as living cells. Raman spectroscopy suffers, however, particularly from a low scattering cross section as well as from interference from naturally occurring luminescence, limiting its applicability to high resolution microscopy. Stimulated Raman scattering – and in particular CARS – can overcome these drawbacks of spontaneous Raman scattering rendering the signal strength required for effective image acquisition in high resolution microscopy.

CARS microscopy was first introduced in the early eighties [28,29], but found limited application at that time, probably due to shortcomings in laser technology. The technique was recently reinvented [19,20], using novel approaches both in terms of the laser apparatus used and the application of non-collinear phase matching geometries at high numerical-aperture focusing conditions. Several approaches have now been developed for the application of CARS microscopy in high resolution microscopy. For instance, Xie and co-workers [30,31] showed that an epi-detection

mode in CARS enhances the sensitivity of the technique to small features, by effectively removing the nonresonant background contribution. In a later publication this group showed that the nonresonant background can alternatively be suppressed by using polarised CARS [32]. Recently, we have shown [33] the potential of multiplex CARS microscopy for imaging the chemical and physical state of biological model system. In multiplex CARS a significant region of the vibrational spectrum is addressed simultaneously – rather than point-by-point as in "conventional" CARS. The general properties of CARS microscopy can be summarised as follows (see [36,37] for recent reviews on CARS microscopy):

- The CARS spectrum generally consists of both a resonant and a nonresonant contribution. The resonant part is proportional to the square of the Raman scattering cross section.
- CARS is a coherent process in which the signal is generated in a forward direction determined by the phasematching geometry ($\boldsymbol{k}_{\mathrm{AS}} = \boldsymbol{k}_{\mathrm{L}} - \boldsymbol{k}_{\mathrm{S}} + \boldsymbol{k}_{\mathrm{P}}$). For high numerical aperture focusing conditions the phasematching condition is significantly relaxed. The divergence of the generated emission scales directly with the lateral waist of the interaction volume.
- Because of the signal's cube dependence on the laser input intensity, CARS microscopy provides inherent optical sectioning, enabling three-dimensional microscopy. The resolution in CARS microscopy is determined by the interaction volume. While sacrificing some of the attainable spatial resolution, nonlinear phase matching configurations [20] can be implemented in high NA microscopy. This mode of operation is essential for the use of time-resolved CARS and may enhance the detection sensitivity in some cases.
- CARS microscopy is a form of "chemical imaging", which – through the Raman spectrum – is particularly sensitive to both the molecular vibrational signature and to intermolecular interactions.

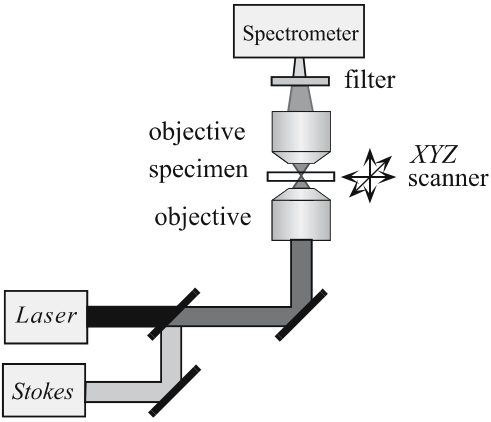


Fig. 10.7. Schematic of the experiment setup for CARS microscopy. The output of two lasers – denoted by *Laser* and *Stokes* – is made collinear and synchronised in time. The beams are focused by a high NA microscopy objective onto the sample, which is moved by piezo scanners in all three dimensions. The generated CARS signal is collected in the forward direction by a second microscope objective and measured on a spectrometer. A CARS spectrum is acquired at each specimen position

10.4.2 Multiplex CARS

A typical CARS experimental setup is shown in Fig. 10.7. Two lasers (pico- and/or femtosecond) are synchronised in time and made collinear before being focused by a high NA microscope objective onto the sample. The CARS image is acquired point-by-point where the sample is scanned in all three spatial directions. The CARS signal is emitted in the forward direction and the, so-called, *Laser* and *Stokes* are blocked by a set of appropriate filters. In the multiplex CARS configuration a narrow bandwidth ($\sim 2\,\mathrm{cm}^{-1}$) picosecond Ti:Sapphire *Laser* laser is used, in combination with a broad bandwidth ($\sim 150\,\mathrm{cm}^{-1}$) femtosecond tunable Ti:Sapphire *Stokes* laser. For every image point, the CARS spectrum is measured with a spectrometer, with typical acquisition times in the order of 20-100 ms. As an example, Fig. 10.8 shows the CARS spectrum of a di-stearoylphosphatidylcholine (DSPC) multilamellar vesicle. The solid line represents a fit of the theoretically expected signal to the data. For the theoretical expression use is made of parameters (linewidths and line positions) obtained from the spontaneous Raman scattering signal (shown as an inset). This effectively reduces the number of free fitting parameters for the CARS signal to the magnitude of both the resonant and nonresonant contribution.

Figure 10.9 shows the various components that contribute to the measured multiplex CARS signal. In Fig. 10.9a both the original data are shown and the signal after dividing out the influence of the *Stokes* spectral intensity profile. The CARS signal is proportional to the absolute square of the nonlinear susceptibility – (10.11) – which in turn consists of a sum of a resonant and a nonresonant contribution [33]:

$$I_{\mathrm{AS}} \propto \left|\chi^{(3)}_{\mathrm{NR}} + \chi^{(3)}_{\mathrm{R}}\right|^2 . \tag{10.15}$$

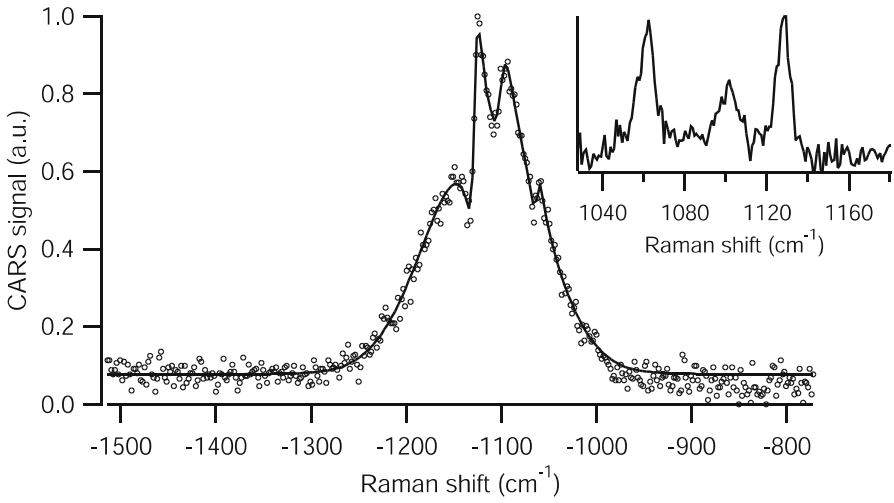


Fig. 10.8. Multiplex CARS spectrum of a DSPC multi-lamellar vesicle. The experimental data (*open circle*) are fitted (*solid line*) to a theoretical expression for the CARS signal which contains three vibrational resonances as deducted from the spontaneous Raman signal (*inset*)

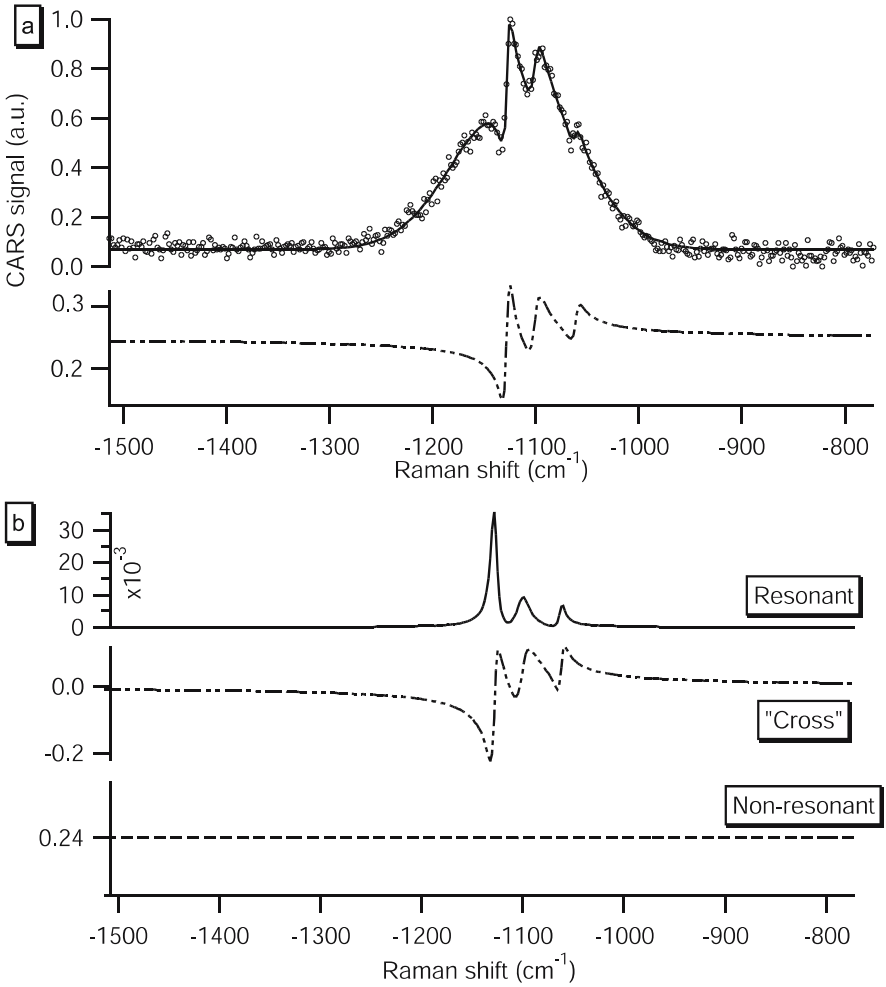


Fig. 10.9. (**a**) Top: Multiplex CARS spectrum of a DSOPC multi-lamellar vesicle. Bottom: CARS signal after dividing out the *Stokes* spectral intensity profile. (**b**) The three contributions that make up the signal of part a, bottom). Top: vibrationally resonant. Middle: "cross"-term. Bottom: vibrationally nonresonant

The total signal thus consists of three parts: a constant nonresonant background, a purely resonant contribution with a Lorentzian shape and amplitude proportional to the square of the Raman scattering cross sections and a cross term which yields a dispersive signal. The different contributions to the total CARS signal of Fig. 10.9a are shown in Fig. 10.9b as they are derived from the fit. Finally, Fig. 10.10 shows the image – one optical section – acquired from the DSPC multi-lamellar vesicle. The 50×50 pixel image was acquired in 250 seconds. This should be compared to a typical acquisition time – for the same excitation power and signal-to-noise ratio –

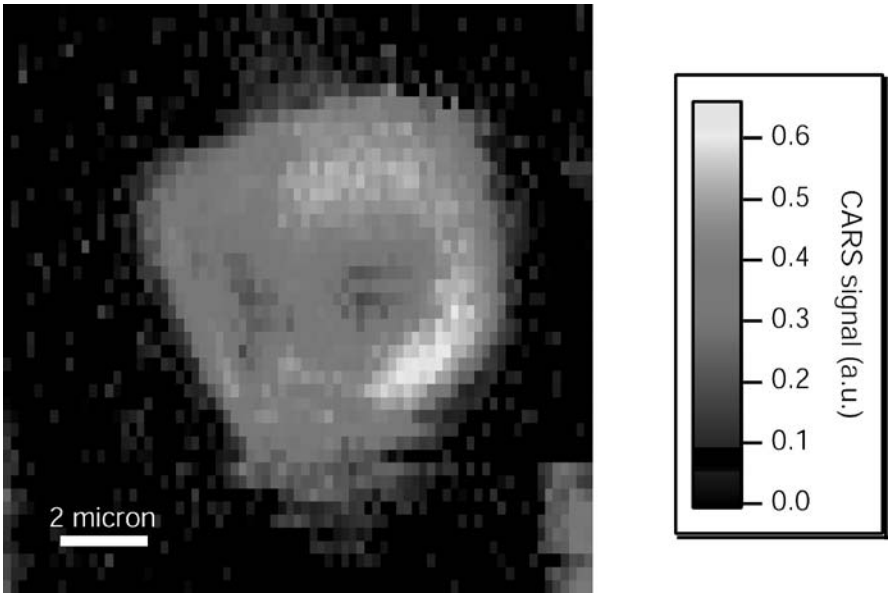


Fig. 10.10. Multiplex CARS images of a DSPC multi-lamellar vesicle. The acquisition time for the full CARS spectrum is 100 ms. Plotted is the amplitude of the $-1101\,\mathrm{cm}^{-1}$ mode as deduced from fitting the data to a theoretical expression for the CARS spectrum

of a single Raman spectrum of such vesicles in the order of 15 minutes. Note that the fitting procedure ensures that laser power fluctuations and time jitter between the lasers do not influence the signal-to-noise of the acquired image. The figure clearly shows that the density of lipids minimises in the centre of the vesicle.

As shown in these figures, multiplex CARS microscopy provides a CARS spectrum over a significant part of the vibrational spectrum in typically 20-100 ms. The signal-to-noise of these spectra is more than sufficient to deduce important information from the specimen. We have shown [33], for example, that within a single experimental run, the information obtained from the so-called skeletal optical mode region of the vibrational spectrum can be used to discriminate between lipid membranes which are in the liquid crystalline phase (i.e. above the phase transition temperature) or in the gel phase (below the phase transition temperature). Interestingly, this region of the spectrum is quite insensitive to the particular chemical structure of the acyl chains, providing a general probe to the physical structure of lipid membranes. Most importantly, multiplex CARS provides a more than four orders of magnitude increase in signal strength compared to spontaneous Raman spectroscopy, permitting high resolution microscopy with realistic (~minutes) image acquisition times.

Coherent anti-Stokes Raman scattering (CARS) microscopy is rapidly developing into an important tool in biophysics and biology research. It provides three-

dimensional imaging with physical [35] and chemical [38] specificity without the need for introducing artificial labels. Due to the relatively large cross-section and density of C-H stretch vibrations, lipid molecules have been used extensively in CARS microscopy studies both in model lipid bilayers [39,40] and cellular membranes [22,41,42]. Also applications of CARS microscopy in the material sciences have recently been reported [43].

10.4.3 Summary

CARS microscopy is a novel approach to obtain spatially resolved information about the chemical composition and physical structure of samples. While retaining the spectral specificity of spontaneous Raman spectroscopy, it provides a signal enhancement through the nonlinear interaction with the specimen that enables high resolution microscopy. The nonlinearity of the process also provides the technique with inherent optical sectioning capability. The contrast in the images is based on vibrational spectral features – i.e. to specimen inherent properties – and does not require the addition of specific labels. Also, since CARS is a parametric process, no energy transfer to the specimen occurs in the generation process of the signal providing generally mild exposure conditions for (biological) specimen and ensuring that no fading of contrast occurs.

In many application, CARS microscopy is accomplished using two picosecond lasers, in which case only a single point in the CARS spectrum can be addressed at a time. To obtain spectral information, one of the lasers needs to be tuned, making the CARS spectrum obtained in this way particularly sensitive to laser-induced fluctuations in the signal strength (e.g. due to power fluctuations, timing jitter, etc.). In the newly developed multiplex CARS microscopy technique, a significant part of the CARS spectrum is obtained simultaneously, through the use of a combination of a picosecond (small spectral bandwidth) laser and a femtosecond (broad bandwidth) laser. This eliminates the problems associated with laser-induced fluctuations, and provides CARS spectra of which the signal-to-noise is limited only by Poisson noise. This mode of operation enables – for the first time – the use of CARS spectral data to image highly specific features of the sample.

10.5 Conclusion

Advances in ultrashort pulsed laser technology has provided the opportunity to combine nonlinear optics with high NA microscopy. This opens the way to a rich field of molecular spectroscopic techniques for spatially resolved studies both in the material sciences and in biology. Through its nonlinearity of response, this class of optical techniques provides inherent optical sectioning, and thus three-dimensional imaging capability.

Here we have considered a sub-class of these techniques: parametric nonlinear optics. In parametric processes no energy transfer between the object and the laser field(s), and *vice versa*, takes place, providing the potential for noninvasive imaging

modes, as well as the absence of fading of contrast in the images. Second harmonic generation microscopy exploits the fact that second-order nonlinear interactions are not observed in specimen that contain a centre of inversion symmetry. Thus this technique is particularly sensitive to surfaces or highly oriented samples. Third harmonic generation on the other hand is in principle a "bulk" phenomenon. However, under strong focusing conditions – as in high NA microscopy – the Gouy phase shift at focus causes destructive interference between the third harmonic radiation that is produced before focus with that produced after the focal plane. This implies that significant third harmonic generation is only observed near interfaces in dispersion or nonlinear susceptibility. This makes the technique highly attractive for material properties characterization studies.

Third harmonic generation is one nonlinear optical technique of the broad class of so-called four-wave mixing techniques that have been applied widely in spectroscopic studies. Recently other members of this class have been introduced to high resolution microscopy. Most prominent of these new additions is coherent anti-Stokes Raman scattering (CARS). CARS microscopy combines the unique vibrational spectroscopic specificity of spontaneous Raman scattering with a signal enhancement through the nonlinear interactions by several orders of magnitude. This signal enhancement enables – for the first time – to exploit the detailed Raman spectral information for imaging purposes. In the recently developed multiplex CARS microscopy configuration, the CARS signal is acquired simultaneously over a significant part of the vibrational spectrum. A consequence of this is that the accuracy with which the spectral information can be measured is limited only by Poisson noise, and not – as in conventional CARS – by laser induced signal strength fluctuations (power fluctuations, time jitter, etc.).

Another recent addition to nonlinear optical microscopy from the class of four-wave mixing techniques is the application of the nonlinear optical Kerr effect [34]. In this application, this novel technique has been applied to study the mobility of intracellular water. The technique permits the determination of rotational relaxation times, which for small molecules such as water are typically in the order of ~800 fs.

As a final remark, a word of caution with respect to the interpretation of images that result from parametric nonlinear optical techniques. All these techniques provide a coherent imaging mode, that is: the generated signals from the specimen interfere on the detector to give rise to a certain detected intensity. In general this may obscure the relation between object and image, since in principle different objects can give rise to similar images. However the situation in nonlinear optical microscopy is more favourable since the signals are generated only within the focal volume of the laser beams. Thus only sub-resolution structure within the specimen may complicate the image interpretation.

Acknowledgement

This research was financially supported in part by the Stichting voor Fundamenteel Onderzoek der Materie (FOM), The Netherlands, under grant no. 94RG02 and by

the Stichting Technische Wetenschappen (STW), The Netherlands, under grant no. ABI.4929.

References

1. W. Denk, J.H. Strickler, W.W. Webb: Science **248**, 73 (1990)
2. Y.R. Shen: *The principles of nonlinear optics* (John Wiley & Sons, New York 1984)
3. M. Schubert, B. Wilhelmi: *Nonlinear optics and quantum electronics* (John Wiley & Sons, New York 1986)
4. R.W. Boyd: *Nonlinear optics* (Academic Press, Inc., Boston 1992)
5. S. Mukamel: *Principles of nonlinear optical spectroscopy* (Oxford University Press, Oxford 1995)
6. C. Flytzanis: 'Theory of nonlinear optical susceptibilities', In: *Quantum Electronics, vol. I: Nonlinear Optics*, ed. by H. Rabin, C. L. Tang (Academic Press, New York 1975) pp. 9–207
7. Y. Prior: IEEE J. Quantum Elect. **20**, 37 (1984)
8. R. Hellwarth, P. Christensen: Opt. Commun. **12**, 318 (1974)
9. J.N. Gannaway, C.J.R. Sheppard: Opt. Quant. Electron. **10**, 435 (1978)
10. R. Gauderon, P.B. Lukins, C.J.R. Sheppard: Opt. Lett. **23**, 1209 (1998)
11. P.J. Campagnola, M. Wei, A. Lewis, L.M. Loew: Biophys. J. **77**, 3341 (1999)
12. L. Moreaux, O. Sandre, M. Blanchard-Desce, J. Mertz: Opt. Lett. **25**, 320 (2000)
13. E.C.Y. Tan, K.B. Eisenthal: Biophys. J. **79**, 898 (2000)
14. L. Moreaux, O. Sandre, S. Charpak, M. Blanchard-Desce, J. Mertz: Biophys. J. **80**, 1568 (2001)
15. Y. Barad, H. Eisenberg, M. Horowitz, Y. Silberberg: Appl. Phys. Lett. **70**, 922 (1997)
16. M. Müller, J. Squier, K.R. Wilson, G.J. Brakenhoff: J. Microsc. – Oxford **191**, 266 (1998)
17. J.F. Ward, G.H.C. New: Phys. Rev. **185**, 57 (1969)
18. J.M. Schins, T. Schrama, J. Squier, G.J. Brakenhoff, M. Müller: JOSA B **19**, 1627 (2002)
19. A. Zumbusch, G.R. Holtom, X.S. Xie: Phys. Rev. Lett. **82**, 4142 (1999)
20. M. Müller, J. Squier, C.A. de Lange, G.J. Brakenhoff: J. Microsc. – Oxford **197**, 150 (2000)
21. J. Cheng, L.D. Book, X.S. Xie: Opt. Lett. **26**, 1341 (2001)
22. G.R. Holtom, B.D. Thrall, B. Chin, D. Colson: Traffic **2**, 781 (2001)
23. J. Squier, M. Müller: Rev. Sci. Instrumen. **72**, 2855 (2001)
24. A. Volkmer, J. Cheng, X.S. Xie: Phys. Rev. Lett. **87**, 23901 (2001)
25. A.C. Millard, P. Wiseman, D.N. Fittinghoff, J.A. Squier, M. Müller: Appl. Optics **38**, 7393 (1999)
26. G.C. Bjorklund: IEEE J. Quant. Electr. **11**, 287 (1975)
27. C.B. Schaffer, E.N. Glezer, N. Nishimura, E. Mazur: 'Ultrafast laser induced microexplosions: explosive dynamics and sub-micrometer structures', In: *Proceedings of SPIE, San Jose Jan. 24-30, 1998*, ed. by M.K Reed (SPIE, Washington 1998) pp. 36–45
28. M.D. Duncan, J. Reijntjes, T.J. Manuccia: Opt. Lett. **7**, 350 (1982)
29. M.D. Duncan: Opt. Commun. **50**, 307 (1984)
30. J. Cheng, A. Volkmer, L.D. Book, X.S. Xie: J. Phys. Chem. B **105**, 1277 (2001)
31. A. Volkmer, J. Cheng, X.S. Xie: Phys. Rev. Lett. **87**, 023901 (2001)
32. J. Cheng, L.D. Book, X.S. Xie: Opt. Lett. **26**, 1341 (2001)

33. R. Barille, L. Canioni, L. Sarger, G. Rivoire: Phys. Rev. E **66**, 67602 (2002)
34. V. Shcheslavskiy, G. Petrov, V.V. Yakovlev: Appl. Phys. Lett. **82**, 3982 (2003)
35. M. Müller, J.M. Schins: J. Phys. Chem. B **106**, 3715 (2002)
36. J.X. Cheng, X.S. Xie: J. Phys. Chem. B **108**, 827 (2004)
37. A. Volkmer: J. Phys. D: Appl. Phys. **38**, R59 (2005)
38. G.W.H. Wurpel, J.M. Schins, M. Müller: Opt. Lett. **27**, 1093 (2002)
39. E.O. Potma, X.S. Xie: J. Raman Spectrosc. **34**, 642 (2003)
40. G.W.H. Wurpel, H.A. Rinia, M. Müller: J. Microsc. **218**, 37 (2005)
41. J. Cheng, Y.K. Jia, G. Zheng, X.S. Xie: Biophys. J. **83**, 502 (2002)
42. X. Nan, J.X. Cheng, X.S. Xie: J. Lipid Res.: D300022-JLR200 (2003)
43. E.O. Potma, X.S. Xie, L. Muntean, J. Preusser, D. Jones, J. Ye, S.R. Leone, W. Schade: J. Phys. Chem. B **108**, 1296 (2004)

11 Second Harmonic Generation Microscopy Versus Third Harmonic Generation Microscopy in Biological Tissues

Chi-Kuang Sun

11.1 Introduction

Second-harmonic generation (SHG) and third-harmonic generation (THG) processes are both nonlinear processes, related to the interaction of intense light with matters. SHG process describes the generation of light wave that is twice the frequency (with half of the original wavelength) of the original one while THG process describes the generation of light wave that triples the frequency (with one third of the original wavelength) of the original one. The harmonic light wave generation is coupled from the excited nonlinear polarization P^{NL} under intense laser excitation. The interaction of nonlinear polarization P^{NL} and the excitation light is usually related through a nonlinear susceptibility χ, as previously described in Chaps. 9 and 10. SHG and THG can be visualized by considering the interaction in terms of the exchange of photons between various frequencies of the fields. According to this picture, which is previously illustrated in Figs. 10.1 (a), (b), two or three photons of angular frequency ω are destroyed and a photon of angular frequency 2ω (for SHG) or 3ω (for THG) is created in a single quantum-mechanical process. The solid lines in the figure represent the atomic ground states, and the dashed lines represent what are known as virtual levels. These virtual levels are not energy eigenlevels of the atoms, but rather represent the combined energy of one of the energy eigenstates of the atom and one or more photons of the radiation field. Due to its virtual level transition characteristics, harmonic generations are known to leave no energy deposition to the interacted matters, since no real transition involved and the emitted photon energy will be exactly the same as the total absorbed photon energy. This virtual transition characteristic provides the optical "noninvasive" nature desirable for microscopy applications, especially for live biological imaging.

Due to its nonlinearity nature, the generated SHG intensity depends on square of the incident light intensity, while the generated THG intensity depends on cubic of the incident light intensity. Similar to multi-photon induced fluorescence process, this nonlinear dependency allows localized excitation and is ideal for intrinsic optical sectioning in scanning laser microscopy. Usually the third-order nonlinear susceptibility $\chi^{(3)}(3\omega : \omega, \omega, \omega)$ needed for THG is much weaker than the second-order nonlinear susceptibility $\chi^{(2)}(2\omega : \omega, \omega)$ needed for SHG, thus THG is harder to observe. However, not all biological materials have second-order nonlinear susceptibility. For centro-symmetric media, the lowest order nonlinear susceptibility will be $\chi^{(3)}$ instead of $\chi^{(2)}$. This is why people have rarely heard of generation of

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

SHG from a bulk glass. The random distribution of oxides in glass creates optical centro-symmetry in optical wavelength scale, thus inhibiting the SHG generation. Only non-centro-symmetry media is allowed to generate SHG. On the other hand, all materials are allowed to create third-order susceptibility, of which magnitudes vary according to material property and wavelength.

According to photon momentum conservation, the generated harmonic photons will be emitted in the same direction as the incident photons, unless noncollinear phase matching process is involved. This forward direction emission property restricts the light collection geometry in microscopy applications. Nowadays most SHG and THG microscopes utilize transmission detection, which is different from the reflection detection in most laser scanning fluorescence microscopes.

11.2 SHG Microscopy

In the past four decades, SHG has been widely applied to the study of SHG materials and interfacial regions without a center of symmetry, and was later combined with a microscope for SHG scanning microscopy [1,2] in 1970s. Today, SHG scanning microscope has been widely used in material studies for surface monolayer detection [3], ferroelectric domain structures [4], and nonlinear crystal characterization with 3D resolution [5], as previously described in Chap. 6. With low frequency electric field breaking the centro-symmetry, SHG can also be generated due to low frequency electric field even in centro-symmetry media, providing a tool to image electric field. This process is usually described as a third-order nonlinear process with the nonlinear polarization described by [6]

$$P^{\mathrm{NL}}(2\omega) = \frac{3}{4}\varepsilon_o \chi^{(3)}(2\omega : \omega, \omega, 0)E(\omega)E(\omega)E(0) \tag{11.1}$$

where $E(0)$ is the low frequency electric field. This process is usually referred as Electric-Field-Induced-Second-Harmonic-Generation (EFISHG). Utilizing EFISHG in gallium nitride, we have successfully demonstrated electric field distribution imaging using SHG scanning microscopy [7,8]. Compared with traditional electro-optical sensor probe technique, EFISHG microscopy of electric field has all the advantage of scanning confocal microscope, with high optical resolution and great 3D sectioning power. It is also a background-free experiment if the EFISHG material possesses centro-symmetric properties. Recently we have successfully achieved 3D electric field distribution mapping in an IC circuit utilizing liquid crystal as EFISHG probe materials with sub-micron resolution [9,10].

SHG generation in biological specimen was first observed in 1971 by Fine and Hansen from collageneous tissues [11] by using a Q-switched ruby laser at 694 nm. In 1986, Freund, Deutsch, and Sprecher demonstrated SHG microscopy in connective tissues [12] based on a 1064 nm Q-switched Nd:YAG laser. They attributed the SHG generation to the polarity (that is one type of non-centro-symmetry) of rat-tail tendon, which was then correlated to polar collagen fibrils. In 1989, J.Y. Huang and his coworkers used SHG to probe the nonlinear optical properties of purple

membrane–poly(vinyl alcohol) films [13]. The SHG is attributed to the naturally oriented dipole layers. Recently Alfano and coworkers have also reported SHG from animal tissues [14]. Even though they attributed the SHG from only the surface term that is due to the broken symmetry at the boundary, they also found some SHG intensity dependence on the tissue constitutes with asymmetric structures such as collagen and keratin. Combining this effect, they demonstrated SHG tomography for mapping the structure of animal tissues by use of 100-fs laser pulses at 625 nm [15].

Taking advantage of EFISHG effect, SHG was shown to have intrinsic sensitivity to the voltage (low frequency electric field) across a biological membrane [16] and was combined with microscopy for membrane potential imaging [17,18]. Membrane-staining dye was also used to demonstrate membrane imaging with SHG microscopy [17–19]. 1064 nm Nd:YAG lasers or 880 nm Ti:sapphire lasers were used in these studies.

We have also extensively studied SHG microscopy on biological tissues. Our recently study indicated that strong SHG images are corresponding to highly organized nano-structures. Numerous biological structures including stacked membranes and arranged protein structures are highly organized in optical scale and are found to exhibit strong optical activities through SHG interactions, behaving similar to man-made nonlinear nano-photonic crystals. Previous observations of SHG on collagen fibrils [11,12,14], purple membrane [13], and muscle fibrils [14,15] are just a few examples of it. This assumption was also supported by the recent SHG microscopy of tooth [20], where the SHG was attributed to the highly ordered structures in enamel that encapsulates the dentine.

11.3 Bio-Photonic Crystal Effect in Biological SHG Microscopy

Recent studies on man-made nano-periodic structures, e.g. super-lattices, indicate strong enhancement in SHG occurring only in noncentro-symmetric media [21]. Another recent study of SHG in a one-dimensional semiconductor Bragg mirror with alternating layers of ~100 nm thickness also supports the hypothesis of forward emission SHG enhancement [22] due to nonlinear photonic crystal effect [23]. It is interesting to notice that the periodicity of the observed nonlinear photonic crystal is not necessarily to be on the order of the wavelength or a fraction of the wavelength in the materials [24], but could be much smaller than the affected wavelength with a nanometer scale [21,25]. This strong SHG enhancement could be understood as the break down of optical isotropy within the materials. Therefore, it is reasonable to speculate that the highly organized biological nano-structures may also break optical isotropy and behave as SHG-active photonic crystals. Examples of highly organized biological nano-structures include stacked membranes, such as myelin sheath, endoplastic reticulum (ER), grana in the chloroplast, Golgi apparatus, microfibrils and collagen bundles. It is thus highly possible that these biological structures act as nonlinear “photonic” crystals and can be detected in SHG microscopy. Recently we have studied this bio-photonic crystal effect using a mul-

timodal nonlinear microscopy, which reveals optical SHG activities in those naturally occurring biophotonic crystalline structures. Multimodal nonlinear microscopy [26], in conjunction with the use of a Cr:forsterite laser [27], is based on a combination of different imaging modalities including second-, third-harmonic generations, and multi-photon fluorescence. Due to optical isotropy created by randomly organized biological structures, we found that SHG microscopic images can reveal specially organized crystalline nano-structures inside biological samples where optical centro-symmetry is broken, similar to the nonlinear photonic crystal effect. Our conclusion is also supported by a recent paper, demonstrating that forward emission SHG in purple membrane is caused by nonlinear photonic crystal properties due orderly patched protein bacteriorhodopsin with nano-scaled periodicity [25]. Not only purple membrane, previous SHG observations on collagen and muscle fibrils [11–15] are all with nano-crystalline structures, thus all related to the nonlinear photonic crystal effect. Different from laser-induced fluorescence processes, only virtual states are involved in the harmonic generations (including both SHG and THG). The marked advantage of this virtual transition during wavelength conversion is the lack of energy deposition, thus no photo-damage or bleaching from the process is expected, can be considered as a truly "noninvasive" imaging modality. The SHG microscopy, together with THG microscopy compared in the next section, thus allows structural visualization with minimal or no additional preparation of the samples. Combining multi-photon fluorescence imaging modes, SHG microscopy is useful for investigating the dynamics of structure-function relationship at the molecular and sub-cellular levels.

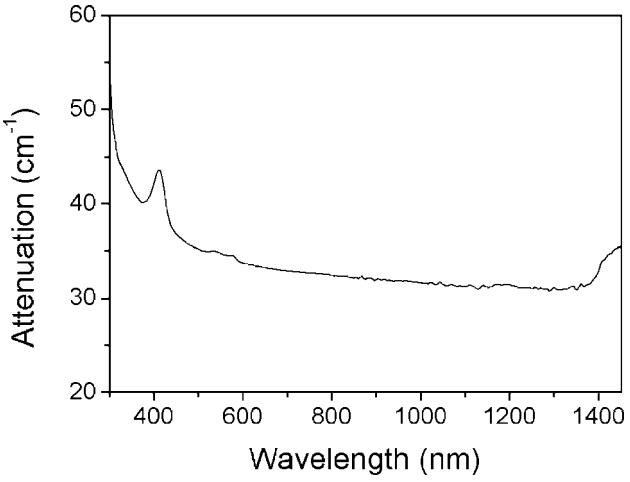


Fig. 11.1. Attenuation spectrum of porcine skin in the visible and near infrared regions. An optical window between 365 nm and 1400 nm is evident

In order to allow both SHG and THG (see next section) within the visible spectrum, but also provide advantages of low attenuation for illumination, we move the excitation wavelength to 1200–1350 nm regimes by using a 100-fs-pulsewidth Cr:forsterite laser. Fig. 11.1 shows that light attenuation (absorption and scattering) in porcine skin reaching a minimum around 1300 nm. Light attenuation in most biological specimen [28,29] reaches minimum in the region of 1300 nm due to the combination of diminishing scattering cross-section with longer wavelength and avoiding resonant molecular absorption of common tissue constituents such as water, protein, and carbohydrates. Due to the high absorption in the visible and near-infrared spectrum, this wavelength is particular useful for imaging plant material [29]. Due to the low attenuation coefficient around 1200–1350 nm, only 10-fold reduction in the signal (for SHG and two-photon fluorescence (2PF)) was observed with 360-μm depths into a maize stem segment fixed in 10% ethanol [26], in good agreement with light attenuation measurement of maize stems [29] that indicates low attenuation coefficient around our illumination with 1230 nm light. This superior depth performance agrees well with previous studies for optical coherent tomography [30] comparing penetration depth between 800 and 1300 nm light sources. In addition, the use of 1300 nm region allows fiber compatibility (due to zero dispersion at 1300 nm) and minimum background detection by Si-based detectors (due to the non-sensitivity of 1200–1350 nm light to Si-based detectors).

Plant cell wall consists of cellulose microfibrils, which are orderly arranged individual to form macrofibrils with a dimension about 10 nm. Within the microfibrils, micelles represent another degree of highly ordered crystalline structure. These crystalline structures produce optical anisotropy and provide active SHG. Figure 11.2 shows the nonlinear emission spectrum that was measured in the transmis-

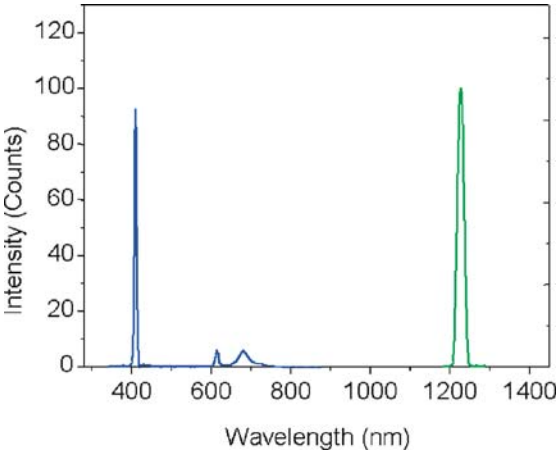


Fig. 11.2. Measured nonlinear emission spectrum from the cell wall of maize stem excited by a femtosecond Cr:forsterite laser with THG centered at 410 nm, SHG centered at 615 nm, and 2PF peaked at 680 nm. The laser spectrum centered at 1230 nm is also provided (not scaled)

sion direction from the cell wall of a parenchyma in maize (*Zea mays*) stem. Symmetric SHG and THG spectra centered at 615 and 410 nm can both be observed, with intensity similar to/or stronger than residual 2PF centered at 680 nm. SHG reflects the highly organized crystalline structures that break the three-dimensional (3D) optical centro-symmetry in the cell wall. The nature of SHG is further confirmed by the strong signal obtained from the stone cell of pear (*Pyrus serotina* R.) fruit (top of Fig. 11.3). The extensive secondary wall development of the sclerenchyma generates significant SHG signals.

Starch granule, a strong birefringent structure, consists of crystalline amylopectin lamellae organized into effectively spherical blocklets and large concentric growth rings. These structural features are presumably responsible for the strong SHG we observed (bottom of Fig. 11.3). For example, the SHG signal from potato

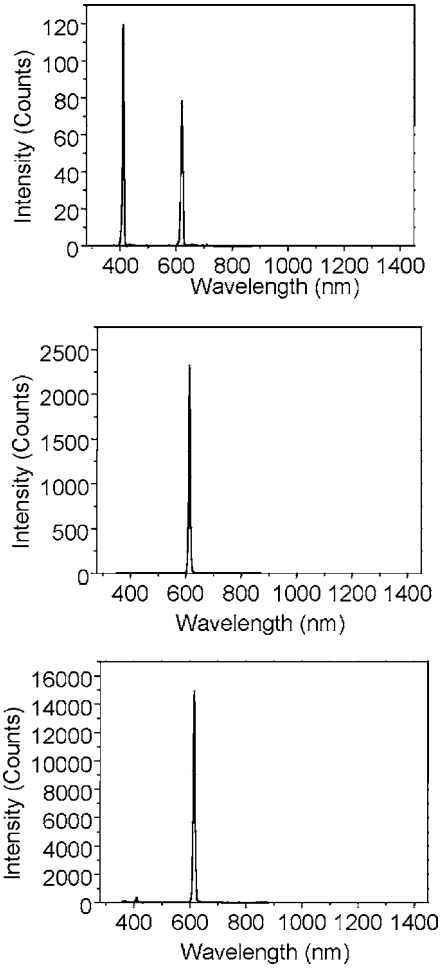


Fig. 11.3. Emission spectra from pear stone cell (*top*), mouse skeletal muscle (*middle*), and potato starch granule (*bottom*) are provided for comparison. All emission spectra (including Fig. 11.2) were taken with similar illumination intensity with normalized 0.1-second integration time

(*Solanum tuberosum* L.) starch granule is so strong that is visible to the naked eyes [26]. Its alternating crystalline and semi-crystalline rings [27] with spatial modulated nonlinear properties could behave as 3D nonlinear photonic "bandgap" crystals [31]. The unexpected strong SHG activity might not only be the result of its super-helical amylopectin nano-structures, but also suggest possible SHG non-collinear phase matching condition provided with its reciprocal lattice basis vectors of the 3D photonic bandgap crystals. Spatial frequency of high order structures between the order of 100 nm up to the order of 10 μm, depending on illumination wavelength and composition materials, could all provide the non-collinear phase matching base-vector for SHG process and can be considered as nonlinear biophotonic "bandgap" crystals [23], providing even stronger SHG activity. This significant enhancement of SHG interactions in a 3D structure could also be treated as a simultaneous availability of a high density of states and improvement of effective coherent length (similar to phase matching effect) due to these wavelength-scale photonic "bandgap" structures [22,24].

Figure 11.4 (a) shows the sectioned (x-y) SHG image taken from the ground tissue of maize stem at a depth of 420 μm from the sample surface. THG (see next section) and 2PF of residue fluorescence images are also shown for comparison. SHG

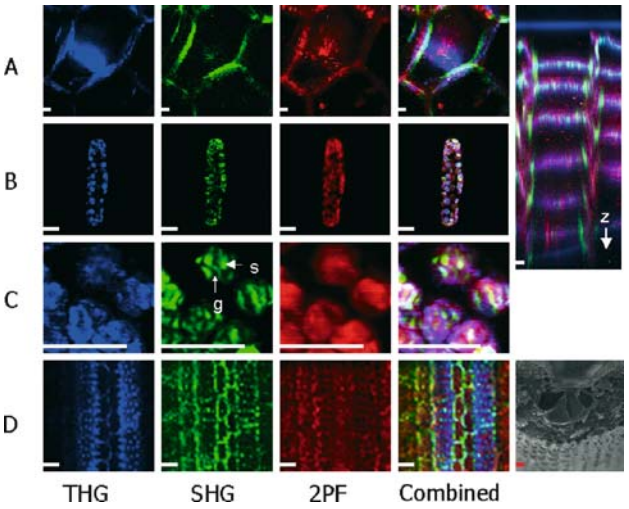


Fig. 11.4. Scanning THG (shown in *blue*), SHG (shown in *green*), 2PF (shown in *red*), and the combined multimodal images. Images were taken from **(a)** (x-y images) ground tissue cells in a maize stem 420 μm from the sample surface **(b)** (x-y images) a live mesophyll cell from *Commelina communis* L. **(c)** (x-y images) enlarged view of (b) **(d)** (x-y images) adaxial surface of rice leaf. The dark region in the middle of THG image corresponds to the bulliform cell layers. Transverse sectional SEM image taken from a rice leaf is provided for correlation purpose (lower right). Multimodal x-z image (with image size of 200 × 540 μm) of ground tissue cells in a maize stem is also shown in the upper right corner. Scale bar: 15 μm. *g*: grana, *s*: starch granule

shows mainly the longitudinal cell walls while 2PF indicates the distribution of the auto-fluorescent molecules. The combined multimodal image is also provided with blue, green, and red colors representing THG, SHG, and 2PF respectively. A combined x-z-λ image (image size: 200×540 μm) of ground tissue cells is also presented in the upper right corner of Fig. 11.4. The intensity contrast of this specific image was processed with log scale so that bottom image corresponding to deeper part of the sample can be easily observed. The correspondence of SHG to longitudinal cell walls can be clearly observed.

A paradermal optical (x-y) section of the adaxial surface of rice (*Oryza sativa* L.) leaf (Fig. 11.4 (d)) also reveals that SHG picks up optically active crystalline structures, including the cuticular papillae and longitudinal cell walls, due to the orderly arrangement of cutin, waxes and cellulose microfibrils respectively. The THG (see next section) and 2PF (corresponding to chlorophyll emission wavelength) images are also shown. For correlation purpose, the lower right corner of Fig. 11.4 gives a cross-sectional SEM (black-and-white) image of a similar rice leaf.

There are numbers of structures in animal tissue that are good candidates for strong SHG. For instance, the orderly arranged sarcomeres in the skeletal muscle have structures fall into the spatial range of strong SHG activity. Fig. 11.5 (a) shows longitudinally sectioned x-y images obtained from the skeletal muscle of mouse. Strong SHG emission was recorded due to orderly packed actin/myosin complexes in sarcomeres (middle of Fig. 11.3). The strong SHG activity from actin/myosin complex can be a useful tool for the study of muscle cell dynamics. Other SHG sources include collagen fiber bundles in connective tissues, as previously reported. Collagen fibrils exhibit a sequence of closely spaced transverse bands those repeat every 68 nm along the length of the fiber, providing necessary condition for biophotonic crystal effect of SHG activity. Figure 11.5 (b) shows cross-sectional x-z image taken from the mouse dermis. The collagen fibrils inside the connective tissue right underneath the epithelial cells can be clearly observed through SHG. Its corresponding x-y section taken at a constant depth of 30 μm was also shown in Fig. 11.5 (c) with wavy structures from collagen fiber bundles. Excellent contrast can be easily obtained through SHG microscopy without staining.

The strength of the SHG signals can vary according to the orientation of the structures which may have different $\chi^{(2)}$ matrix components. This mechanism provides opportunity for structural orientation studies using SHG with controlled illumination polarization. For instance, by varying the incident light polarization (Fig. 11.5 (d)–(f)), the concentric silica deposition in the dumb-bell-shaped silica cells of rice leaf produces orientated SHG images in respecting to the orientation of the illumination polarization (shown as arrows). For comparison, a paradermal SEM (black-and-white) image showing silica cells taken from a similar rice leaf is included in lower right corner of Fig. 11.5.

Apart from mineral deposition in plant cells, laminated membrane structures are also potential candidates for producing strong SHG activity. In chloroplasts, in addition to the 2PF signals generated from the highly auto-fluorescing photosynthetic pigments, SHG appears in different sub-organelle compartments (Fig. 11.4 (c), (d)).

Matching with TEM images of similar specimens, we concluded that the signals of SHG are the result of the orderly stacked thylakoid membranes in grana and the highly birefringent starch granules in the chloroplasts. The stacked thylakoid membranes of grana and the orderly deposited amylopectin in the starch granules provide the structural requirement for efficient SHG, resembling the behavior of photonic crystals.

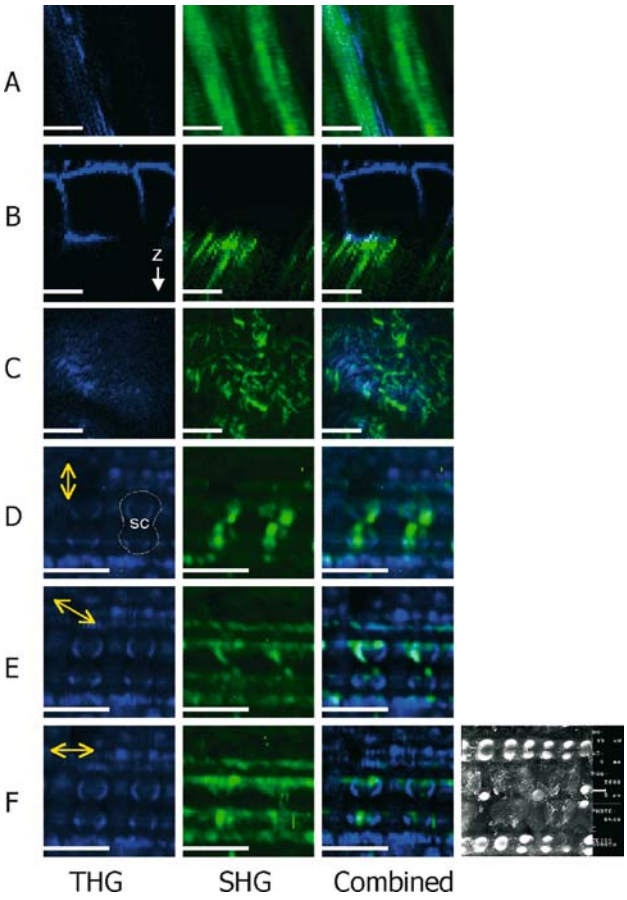


Fig. 11.5. Scanning THG (shown in *blue*), SHG (shown in *green*), and the combined multimodal images. Laser induced fluorescences are too weak to detect in these samples under our imaging condition. Images were taken from **(a)** (x-y images) longitudinal optical section of mouse skeletal muscle **(b)** (x-z images) optical cross-section of mouse dermis **(c)** (x-y images) corresponding to (b) at a constant depth of 30 μm with wavy structures showing collagen fiber bundles. **(c)**, **(e)**, **(f)** adaxial surface of rice leaf showing dumb-bell-shaped silica cells. Paradermal SEM image taken from a rice leaf is provided for correlation purpose (lower right). Scale bar: 15 μm. *Yellow arrows* in (**d**), (**e**), (**f**): direction of illumination polarization. *sc*: silica cell

Since SHG occurs from the nano-crystalline structures under study, it is highly desirable to use other noninvasive harmonic generation effect to image general cellular features so that proper correlation of structures can be obtained. Due to the optical dispersion property in biological tissues, THG was proven to be generated from regions with optical inhomogeneity [32–34] and was applied to image general cellular structures [26,35–37]. In the next section, we will discuss and compare THG microscopy with SHG microscopy in biological specimens.

11.4 THG Microscopy

THG is generally a weak process but is dipole allowed; therefore it occurs in all materials, including materials with inversion symmetry. In 1995, Tsang [32] reported that when using focused high intensity ultrashort laser pulses, this normally weak THG process becomes highly operative at a simple air-dielectric interface and is much stronger then the bulk of most dielectric materials. It was also reported [32] that the interface THG is a fundamental physical process occurring at all interfaces free from the constraint of a phase-matching condition and wavelength restriction. This could be explained by Boyd in his book "Nonlinear Optics" [33] by treating THG process with a focused Gaussian beam. According to Boyd, the efficiency of THG vanishes when focused inside a uniform sample for the case of positive phase mismatch ($\Delta k = k_3 - 3k_1 > 0$, true for normally dispersive materials) or even phase matching ($\Delta k = 0$). Efficient THG will only occur when the laser beam was focused in some inhomogeneous regions like interfaces, unless the materials possessing anomalous dispersion. Taking advantage of this characteristic, in 1997 Silberberg and his coworkers demonstrated THG microscopy for interface images in an optical glass fiber [34]. THG microscopy was then applied to image laser-induced breakdown in glass [38], liquid crystal structures [39], and defect distribution in GaN semiconductors [7].

For THG applications in biology, in 1996 Alfano and his coworkers found THG in chicken tissues using a 10 Hz 1064 nm Nd:YAG laser with 30 ps pulsewidth [14]. After the first demonstration of THG microscopy [34], it was then quickly applied to image biological specimen due to its interface sensitivity, by using 1200 nm femtosecond light from an optical parametric amplifier after a Ti:sapphire amplifier with 250 kHz repetition rate [34,35], 1500 nm femtosecond light from an optical parametric oscillator synchronously pumped by a Ti:sapphire laser with 80 MHz repetition rate [36], or with 1230 nm femtosecond light directly from a Cr:forsterite laser with 110 MHz repetition rate [26]. Detailed cellular and subcellular organelles can be clearly resolved with THG even with dynamic imaging. With THG imaging on Chara plant rhozoids, non-fading image can be achieved with continuous viewing, indicating prolonged viability under the viewing condition with 1200 nm light [35]. Recently THG microscopy even allows one to temporally visualize the release of Ca^{2+} [40]. With cubic dependence on the illumination intensity, THG provides even better optical sectioning resolution than SHG or 2PF using the same illumination wavelength, but is more sensitive to attenuation to the illumination light [26].

In order to allow THG within the visible spectrum, but also provide advantages of low attenuation for illumination, we move the THG excitation wavelength to 1230 nm regimes by using a 100-fs-pulsewidth Cr:forsterite laser. Like SHG microscopy, transmission detection in THG microscopy is preferred. Fig. 11.6 shows a THG image of rat hepatocytes taken with 1230 nm light from a Cr:forsterite laser. General cellular and subcellular organelles can be clearly resolved with THG due to its superior capability to image interfaces and inhomogeneity.

It is also interesting to compare THG with its corresponding SHG images. In general, SHG reflects the orderly arranged nano-structures while THG gives general interfacial profiles but requires inhomogeneity. For example, strong THG is induced due to optical inhomogeneity within and surrounding the cell walls (Fig. 11.4 (a)), providing images not just from the longitudinal cell walls but especially from the transverse cell walls due to their high spatial fluctuations in the light propagation direction (z-direction), with the ability to pick up the whole cell profile. In Fig. 11.5 (b), THG reveals the cell profile of epithelial cells. In Figs. 11.4 (d) and 11.5 (d)–(f), THG picks up structural interfaces, such as the papillae from the cuticular layer and the boundary of silica cells. THG can even pick up the bottom surface of cover glass as shown in the top of the upper-right corner image of Fig. 11.4. Not just cell profiles and structure interfaces, THG also provides information on the various sub-organellar interfaces (Fig. 11.5 (b), (c) and Fig. 11.6).

On the other hand, in highly organized skeleton muscles and connective tissues, SHG seems to provide better signal intensity (middle of Fig. 11.3) and contrast (Fig. 11.5 (a)–(c)) compared with THG. Due to optical homogeneity inside some animal tissues, very weak THG (middle of Fig. 11.3) can only be picked up in interfaces like the longitudinal interface of myofibril and surrounding cytoplasm due to induced optical inhomogeneity (Fig. 11.5 (a)). It is also interesting to notice that, since THG is allowed for isotropic materials and can be applied to image general interfaces, thus has weaker dependence on incident light polarization. No polarization dependency was found in THG images of Fig. 11.5 (d)–(f).

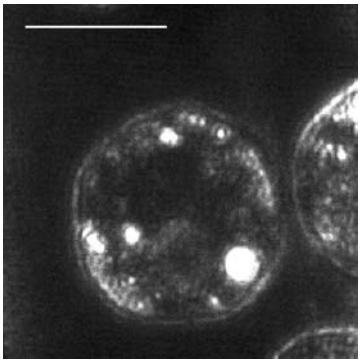


Fig. 11.6. Scanning (x-y) THG image of rat hepatocytes. General cellular and subcellular organelles can be clearly resolved. Scale bar: 20 μm

11.5 Conclusion

Light microscopy, fluorescence in particular, has been used extensively in the correlation between structures, molecules, and functions in modern biological sciences. Either intrinsic or extrinsic fluorescent probes are needed in fluorescence microscopy using single- or multi-photon excitation schemes. Hence, dye penetration, probe toxicity, and photo-bleaching/damage are the limiting factors frequently encountered. Different from fluorescence processes, SHG and THG processes are based on virtual transitions. The marked advantage of this virtual transition during wavelength conversion is the lack of energy deposition, thus no photo-damage or bleaching from the process is expected, can be considered as a truly "noninvasive" imaging modality. Due to its virtual nature, no saturation or bleaching in the harmonic generation signal is expected, thus ideal for live samples without preparation or with minimum staining under prolonged viewing condition. SHG microscopy can provide images on stacked membranes and arranged proteins with organized nano-structures due to the bio-photonic crystal effect or membrane potential imaging. On the other hand, THG microscopy can provide general cellular or subcellular interface imaging due to optical inhomogeneity.

In order to accomplish the "non-invasive" characteristic of harmonic generation imaging on live specimen, it is also essential to avoid other side effects induced by the excitation light. Our wavelength selection is based on previous studies of the attenuation spectra of living specimen indicating a light penetration window between 400–1300 nm. In order to have THG in the visible range, we choose the excitation light in the 1200–1300 nm range. Longer excitation wavelength will cause more water absorption and reduce its penetration capability. The quest for optimized wavelength and viewing condition will be a challenge for THG microscopy. Our recent studies comparing photo-damages induced by 1230 nm femtosecond light and 780 nm femtosecond light indicate suppressed multi-photon autofluorescence and suppressed multi-photon damages with longer 1230 nm light in plant materials [41], due to the fact that multi-photon absorption cross-section decreases with longer wavelength.

Sometimes it is also important to combine harmonic generation imaging modalities with multi-photon fluorescence modality due to the fact that the later is equipped with the capability to image different molecular distributions. For example, combining multi-photon fluorescence imaging modes, SHG microscopy is useful for investigating the dynamics of structure-function relationship at the molecular and sub-cellular levels. However, with an excitation wavelength longer than 1200 nm, the multiphoton absorption cross-section will also be decreased. For instance, blue-green fluorescence dyes and green-fluorescence protein will require 3-photon excitation, thus with much weaker fluorescence intensity [42]. It is therefore also important to develop efficient long excitation wavelength or long emission wavelength dyes for simultaneous multi-modal microscopy. Our recent study has indicated efficient multi-photon fluorescence for some common dyes under 1230 nm excitation condition [43].

Other restriction of harmonic generation microscopy includes its imaging geometry. Due to momentum conservation law, transmission detection is preferred for SHG and THG microscopies. Even though this might cause difficulties for some applications requiring reflection mode, it also possesses some advantages. The forward emission properties of harmonic generation allow efficient collection in transmission geometry, in contrast to inefficient collection of 4π fluorescence emission. The excitation light and the emission light will take different paths in transmission detection, thus allows thicker penetration depth with controlled sample thickness. On the other hand, in a highly scattering media, the reflection detection of the short-wavelength harmonic-generation light should also be possible due to the fact that light scattering increases with shorter wavelength.

References

1. R. Hellwarth and P. Christensen: Opt. Commun. **12**, 318 (1974)
2. C. J. R. Sheppard, R. Kompfner, J. Gannaway, and D. Walsh: IEEE J. Quantum Electron. **13**, 100 (1977)
3. G. T. Boyd, Y. R. Shen, and T. W. Hänsch: Opt. Lett. **11**, 97 (1986)
4. Y. Uesu, S. Kurimura, and Y. Yamamoto: Appl. Phys. Lett. **66**, 2165 (1995)
5. R. Gauderon, P. B. Lukins, and C. J. R. Sheppard: Opt. Lett. **23**, 1209 (1998)
6. We follow the notation of H. A. Haus: *Waves and Fields in Optoelectronics* (Prentice-Hall, Inc., New Jersey 1984)
7. C.-K. Sun, S.-W. Chu, S.-P. Tai, S. Keller, U. K. Mishra, and S. P. DenBaars: Appl. Phys. Lett. **77**, 2331 (2000)
8. C.-K. Sun, S.-W. Chu, S. P. Tai, S. Keller, A. Abare, U. K. Mishra, and S. P. DenBaars: J. Scanning Microscopies **23**, 182 (2001)
9. F. Bresson, F. E. Hernandez, J.-W. Shi, and C.-K. Sun: 'Electric-field induced second harmonic generation in nematic liquid crystals as a probe for electric field', In: *Proceeding of Optics and Photonics Taiwan '01*, ed. by (Kaohsiung, Taiwan ROC 2001) pp. 713–715
10. F. Bresson, C.C. Chen, F.E. Hernandez, J.-W. Shi, and C.-K. Sun: submitted to Rev. Sci. Instrum.
11. S. Fine and W. P. Hansen: Appl. Optics **10**, 2350 (1971)
12. I. Freund, M. Deutsch, and A. Sprecher: Biophys. J. **50**, 693 (1986)
13. J. Y. Huang, Z. Chen, and A. Lewis: J. Phys. Chem. **93**, 3314 (1989)
14. Y. Guo, P. P. Ho, A. Tirksliunas, F. Liu, and R. R. Alfano: Appl. Optics **35**, 6810 (1996)
15. Y. Guo, P. P. Ho, H. Savage, D. Harris, P. Sacks, S. Schantz, F. Liu, N. Zhadin, and R. R. Alfano: Opt. Lett. **22**, 1323 (1997)
16. O. Bouevitch, A. Lewis, I. Pinevsky, J. P. Wuskell, and L. M. Loew: Biophys. J. **65**, 672 (1993)
17. P. J. Campagnola, M. D. Wei, A. Lewis, and L. M. Loew: Biophys. J. **77**, 3341 (1999)
18. G. Peleg, A. Lewis, M. Linial, and L. M. Loew: P. Natl. Acad. Sci. USA **96**, 6700 (1999)
19. L. Moreax, O. Sandre, and J. Mertz: J. Opt. Soc. Am. B **17**, 1685 (2000)
20. F.-J. Kao, Y.-S. Wang, M.-K. Huang, S. L. Huang, P.-C. Cheng: P. Soc. Photo–Opt. Inst. **4082**, 119 (2000)
21. T. Zhao, Z.-H. Chen, F. Chen, W.-S. Shi, H.-B. Lu, and G.-Z. Yang: Phys. Rev. B **60**, 1697 (1999)

22. Y. Dumeige, P. Vidakovic, S. Sauvage, I. Sagnes, J. A. Levenson, C. Sibilia, M. Centini, G. D'Aguanno, and M. Scalora: Appl. Phys. Lett. **78**, 3021 (2001)
23. V. Berger: Phys. Rev. Lett. **81**, 4136 (1998)
24. N. G. R. Broderick, G. W. Ross, H. L. Offerhaus, D. J. Richardson, and D. C. Hanna: Phys. Rev. Lett. **84**, 4345 (2000)
25. K. Clays. S. Van Elshocht, M. Chi, E. Lepoudre, and A. Persoons: J. Opt. Soc. Am. B **18**, 1474 (2001)
26. S.-W. Chu, I-H. Chen, T.-M. Liu, B.-L. Lin, P. C. Cheng, and C.-K. Sun: Opt. Lett. **26**, 1909 (2001)
27. T.-M. Liu, S.-W. Chu, C.-K. Sun, B.-L. Lin, P. C. Cheng, and I. Johnson: J. Scanning Microscopies **23**, 249 (2001)
28. R. R. Anderson and J. A. Parish: J. Invest. Dermat. **77**, 13 (1981)
29. P. C. Cheng, S. J. Pan, A. Shih, K.-S. Kim, W. S. Liou, and M. S. Park: J. Microscopy **189**, 199 (1998)
30. B. E. Bouma, G. J. Tearney, I. P. Bilinsky, B. Golubovic, and J. G. Fujimoto: Opt. Lett. **21**, 1839 (1996)
31. D. J. Gallant, B. Bouchet, and P. M. Baldwin: Carbohydrate Polymers **32**, 177 (1997)
32. T. Y. F. Tsang: Phys. Rev. A **52**, 4116 (1995)
33. R. W. Boyd: *Nonlinear Optics* (Academic Press, San Diego, CA 1992)
34. Y. Barad, H. Eisenberg, M. Horowitz, and Y. Silberberg: Appl. Phys. Lett. **70**, 922 (1997)
35. M. Müller, J. Squier, K. R. Wilson, and G. J. Brakenhoff: J. Microscopy **191**, 266 (1998)
36. J. A. Squier, M. Müller, G. J. Brakenhoff, and K. R. Wilson: Opt. Express **3**, 315 (1998)
37. D. Yelin and Y. Silberberg: Opt. Express **5**, 169 (1999)
38. J. A. Squier and M. Müller: Appl. Optics **38**, 5789 (1999)
39. D. Yelin, Y. Silberberg, Y. Barad, and J. S. Patel: Appl. Phys. Lett. **74**, 3107 (1999)
40. L. Canioni, S. Rivet, L. Sarger, R. Barille, P. Vacher, and P. Voisin: Opt. Lett. **26**, 515 (2001)
41. I-H. Chen, S.-W. Chu, C.-K. Sun, P. C. Cheng, and B.-L. Lin: 'Wavelength dependent cell damages in multi-photon confocal microscopy'. In: *Optical and Quantum Electronics* (in press 2002)
42. T.-M. Liu, S.-W. Chiu, I-H. Chen, C.-K. Sun, B.-L. Lin, W.-J. Yang, P.-C. Cheng, and I. Johnson: 'Multi-photon fluorescence of green fluorescence protein (GFP) and commonly used bio-probes excited by femtosecond Cr:forsterite lasers', In: *Proceedings of 2nd International Photonics Conference*, ed. by Hsinchu (TAIWAN, 2000 119) pp. 121–
43. T.-M. Liu, S.-W. Chu, C.-K. Sun, B.-L. Lin, P. C. Cheng, and I. Johnson: J. Scanning Microscopies **23**, 249 (2001)

Part III

Miscellaneous Methods in Optical Imaging

12 Adaptive Optics

J. Chris Dainty

12.1 Introduction

The principle of adaptive optics is straightforward and is depicted in Fig. 12.1 in the context of astronomical imaging. Because of atmospheric turbulence, the wavefront from a pointlike object is continuously distorted. The purpose of the adaptive optics (AO) system is to remove this distortion and provide diffraction-limited imaging if possible. This is achieved by sensing the distortion using a *wavefront sensor* and compensating for it using a *deformable mirror*. The whole system operates in closed loop with the aid of a suitable *control system*. These are the three key elements of an adaptive optics system. The principle of operation is simple: the control system drives the deformable mirror so that the error in the wavefront is minimised. The "devil lies in the detail" of every individual part of the system and of the way they interact with each other and with the incoming dynamically distorted wave.

In this chapter, I review the basic elements of adaptive optics with special reference to the fundamental limitations rather than to the technology. This is preceeded by a brief (and selective) discussion of the history of adaptive optics. In order to

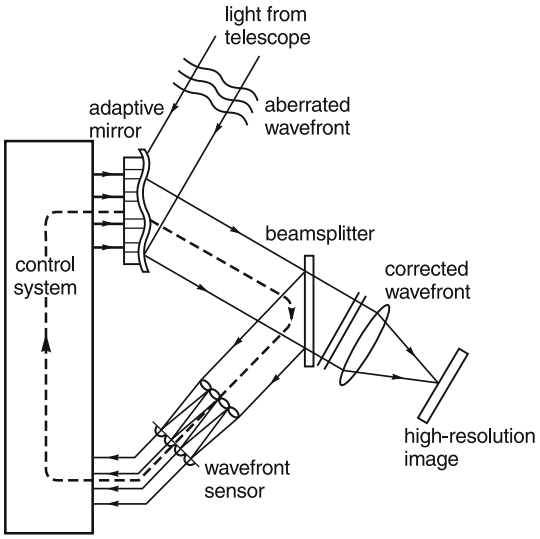


Fig. 12.1. Schematic arrangement of an adaptive optics system

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

move the subject into the mainstream of optics – so that it becomes another tool for the construction of optical system – it is important to understand why some people felt (and others still believe) that it was necessary to spend millions of dollars to build adaptive optics systems. This chapter concludes with a short description of two of the key current issues in astronomical adaptive optics (laser guide stars and multiconjugate AO) and of the potential applications of adaptive optics in the human eye.

12.2 Historical Background

The origins of adaptive optics in the "modern" era can be traced back nearly fifty years to a paper by an astronomer, Horace Babcock [1]. The angular resolution of groundbased telescopes is limited by the optical effects of atmospheric turbulence. The loss in resolution is very large as the following discussion illustrates.

The statistics of the phase perturbation due to atmospheric turbulence are characterised by a scale length r_0, the Fried parameter [2] . A telescope with a diameter equal to the Fried parameter r_0 has a mean square (piston-removed) phase fluctuation of approximately 1 rad^2, and the image quality, provided that the two tilt components are removed, is high (a Strehl ratio [3] of ≈0.87, as shown by Noll [4]). The quantity r_0 is therefore also approximately equal to the diameter of a diffraction-limited telescope that gives the same angular width of the point spread function as that produced when imaging through turbulence whose phase statistics are characterised by r_0. If the atmospheric turbulence is characterised by a small value of r_0, then the image produced in a large telescope in the presence of turbulence will only have an angular image size given by that of a small telescope (of diameter r_0).

As is so often the case in adaptive optics, the key issues depend critically on the numerical values of certain parameters ("the devil lies in the detail"). What is a typical value for the Fried parameter r_0? Using the Kolmogorov model [5] for the refractive index fluctuations in the atmosphere and the assumption that the turbulence is "weak", it can be shown [6] that the Fried parameter is given by the expression:

$$r_0 = \left(0.42k^2\,(\cos\gamma)^{-1}\int_0^{\infty} C_{\mathrm{N}}^2(z)dz\right)^{-3/5} \tag{12.1}$$

In this equation, k is the wavenumber ($= 2\pi/\lambda$), γ is the angle of observation measured from the zenith and $C_{\mathrm{N}}^2(z)$ in the refractive index structure function "constant" as a function of altitude z. One of the consequences of (12.1) is that r_0 is proportional to the six-fifths power of the wavelength:

$$r_0 \propto \lambda^{6/5}\,. \tag{12.2}$$

At an excellent observing site, the numerical value of r_0 in the visible region of the spectrum is between 10 and 20 cm, and this corresponds to values in the range 50–100 cm in the K band (≈ 2.2 μm). So, returning now to Babcock's era, the 5 m diameter telescope at Mt. Palomar (where r_0 in the visible would typically be ≈ 10 cm)

would have an angular resolution roughly equal to that of a 10 cm diameter amateur telescope, that is, about 1 arc-second, compared to its theoretical limit some fifty times smaller, 0.02 arc-sec. The effect of atmospheric turbulence in large telescopes totally dominates their angular resolution, and the degradation in resolution caused by turbulence is very severe indeed.

In order to overcome the deleterious effects of atmospheric turbulence, Babcock suggested the closed loop arrangement schematically shown in Fig. 12.1 with a Schlieren optical system as the wavefront sensor and an Eidophor (a system using electric charge to deform an oil film) as the corrective device. A working system was not built. At least two problems existed at the time: firstly, the technology, particularly that of the corrective device was simply not up to the job, and secondly there was the suspicion that this principle would only work on bright stars. We return to this latter issue in the following Section.

Although Babcock did not make an adaptive optics system, he had suggested how the effects of atmospheric turbulence might be overcome in real-time in astronomy and indeed in any application which involved imaging through turbulence. As illustrated above, the potential gain in angular resolution is very large, and this benefit helped to justify the large investment made by the US Government from the early 1970s onwards in adaptive optics. Some details of this are given by J.W. Hardy [7] of Itek Corporation, who pioneered the construction first of a 21-actuator "real time atmospheric compensator" and later of a 168-actuator "Compensated Imaging System" on the 1.6 m diameter satellite tracking telescope at Mt. Haleakala in Maui, Hawaii, commissioned in 1982. Even by modern standards, this is an impressive system, although its initial limiting magnitude – an issue of crucial importance for astronomy – was only about $m_v \approx 7^{th}$.

The challenge of implementing adaptive optics in astronomy was first taken up by a group of French observatories working together with the European Southern Observatory. The first system was a 19-actuator system called COME-ON [8] installed at the 3.6 m diameter telescope at La Silla, Chile, later upgraded to 52-actuators and with user-friendly software (ADONIS) with a limiting magnitude on the order of $m_v \approx 13^{th}$. These systems, in common with all other astronomical AO systems built to date, operate primarily in the infrared. As a result of the wavelength dependence of r_0 (12.2), and also because the required bandwidth is smaller and the isoplanatic angle is larger, operation at longer wavelengths is easier, and, in particular, brighter guide stars can be used. Figure 12.2 shows one image obtained from this system [9], illustrating the remarkable increase in information that the increased resolution – even in the infrared – that adaptive optics brings to astronomy.

Not surprisingly, all large telescopes are now equipped with adaptive optics systems operating in the near infrared. The case for implementing adaptive optics is overwhelming, and justifies the relatively large cost of $1–5M. Astronomical AO systems do not yet operate routinely in the visible region of the spectrum and in a later section we explore some of the current research issues in astronomical AO which are working towards wide field, full sky coverage in the visible.

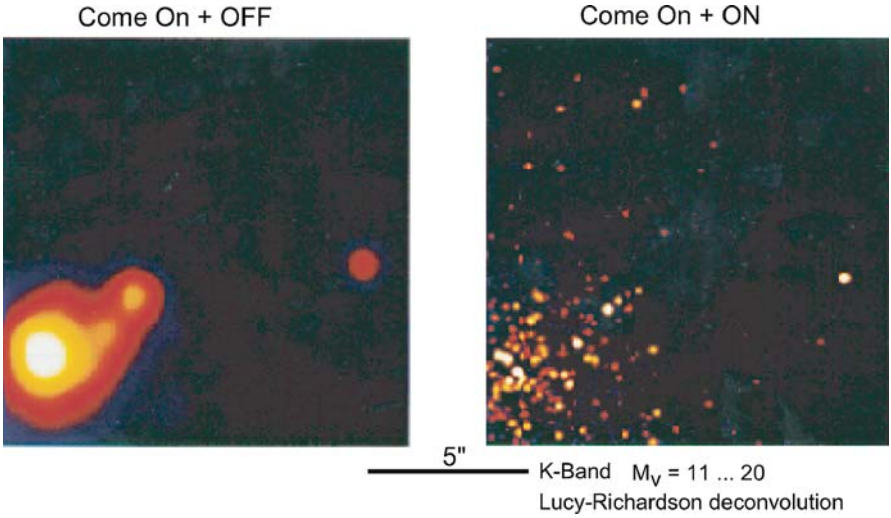


Fig. 12.2. Starburst region R136 in Doradus 30 [9]. *Left*: Adaptive optics off, *Right*: Adaptive optics on, with additional Richardson–Lucy deconvolution. Courtesy of Bernhard Brandl and Andreas Quirrenbach

The adaptive optics systems built for astronomy and for military systems are individually designed and constructed "one-off" instruments, operating to the highest possible performance given current technology. Particularly in astronomy, the building of an AO system is a high-profile project that cannot be allowed to fail. In addition, AO systems in astronomy have to be "user-instruments", i.e. be able to be operated by visiting astronomers not familiar with the detailed workings of adaptive optics: this means that the user-interface software has to be of a very high standard. All these facts contribute to the very large cost of astronomical AO systems, but these high costs are justifiable in terms of the scientific productivity of the finished instruments.

However, there are many subjects other than astronomy that would benefit enormously from adaptive optics, but in order to justify the use of AO in these cases it is essential that the cost be reduced dramatically, certainly to thousands if not hundreds of dollars. In order to reduce the costs by several orders of magnitude needs both a proper understanding of the essentials of adaptive optics as well as a different approach to that used in astronomy and in military applications. In 2000, two groups demonstrated that low cost systems could be built successfully [10,11] and now it is generally accepted that high quality, high bandwidth adaptive optics is achievable at much lower cost than previously imagined. This is due in part to developments in technology and in part to a more risk-oriented approach to research not found in the big-science area of astronomy.

12.3 Strehl Ratio and Wavefront Variance

The effect of a *small* aberration in an imaging system is to lower the central intensity of the point spread function and redistribute the energy, in particular, filling in the zeros, as shown schematically in Fig. 12.3. The **Strehl ratio** S [3] is the ratio of the central intensity with the aberration to the central intensity in the absence of the aberration and a value of 0.8 was proposed by Strehl as providing a suitable criterion for the quality of an imaging system: if $S > 0.8$, the system is *effectively* diffraction-limited whereas if $S < 0.8$ the aberration is significant. For *small* aberrations, the Strehl ratio S and the wavefront phase variance σ_ϕ^2 are related by

$$S \approx 1 - \sigma_\phi^2 \tag{12.3}$$

and thus a value $S = 0.8$ corresponds to a wavefront phase variance $\sigma_\phi^2 = 0.2\,\mathrm{rad}^2$, or an actual wavefront variance $\sigma_\mathrm{W}^2 = \lambda^2/200$.

In adaptive optical systems, like other high-quality imaging systems, it is customary to use the Strehl ratio and/or the wavefront variance as a measure of the image quality. The sources of error – and there are many of them – in an AO system are frequently uncorrelated, so that the total variance of the wavefront error is the sum of the variances from each source of error, and because $1 - \sigma^2 \approx \exp -\sigma^2$, the Strehl factors for each source of error are multiplicative. (It should be stressed that this latter result is true only for small errors.)

It is very difficult to make an AO system for astronomical use with a Strehl ratio as high as 0.8, and more typical values for these systems are in the range 0.2 – 0.6.

In working out how to design an AO system, it is useful to know how different components affect the wavefront error. One might start with an understanding of the wavefront to be corrected. In his classic paper, Noll [4] evaluated the magnitudes of the mean square Zernike components for Kolmogorov turbulence, and then calculated the residual mean square phase error that would result if the Zernike components in the wavefront were successively corrected. Table 12.1 summarises his results. For example, if the telescope diameter $D = 5r_0$, i.e. $(D/r_0)^{5/3} \approx 15$, then exact correction of the first 10 Zernikes would yield a residual wavefront error of $\approx 15 \times 0.04 \approx 0.60\,\mathrm{rad}^2$.

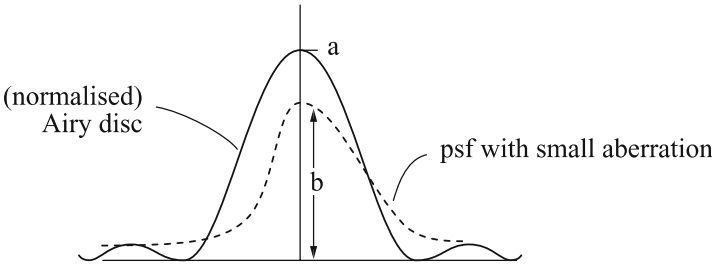


Fig. 12.3. Effect of a small aberration: Strehl ratio

Table 12.1. The residual phase variance (in rad^2) when the first J Zernike coefficients of Kolmogorov turbulence are corrected exactly (after Noll [4]). The values of the phase variance given are for $D = r_0$ and scale as $(D/r_0)^{5/3}$

J	Squared error	J	Squared Error
1	1.030	12	0.0352
2	0.582	13	0.0328
3	0.134	14	0.0304
4	0.111	15	0.0279
5	0.0880	16	0.0267
6	0.0648	17	0.0255
7	0.0587	18	0.0243
8	0.0525	19	0.0232
9	0.0463	20	0.0220
10	0.0401	21	0.0208
11	0.0377		

Whilst Noll's paper sheds light on the idea of correcting successive components of the aberration, and thus improving the final image quality in an adaptive optics system, it has also been the source of some misunderstanding, since it focuses on the Zernike expansion of the wavefront. It is well-known that the Zernike expansion is particularly suited to the description of the aberrations of a rotationally symmetrical optical system, and for this reason it has a special place in traditional optics. However, as we shall see in Sec. 12.6, the Zernike expansion has no special role in adaptive optics, and other expansions, such as those involving sensor or mirror modes have a much more important role. In particular, when operating a closed loop AO control system, there is no benefit or need to consider Zernike modes.

12.4 Wavefront Sensing

The purpose of the wavefront sensor in an adaptive optics system is to measure the wavefront with the required spatial and temporal sampling. A generic wavefront sensor consists of some kind of optical system followed by a detector (or set of detectors) and gives rise to (at least) three sources of error, each specified as a wavefront variance: a spatial wavefront sampling error σ_s^2, a temporal bandwidth error σ_t^2 and a detector noise error σ_d^2. The temporal error is usually lumped together with other temporal errors (for example, those in the control system) as a bandwidth error σ_{bw}^2, and the detector error frequently divided into several terms, such as the errors due to CCD read noise, photon noise, etc. The important point here is to realise that these errors are of crucial importance in determining the effectiveness of the adaptive optics system, and in general the variance of each error has to be *very* much smaller than one rad^2, so that the combined error from all sources adds up to less than approximately one radian squared.

The *ideal* wavefront sensor would have no detector noise, would not introduce any time delay into the measurement process, and would be capable of sensing only

those wavefront deformations that the mirror can produce. The first two requirements are clearly unattainable. The third requirement can be met in some cases by closely matching the wavefront sensor to the deformable mirror, for example using a wavefront curvature sensor for a bimorph or membrane mirror. The sensor must be capable of sensing all the mirror modes but there is little point in it being capable of sensing modes that the mirror cannot correct (see Sec. 12.6). Thus a wavefront sensor optimised for a particular closed-loop AO system is usually quite different from a wavefront sensor designed simply for measuring wavefronts in an open loop system (e.g. a wavefront sensor designed for metrology). In the latter case, one typically would use a much higher degree of spatial sampling.

There are a number of wavefront sensors that have been used in AO systems and several others have been suggested. One way of classifying them is by the quantity that is directly measured, and on this basis one can make the following classification:

- **Wavefront**: Interferometry, Common Path Interferometry [12]
- **Wavefront Slope**: Shearing Interferometer [13], Shack–Hartmann Sensor [14], Pyramid Sensor [15]
- **Wavefront Curvature**: Curvature Sensor [16], Phase Diversity [17,18]
- **Other**: Modal Sensing [19,20]

This is only a partial list, and special devices might be more suitable for specialist AO systems. Furthermore, devices can be used in different ways: for example, the Shack–Hartmann sensor can be modified to yield the wavefront curvature as well as the wavefront slope [21].

From an operational point of view, a better classification might be whether the wavefront sensor is adaptable in its sensitivity as an AO loop is closed. On this basis, the pyramid and curvature sensors would be in a class of their own, and this variable sensitivity allows them to reach limiting magnitudes up to ten times fainter in astronomical applications. The most common wavefront sensor in adaptive optics is the Shack–Hartmann device, largely because of its simplicity to implement and to understand. The principle of operation is shown in Fig. 12.4. It consists of a lenslet array in front of a pixellated detector such as a CCD. A plane wave incident upon the device produces an array of spots on a precise matrix determined by the geometry of the lenslet array, and the centroids of these spots can be found by computation. A distorted wave also gives an array of spots but the centroid of each spot in now determined by the average slope of the wavefront over each lenslet, and hence by finding the centroids of every spot, the wavefront slope at each point can be found. The wavefront can be reconstructed from the slope information by a least squares or other method [22]. Figure 12.5 gives an example of a spot pattern and the reconstructed wavefront for the case of measuring the wavefront aberration of the eye. The detector error in wavefront sensing is one that dominates the error budget of astronomical adaptive optics systems. In this case, the read noise of the CCD is significant, and special high-speed, low noise multi-port CCDs are specially developed for this purpose. The effect of detector noise on wavefront sensing is discussed in [23]. As a rule-of-thumb for astronomy, at least 100 detected photons per spot are

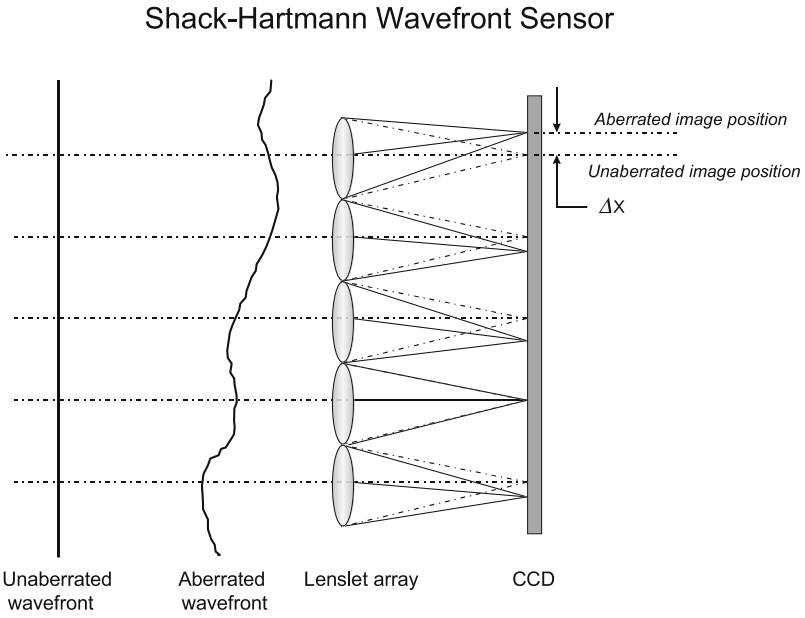


Fig. 12.4. The Shack–Hartmann wavefront sensor

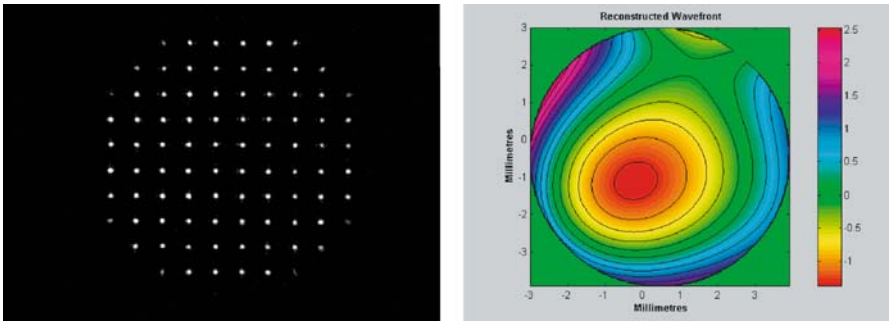


Fig. 12.5. Sample spot pattern in a Shack–Hartmann sensor (*left*) and the reconstructed wavefront (*right*). Courtesy of D. P. Catlin

required to close the loop with a low wavefront variance due to photon noise: however, one can still, in principle, close the loop with as few as 10 detected photons per spot provided one accepts a higher wavefront variance due to photon noise. In other applications, noise sources other than detector noise may be more important, such as the speckle noise in the case of wavefront sensing in the eye at high speed.

The spatial sampling error of the Shack–Hartmann sensor is governed by the lenslet dimensions. As a guideline, in astronomical applications, the lenslet side is on the order of the Fried parameter r_0, and the uniform mesh of the lenslet array is well-suited to the isotropic nature of the atmophere-induced aberration to be cor-

rected. In contrast, for aberrations that are non-isotropic, e.g. increase in magnitude towards the edge of the pupil, as in the eye, the uniform sampling is not ideal. In non-astronomical applications, there are probably significant new methods of wavefront sensing to be invented for adaptive optics.

12.5 Deformable Mirrors and Other Corrective Devices

There are many possible devices that can be used to modulate the wavefront and hence provide correction for dynamic aberrations. These include:

- Zonal Continuous Facesheet Mirrors [24]
- Bimorph Mirrors [25,26]
- Membrane Mirrors [27]
- MEMS Mirrors [28,29]
- Liquid Crystal Devices [30–32]

In a zonal mirror, the influence function $r(x, y)$ of the mirror is localised around the location of the actuator, whereas in a modal device (such as a membrane or bimorph mirror) each actuator has an influence function that extends over the whole mirror. For low order AO, there is general agreement that modal corrective devices offer a number advantages; for example, focus correction can be very simply implemented. Until recently, the cost of a deformable and its power supply was formidable, in excess of $1000 per actuator but now low cost membrane mirrors [27] are available at approximately $100 per actuator for a 37-actuator device. These mirrors are rather fragile, and another modal technology is the bimorph mirror which has the potential for very low cost. In one form of bimorph mirror, shown in Fig. 12.6, a slice of piezo-electric material with an electrode pattern etched onto one of its conducting surfaces, is bonded to a glass substrate. This device tends to be temperature sensitive due to the differing coefficients of expansion of the glass and piezo material, and in an alternative structure two piezo slices are bonded to each other and an optical mirror replicated onto one of the surfaces. There is also great interest in the development of MEMS mirrors, although these are still in development and are not yet low cost devices [29,33].

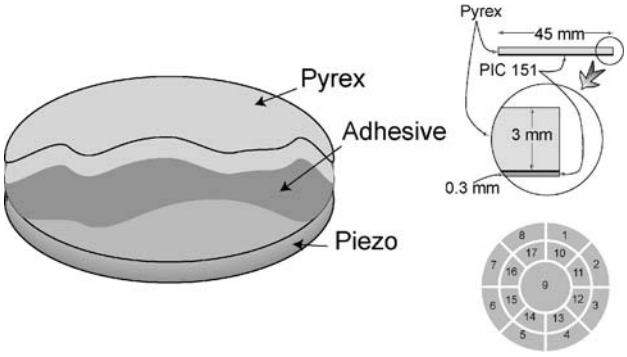


Fig. 12.6. The construction of a bimorph mirror

How many actuators are required to correct a given type of aberration? This is not a simple question to answer, the result clearly depending on the statistics of the aberration and the influence functions of the mirror (or liquid crystal) actuators. For atmospheric turbulence, there is a rule-of-thumb that the number N of actuators (and Shack–Hartmann lenslets in the wavefront sensor) is given by $N \approx (D/r_0)^2$, but this is a crude approximation that ignores the effectiveness of the influence functions to correct Kolmogorov wavefronts.

In order to be more specific, consider a corrective device with N_A actuators whose influence functions are $r_j(x, y)$, one for each actuator, labelled j. For a set of control signals, w_j applied to these actuators, the resulting phase[1] given to the surface of the mirror can be assumed to be the linear superposition of the influence functions

$$\Phi_\mathrm{M}(x, y) = \sum_{j=1}^{N_\mathrm{A}} r_j(x, y) w_j \,. \tag{12.4}$$

The phase $\Phi_\mathrm{M}(x, y)$ can be expanded in an orthonormal series and for convenience we use the Zernike basis as the pupils (and the mirrors) are usually circular and this basis is orthonormal over the unit circle. The Zernike coefficents ϕ_i are given by

$$\begin{aligned}\phi_i &= \iint \mathcal{W}(x, y)\Phi_\mathrm{M}(x, y)Z_i(x, y)\mathrm{d}x\mathrm{d}y \\ &= \sum_{j=1}^{N_\mathrm{A}} \left[\iint \mathcal{W}(x, y)Z_i(x, y)r_j(x, y)\mathrm{d}x\mathrm{d}y\right] w_j\end{aligned}$$

or in matrix form

$$\boldsymbol{\phi} = \mathbf{M}\,\boldsymbol{w} \tag{12.5}$$

where $\mathcal{W}(x, y)$ is the circular aperture function. The matrix $\mathbf{M}$ is called the influence matrix, and in this case each column of $\mathbf{M}$ is the Zernike expansion coefficients of a single influence function of the mirror. It is easily measured, by applying a fixed signal to each electrode sequentially, recording the wavefront phase, and computing the Zernike expansion of each influence function. Note that this particular version of the influence matrix $\mathbf{M}$ uses the Zernike expansion, but this is not esential.

Given a particular wavefront aberration whose Zernike expansion is specified by the vector $\boldsymbol{\phi}_0$, the control signals $\boldsymbol{w}$ required to give the best least squares fit are given by

$$\boldsymbol{w} = \mathbf{M}^{-1*}\boldsymbol{\phi}_0 \,,$$

where $\mathbf{M}^{-1*}$ is the least squares inverse of $\mathbf{M}$. The residual wavefront error after correction is

$$\boldsymbol{\phi}_\mathrm{e} = \mathbf{M}\,\mathbf{M}^{-1*}\boldsymbol{\phi}_0 - \boldsymbol{\phi}_0 = \left[\mathbf{M}\,\mathbf{M}^{-1*} - \mathbf{I}\right]\boldsymbol{\phi}_0 \,. \tag{12.6}$$

[1] We consider the continuous phase in the interval $-\infty$ to $+\infty$.

Equation (12.6) forms the basis of the assessment of a given mirror to correct given wavefronts or given statistics of wavefronts and the mirror fitting error σ^2_{fit} can be evaluated. It turns out that the "best" set of influence functions for any mirror is that which spans the subspace covered by the first N_{A} Karhunen–Loève functions of the aberration space. For atmospheric turbulence with Kolmogorov statistics, curvature mirrors (membrane, bimorph) have sets of influence functions which are much closer to the Karhunen–Loève functions than segmented or other zonal mirrors.

12.6 The Control System

The control system involves both spatial and temporal aspects, and in general these are coupled and should be considered together. Here, for simplicity, we separate the two aspects.

Before discussing the mathematics of the control system, some remarks on the hardware required are relevant. The first "controllers" were custom built circuits housed in several 19-inch racks. A second generation, found in a number of current astronomical AO systems, used many (20+) Texas C-40 or similar digital signal processing (DSP) chips. Such an approach may still be needed for systems with a very high bandwidth and a large number (several hundred) of actuators, but nowadays a single processor can handle the control for modest bandwidths and numbers of actuators. For example, in [11], a 500 MHz Pentium chip was used for a 5×5 Shack–Hartmann wavefront sensor and a 37-actuator mirror at a frame rate of 780 Hz: only a fraction of the processor time was actually used and calculations indicate that this could have handled a 10×10 Shack–Hartmann array (with an appropriate increase in the number of mirror actuators) at the same frame rate. In fact the major practical problem is the input rate (and, more importantly, any latency) from the wavefront sensor detector, rather than the matrix operations of the control system.

The operation of a wavefront sensor on an aberration described by the vector $\boldsymbol{\phi}_0$ (of, for example, Zernike coefficients) can be expressed by the matrix equation

$$\boldsymbol{s} = \mathbf{S}\boldsymbol{\phi}_0 \,, \tag{12.7}$$

where $\boldsymbol{s}$ is the vector of sensor signals and $\mathbf{S}$ is the wavefront sensor response matrix. For example, for a Shack–Hartmann sensor, the components of $\boldsymbol{s}$ are the $x-$ and $y-$slopes of the wavefront averaged over each lenslet (for N lenslets the vector has length $2N$), and the elements S_{ij} of $\mathbf{S}$ are the average $x-$ and $y-$differentials of the Zernike polynomials for the i^{th} lenslet and the j^{th} polynomial averaged over each lenslet.

The actuator control signals w_j form a vector $\boldsymbol{w}$ that are assumed to be a linear combination of the sensor signals, i.e.

$$\boldsymbol{w} = \mathbf{C}\boldsymbol{s} \tag{12.8}$$

where $\mathbf{C}$ is the control matrix.

There are many approaches to finding **C**. One approach, the least squares method, is outlined in the following. In a closed loop AO system, the wavefront entering the sensor is the sum of the aberrated wave $\boldsymbol{\phi}_0$ and the mirror-induced phase $\boldsymbol{\phi} = \mathbf{M}\boldsymbol{w}$. In the absence of noise, the wavefront sensor signal that results is given by

$$\boldsymbol{s} = \mathbf{S}\left[\boldsymbol{\phi}_0 + \mathbf{M}\boldsymbol{w}\right] = \boldsymbol{s}_0 + \mathbf{B}\boldsymbol{w} \tag{12.9}$$

where $\mathbf{B} = \mathbf{S}\,\mathbf{M}$ is the response matrix for the mirror-sensor system. The matrix **B** plays an important rôle in least squares control. Note that although both the sensor response matrix **S** and the actuator response matrix **M** were defined, in the above, in terms of a Zernike expansion, the overall response matrix is independent of this choice and does not involve Zernike polynomials. The matrix **B** is found by applying a fixed signal to each electrode of the mirror in turn and recording the sensor signals (for example, the $x-$ and $y-$slopes or simply the spot positions in a Shack–Hartmann sensor).

Since the wavefront sensor measures the net wavefront aberration after correction, we shall use this as a measure of the correction error. Using a standard least squares approach, it can be shown that the control matrix **C** that minimises the wavefront sensor error is

$$\mathbf{C} = -\left[\mathbf{B}^T\mathbf{B}\right]^{-1}\mathbf{B}^T \tag{12.10}$$

where the quantity $\left[\mathbf{B}^T\mathbf{B}\right]^{-1}\mathbf{B}^T$ is known as the least squares inverse, or psuedo-inverse, of **B**.

From (12.9) it can be seen that the matrix **B** defines a mapping between the actuator ($\boldsymbol{w}$, length N_A) and sensor ($\boldsymbol{s}$, length N_S) vectors. Any matrix of dimensions $N_\mathrm{S} \times N_\mathrm{A}$ can be written as a product of three matrices, the so-called *singular value decomposition*:

$$\mathbf{B} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^T \tag{12.11}$$

where (1) **U** is an $N_\mathrm{S} \times N_\mathrm{S}$ orthogonal matrix, (2) **V** is an $N_\mathrm{A} \times N_\mathrm{A}$ orthogonal matrix and (3) $\boldsymbol{\Lambda}$ is an $N_\mathrm{S} \times N_\mathrm{A}$ diagonal matrix.

Equation (12.11) can be expanded as

$$\mathbf{B} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \ldots \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ldots & \\ & & & \lambda_N \end{pmatrix} \begin{pmatrix} \boldsymbol{v}_1^T \\ \boldsymbol{v}_2^T \\ \ldots \end{pmatrix} \tag{12.12}$$

where the vectors $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ form complete sets of modes for the sensor signal and mirror control spaces respectively. The diagonal elements of $\boldsymbol{\Lambda}$, λ_i are the singular values of the matrix **B**. Each non-zero value of λ_i relates the orthogonal basis component $\boldsymbol{v}_i$ in $\boldsymbol{w}$, the control signal space, to an orthogonal basis component $\boldsymbol{u}_i$ in $\boldsymbol{s}$, the sensor signal space.

We can now distinguish three possible situations:

- **Correctable Modes** For these modes, $\lambda_i \neq 0$, the actuator control signal $\boldsymbol{w} = \boldsymbol{v}_i$ results in the sensor signal $\boldsymbol{s} = \lambda_i \boldsymbol{u}_i$. This mode can be corrected by applying the actuator model $\boldsymbol{w} = \lambda_i^{-1} \boldsymbol{v}_i$, the singular value λ_i being the sensitivity of the mode (clearly we not want λ_i to be too small).
- **Unsensed Mirror Modes** These are modes $\boldsymbol{v}_i$ for which there is no non-zero λ_i. Unsensed mirror modes would cause a big problem in an AO system and are to be avoided if at all possible by proper design of the wavefront sensor: some of the early zonal AO systems suffered from this defect, producing so-called "waffle" modes in the mirror.
- **Uncorrectable Sensor Modes** These are modes $\boldsymbol{u}_i$ for which there is no non-zero λ_i. Nothing the mirror does affects these modes, and arguably there is no point in measuring them.

The control matrix $\mathbf{C}$ is now given by

$$\mathbf{C} = -\mathbf{B}^{-1*} = -\mathbf{V}\boldsymbol{\Lambda}^{-1*}\mathbf{U}^T \tag{12.13}$$

where $\boldsymbol{\Lambda}^{-1*}$ is the least-squares pseudo-inverse of $\boldsymbol{\Lambda}$ formed by transposing $\boldsymbol{\Lambda}$ and replacing all non-zero diagonal elements by their reciprocals λ_i^{-1}, which now can be interpreted as the *gains* of the system modes. From a practical point of view, one discards modes which have a small value for λ_i as clearly they are susceptable to noise. An example of the effect of discarding modes is presented in Sec. 12.7.

The least squares approach is simple and does not require much prior information about either the AO system or the statistics of the aberration to be corrected. It also has the advantage that the key matrix $\mathbf{B}$ can easily be measured. But surely prior knowledge, for example about the aberration to be corrected or the noise properties of the sensor, could be an advantage in a well-designed control system? There are several other approaches to "optimal reconstructors", originating with the work of Wallner [34], and described in [35]. These are not so easy to implement but may offer improved performance. The approach to understand the temporal control aspects is through the transfer functions of each component [36]. Figure 12.7 shows four of the key elements which can play an important rôle in the time behaviour.

The wavefront sensor detector integrates for a time T to record the data, and therefore has a transfer function (Laplace transform),

$$G_{\mathrm{WS}}(s) = \frac{1 - \exp(-Ts)}{Ts}$$

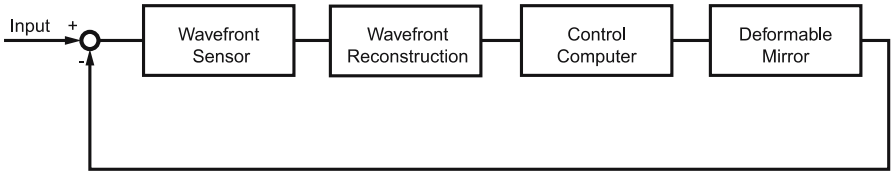


Fig. 12.7. Schematic of a closed loop adaptive optics system

whereas the wavefront reconstructor introduces a lag τ:

$$G_{\mathrm{WR}}(s) = \exp(-\tau s)\,.$$

We assume that the deformable mirror has a unit transfer function (i.e. it responds immediately)

$$G_{\mathrm{DM}}(s) = 1\,.$$

Finally, the control computer, if it implements a simple integrator, has a transfer function

$$G_{\mathrm{C}}(s) = \frac{K}{s}\,,$$

where K is the gain of the integrator. The open loop transfer function G_{ol} is the product of the transfer functions:

$$\begin{aligned} G_{\mathrm{ol}}(s) &= G_{\mathrm{WS}}(s) \times G_{WR}(s) \times G_{\mathrm{DM}}(s) \times G_{\mathrm{C}}(s) \\ &= \frac{K\left[1 - \exp(-Ts)\right]\exp(-\tau s)}{Ts^2} \end{aligned} \tag{12.14}$$

and the closed loop transfer function $G_{\mathrm{cl}}(s)$ is given by:

$$G_{\mathrm{cl}}(s) = \frac{G_{\mathrm{ol}}(s)}{1 + G_{\mathrm{ol}}(s)H(s)} \tag{12.15}$$

where $H(s)$ is the feedback parameter.

The closed loop bandwidth can be defined in a number of ways. One way is to define it as the frequency at which the closed loop transfer function falls to the −3 dB level. However a more informative measure is found by plotting the ratio of the input and output signals as a function of frequency, and defining the closed loop bandwidth as the frequency at which this ratio is unity (i.e. above this frequency the system makes no correction). In general, this frequency is lower than the one given by the −3 dB definition.

12.7 Low Cost AO Systems

It is possible to build a low cost adaptive optics system from commercially available components, as demonstrated in [10,11]. Depending upon the requirements, a reasonable system based around a 37-actuator membrane mirror [27] costs in the region of \$10K – \$20K in parts. Given this choice of deformable mirror, and assuming (following an analysis based on (12.6)) that this mirror has sufficient degrees-of-freedom to provide useful improvements (i.e. that its fitting error variance is less than approximately 1 rad^2), there are (at least) three further decisions to be made:

- **Type of Wavefront Sensor** A curvature sensor is best matched to the membrane mirror. However, both the systems referred to above used a Shack–Hartmann sensor, probably resulting in sub-optimal performance (see Fig. 12.10).

- **Speed of Operation** A system based on a video-rate detector in the wavefront sensor will be much less expensive to build than one based on a high speed CCD such as the Dalsa camera used in [11], which had a frame rate of 780 fps. The choice of frame grabber is important, as this could introduce additional delay into the control loop.
- **Choice of Control Computer** A standard PC (Pentium or Power PC chip) will easily handle the control of a 37-actuator membrane mirror and a well-matched wavefront sensor for frame rates up to 1 kHz. A choice has to be on the operating system (e.g. Linux or Windows). Alternatively, a compact design might use an embedded processor using a DSP chip. We have recently taken delivery of a custom-designed CMOS detector and control computer package based on SHARC DSPs that is very compact.

For testing the performance of an AO system we have used a wavefront generator that uses a ferro-electric liquid crystal device to create dynamic binary holograms of any desired phase aberration [37]. This is used as shown in Fig. 12.8. The virtue of using a wavefront generator is that the spatial and temporal performance of the AO system can be studied quantitatively. Figures 12.9 and 12.10 show typical results from the first system built in our laboratory.

In Fig. 12.9 the open loop frame rate was 270 Hz, $D/r_0 \approx 7.5$ and $v/r_0 \approx 5$ Hz. The maximum Strehl ratio was approximately 0.4 in this case, compared to a value for the reference arm of approximately 0.8: we have not carried out detailed modeling of our system, but this Strehl is of the same order-of-magnitude that we would expect for this value of D/r_0 and a 37-actuator membrane mirror.

Figure 12.10 provides a more quantitative description of the performance of the system. The left hand side shows the effect of discarding higher order modes, for a value of D/r_0 of 7.5. We typically find that it is optimum to retain 20-25 modes in this system. The right hand side shows the temporal behaviour for an 780 Hz frame rate and $D/r_0 \approx 5$ (note the higher Strehl value at the origin). There is still a good

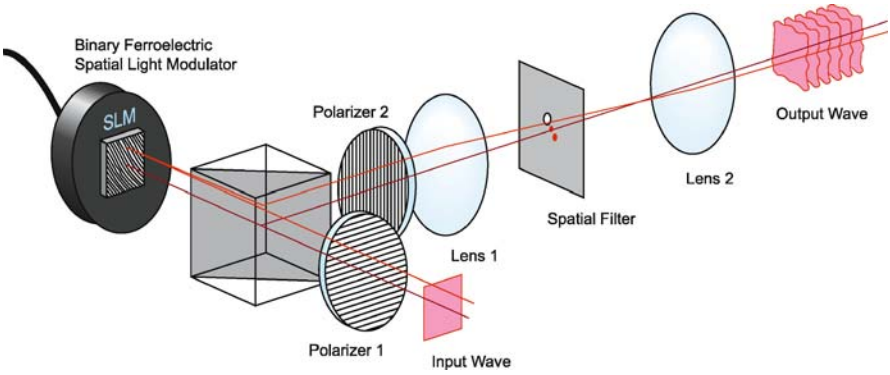


Fig. 12.8. A wavefront generator for testing AO systems [37]

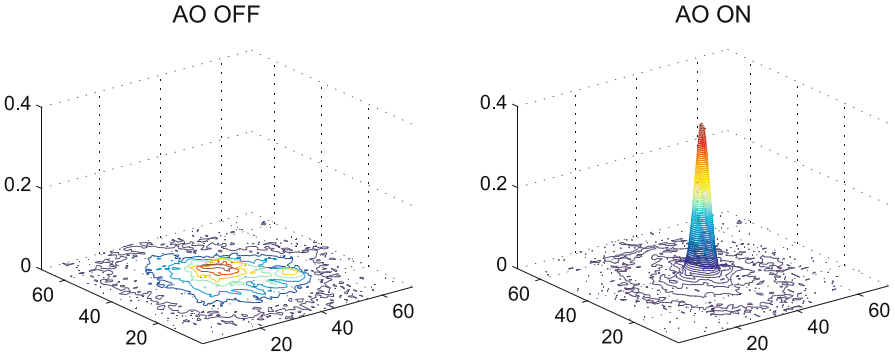


Fig. 12.9. The time-averaged point spread function. *Left*: AO system off. *Right*: AO system on (see text for parameters)

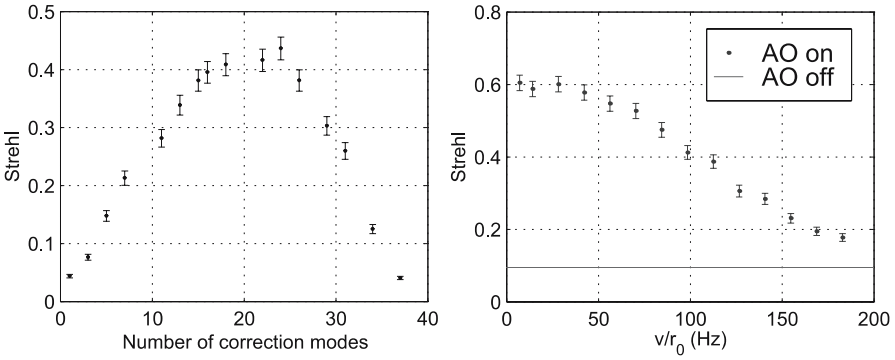


Fig. 12.10. *Left*: Strehl ratio as a function of the number of modes used in the control loop. *Right*: Strehl ratio as a function of the "wind speed" (see text for parameters)

Strehl ratio at value of $v/r_0 \approx 100\,\mathrm{Hz}$, corresponding to a wind speed of $50\mathrm{ms}^{-1}$ for a value of $r_0 \approx 0.5\,\mathrm{m}$ (typical of a good observing site in the near IR).

12.8 Current Research Issues in Astronomical Adaptive Optics

Adaptive optics systems that operate in the near infrared are now installed on all the World's major groundbased optical/IR telescopes. The reason that these first systems all operate in the near IR is threefold:

- As shown in (12.2), the Fried parameter r_0 is proportional to $\lambda^{6/5}$. This is means that the lenslet area in a Shack–Hartmann sensor, and hence the flux collected by a lenslet, is proportional to $\lambda^{2.4}$: this in turn means that fainter guide stars can be used in the IR.

- The correlation time of the atmospherically induced phase pertubation is proportional to $\lambda^{6/5}$: thus longer sample times, and fainter guide stars, can be used in the IR.
- The isoplanatic angle is also proportional to $\lambda^{6/5}$, giving an increased area (proportional to $\lambda^{2.4}$) over which a guide star can be found. The isoplanatic angle (see the left hand side of Fig. 12.11) is the angle over which the variance of the wavefront aberration due to differing angles of incidence varies by approximately 1 rad^2. The isoplanatic angle is strongly influenced by the $C_N^2(z)$ profile of the refractive index structure function "constant".

The first two factors combine to make the required brightness of the guide star proportional to $\lambda^{3.6}$, so that, for a AO system to operate at 550 nm, the guide star has to be approximately $4^{3.6} = 150$ times brighter than at 2.2 μm. Futhermore, because of the third factor above, the region of the sky over which this much brighter guide star has to be found is approximately 28 times smaller in area.

This question of the required guide star brightness at different wavelengths translates to one of sky coverage – the fraction of the sky over which it is possible to find a bright enough natural guide star within the isoplanatic patch. For observation in the 1–2 μm region, the sky coverage, with an extremely high efficiency, low noise detector in the wavefront sensor, is on the order of 1% – 10%, an acceptable value. However, in the visible, the sky coverage is so low that only "targets of opportunity" can be observed using an AO system.

A related problem is the actual value of the isoplanatic angle: this is only a few arc-seconds in the visible (and is proportional to $\lambda^{6/5}$), and thus the field of view of an AO system is so small as to be of little interest to astronomers as a general purpose enhancement of image quality.

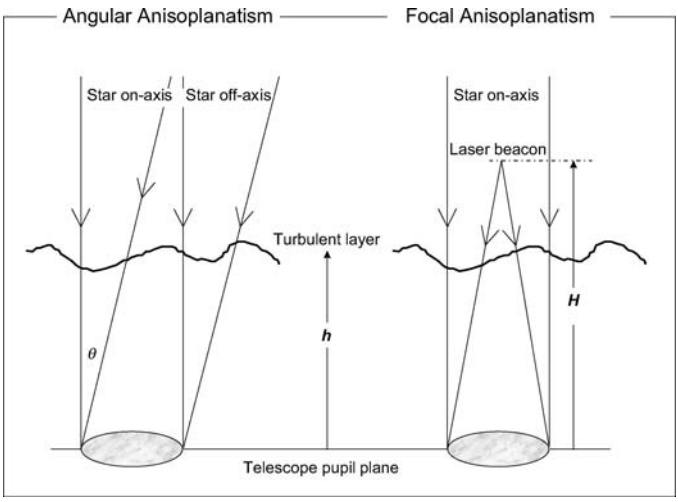


Fig. 12.11. *Left*: Angular anisoplanatism. *Right*: Focus anisoplanatism

The solution is to make an artificial point source – a laser guide star – either by Rayleigh scattering or by resonantly exciting atomic sodium in the mesosphere. In each case, it is technically feasible to make a bright enough source but a new problem is now introduced, that of focus anisoplanatism. This is illustrated on the right hand side of Fig. 12.11: the wavefront sensed by the laser guide star is not the same as that experienced by the light from the astronomical object under observation. Since the mesospheric sodium guide star is at a much higher altitude (≈ 85 km) than the Rayleigh guide star (up to ≈ 20 km), the former minimises this effect. Unfortunately, a high power (10–20 W) laser precisely tuned to the sodium D_2 line is rather complex and expensive.

In the visible, focus anisoplanatism severely reduces the Strehl ratio for a 8 m diameter telescope and must be overcome. By using several laser guide stars, one can reduce the error due to focus anisoplanatism considerably, and, as a bonus, slightly increase the isoplanatic angle. Several laser guide stars means several wavefront sensors, and thus the system complexity is greatly increased. By further adding additional deformable mirrors, each conjugate to a different height in the atmosphere, i.e. doing so-called *multi-conjugate adaptive optics*, MCAO, one can increase the field of view significantly.

At present, MCAO (using either laser or natural guide stars) is the key research area in astronomical adaptive optics. Two approaches are being explored, one in which there is one wavefront sensor for each guide star ("conventional" MCAO) and the other in which all the guide stars contribute to each sensor, which is conjugated to a certain height above the telescope (the "layer-oriented" approach). In both cases, the AO system will be very complex.

Looking beyond the time when MCAO is implemented in astronomy, there will be the need to make MCAO work in low cost applications, for example, AO-assisted binoculars. Clearly, the complexity of the astronomical systems will have to be eliminated: there is still much to invent in this area before it becomes a reality.

12.9 Adaptive Optics and the Eye

The concepts of wavefront control, wavefront sensing and adaptive optics have many applications apart from astronomy, although in order to justify the use of AO the costs have to be several orders of magnitude smaller. Some of these potential applications are:

- *Line of sight optical communication* Building to building line of sight optical communication is limited in part by atmospherically induced intensity fluctuations (scintillation). AO may offer a means of reducing this and hence improving the reliability and bandwidth of the link.
- *Confocal microscopy* In confocal microscopy, the scanning probe is degraded as it propagates into the depth of an object, and also spherical aberration is induced. AO may offer a means of reducing this effect and thus improving the transverse resolution, the depth discrimination and signal-to-noise ratio of the image. Since

optical data storage may use confocal imaging, there are potential applications in his field as well.

- *High power lasers* Thermally induced effects in high power lasers reduce the beam quality and range of operating conditions. Intracavity AO may offer improved beam quality and a greater range of operating conditions.
- *Beam shaping* In lasers of all powers, AO may allow minor beam imperfections to be cleaned up and permit the dynamic shaping of beams, for example in laser materials processing.
- *Scanned imaging systems* AO may be usede to maintain the integity of a scanning probe, for example, in a laser printing system.

Another particularly promising area of application of adaptive optics and its technologies is in the eye. At the simplest level, wavefront sensing can be used to determine the exact refractive state of the eye, and several commercial wavefront sensors are becoming available for use in connection with laser refractive surgery (Lasik, PRK). Closed loop adaptive optics has two main areas of application in the eye. The first is for retinal imaging, and the second is to enhance vision.

When focus and astigmatism errors are well-corrected, the eye is more-or-less diffraction limited for a pupil diameter of approximately 2 mm. With this pupil size, the point spread function is $\approx 5\,\mu$m is diameter, compared to the cone photoreceptor spacing at the fovea on the order of $2\,\mu$m: this spacing is consistent with the Shannon or Nyquist sampling theorem. In order to image the cones at the fovea, one would need to dilate the pupil and then ensure that the aberrations over the pupil, which are typically severe for a dilated pupil, are corrected using adaptive optics. When the aberrations are corrected, we should have diffraction-limited retinal imaging, and also we may have enhanced visual acuity: indeed there are a number of interesting psychophysical experiments that can be carried out on AO-corrected eyes.

Initial studies in AO-assisted retinal imaging used open-loop adaptive optics, and combining AO with the technique of retinal densitometry, Roorda and Williams [38] were able to obtain remarkable images showing the spatial arrangement of the short, medium and long wavelength sensitive cones about 1 deg from the fovea, reproduced in Fig. 12.12. The temporal variation of the wavefront aberration is much slower in the eye than for atmospheric turbulence [39] and so that aspect of building a closed loop system is relatively straightforward. However, alignment and other issues due to the fact that a human subject is involved make adaptive optics in the eye quite difficult to achieve. Recently, three groups have published results showing closed-loop AO systems operating in the eye [40–42].

Finally, two groups [43,44] have implemented "poor man's AO", using wavefront data combined with direct retinal imagery in a deconvolution scheme [45].

Adaptive optics for retinal imaging and vision enhancement is on the threshold of a number of breakthroughs and improvements that can be expected to have a significant impact in clinical vision science.

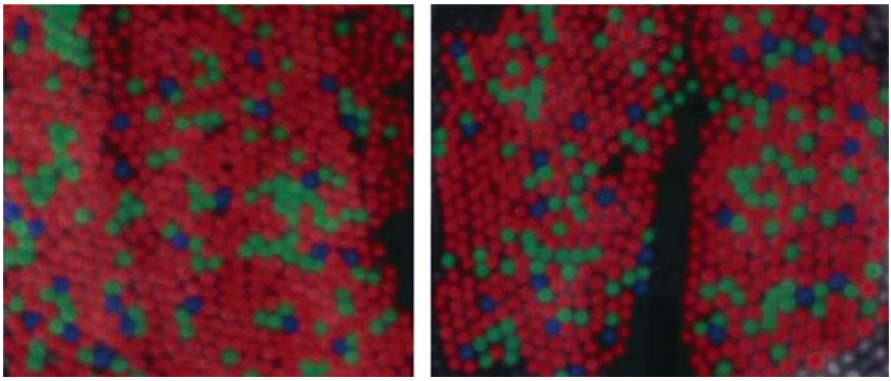


Fig. 12.12. Spatial arrangement of short (S), medium (M) and long (L) wavelength sensitive cones using retinal densitometry combined with open loop adaptive optics [38]. *Left*: 1 deg temporal. *Right*: 1 deg nasal. Courtesy of Austin Roorda and David Williams

Acknowledgements

I am grateful to my present and former students and post-docs for increasing my understanding of adaptive optics, particularly to Dr Carl Paterson who gave me a basic understanding of some aspects of control systems and whose lecture notes form the basis of part of Sec. 12.5 and all of Sec. 12.6, and to Dr Luis Diaz and Ian Munro who built our first closed loop AO system for the eye.

Our research on adaptive optics is funded by the European Union, the UK Engineering and Science Research Council, the UK Particle Physics and Astronomy Research Council, The Wellcome Trust, and several industrial partners, and I am grateful for their support.

References

1. H.W. Babcock: Pub. Astron. Soc. Pac. **65**, 229 (1953)
2. D.L. Fried: J. Opt. Soc. Am. **56**, 1372 (1966)
3. M. Born, E. Wolf: *Principles of Optics* (Cambridge University Press, Cambridge 1999)
4. R.J. Noll: J. Opt. Soc. Am. **66**, 207 (1976)
5. V.I. Tatarski: *Wave Propagation in a Turbulent Medium* (Dover Press, New York 1967)
6. F. Roddier: 'Effects of Atmospheric Turbulence in Astronomy', In: *Progress in Optics*, ed. by E. Wolf (Elsevier, Amsterdam 1981) pp. 281–376
7. J.W. Hardy: *Adaptive Optics for Astonomical Telescopes* (Oxford University Press, New York 1998)
8. G. Rousset et al.: Astron. Astrophys. **230**, L29 (1990)
9. B. Brandl et al.: Astrophys. J. **466**, 254 (1996)
10. D. Dayton et al.: Opt. Commun. **176**, 339 (2000)
11. C. Paterson, I. Munro, J.C. Dainty: Opt. Express **6**, 175 (2000)
12. J.R.P. Angel: Nature **368**, 203 (1)994

13. C.L. Koliopoulis: Appl. Optics **19**, 1523 (1980)
14. R.V. Shack, B.C. Platt: J. Opt. Soc. Am. **46**, 656 (1971) (abstract only)
15. R. Ragazzoni: J. Mod. Optics **43**, 289 (1996)
16. F. Roddier: Appl. Optics **29**, 1402 (1990)
17. R.A. Gonsalves: Opt. Eng. **21**, 829 (1982)
18. P.M. Blanchard et al.: Appl. Optics **39**, 6649 (2000)
19. M.A.A. Neil, M.J. Booth, T. Wilson: J. Opt. Soc. Am. A **17**, 1098 (2000)
20. E.N. Ribak, S.M. Ebstein: Opt. Express **9**, 152 (2001)
21. C. Paterson, J.C. Dainty: Opt. Lett. **25**, 1687 (2000)
22. R. Cubalchini: J. Opt. Soc. Am. **69**, 972 (1979)
23. G. Rousset: 'Wavefront Sensing', In: *Adaptive Optics for Astronomy*, ed. by D.M. Alloin, J.-M. Mariotti (Kluwer Academic Publishers, Dordrecht 1994) pp. 115–137
24. See `http://www.xinetics.com/` and `http://www.cilas.com/englais3/index`
25. C. Schwartz, E. Ribak, S.G. Lipson: J. Opt. Soc. Am. **11**, 895 (1994)
26. A.V. Kudryashov, V.I. Shmalhausen: Opt. Eng. **35**, 3064 (1996)
27. G. Vdovin, P.M. Sarro, S. Middelhoek: J. Micromech. & Microeng. **9**, R8 (1999) (See also `http://www.okotech.com/`)
28. M. Horenstein et al.: J. Electrostatics **46**, 91 (1999)
29. T. Weyrauch et al.: Appl. Optics **40**, 4243 (2001)
30. See `http://www.meadowlark.com/index.htm`
31. T.L. Kelly, G.D. Love: Appl. Optics **38**, 1986 (1999)
32. S.R. Restaino et al.: Opt. Express **6**, 2 (2000)
33. M.A. Helmbrecht et al.: 'Micro-mirrors for adaptive optics arrays', In: *Transducers*, (Proc. 11th International conference on Solid State Sensors and Actuators, Munich, Germany 2001) pp. 1290–1293
34. E.P. Wallner: J. Opt. Soc. Am. **73**, 1771 (1983)
35. M.C. Roggemann, B. Welsh: *Imaging through Turbulence* (CRC Press, Boca Raton, Florida 1996)
36. M. Demerlé, P.Y. Madec, G. Rousset: 'Servo-loop analysis for adaptive optics', In: *Adaptive Optics for Astronomy*, ed. by D.M. Alloin, J.–M. Mariotti (Kluwer Academic Publishers, Dordrecht 1994) pp. 73–88
37. M.A.A. Neil, M.J. Booth, T. Wilson: Opt. Lett. **23**, 1849 (1998)
38. A. Roorda, D.R. Williams: Nature **397**, 520 (1999)
39. H. Hofer et al.: J. Opt. Soc. Am. A **18**, 497 (2001)
40. E.J. Fernández, I Iglesias, P Artal: Optics Lett. **26**, 746 (2001)
41. H. Hofer et al.: Opt. Express **18**, 631 (2001)
42. J.F. Le Gargasson, M. Glanc, P. Lena: C. R. Acad. Sci. Paris **t. 2 Série IV**, 1131 (2001)
43. I. Iglesias, P. Artal: Optics Lett. **25**, 1804 (2000)
44. D.P. Catlin: *High Resolution Imaging of the Human Retina* (PhD Thesis, Imperial College, London 2001)
45. J. Primot, G. Rousset, J.C. Fontanella: J. Opt. Soc. Am. A **7**, 1598 (1990)

13 Low-Coherence Interference Microscopy

C.J.R. Sheppard and M. Roy

13.1 Introduction

Confocal microscopy is a powerful technique that permits three-dimensional (3D) imaging of thick objects. Recently, however, two new techniques have been introduced, which are rivals for 3D imaging in reflection mode. Coherence probe microscopy (CPM) is used for surface inspection and profiling, particularly in the semiconductor device industry [1–10]. Optical coherence tomography (OCT) is used for medical diagnostics, particularly in ophthalmology and dermatology [11,12]. These are both types of low coherence interferometry (LCI), in which coherence gating is used for optical sectioning. LCI has many advantages over the conventional (narrow-band source) interferometric techniques, including the ability to reject strongly light that has undergone scattering outside of a small sample volume, thus allowing precise non-invasive optical probing of dense tissue and other turbid media. LCI can be used to investigate deep, narrow structures. This feature is particularly useful for in-vivo measurement of deep tissue, for example, in transpupilliary imaging of the posterior eye and endoscoping imaging. Another application area of LCI is in optical fibre sensors [13–15].

Although CPM and OCT are both forms of LCI, their implementations differ. In CPM, an interferometer is constructed using a tungsten-halogen or arc lamp as source, Fig. 13.1. Interference takes place only from the section of the sample located within the coherence length relative to the reference beam. This technique is usually used for profilometry of surfaces. A series of 2D images is recorded using a CCD camera as the sample is scanned axially through focus. For each pixel an interferogram is recorded, which can be processed to extract the location of the surface either from the peak of the fringe visibility, or from the phase of the fringes. Use of a broad-band source avoids phase-wrapping ambiguities present in laser interferometry with samples exhibiting height changes of greater than half a wavelength. If a microscope objective of high numerical aperture is employed in CPM, interference occurs only for regions of the sample which gives rise to a wave front whose curvature matches that of the reference beam [16,17]. For this reason CPM has been termed phase correlation microscopy[4]. An optical sectioning effect results which is analogous to that in confocal microscopy. This phenomenon been used by Fujii [18] to construct a lensless microscope, and by Sawatari [19] to construct a laser interference microscope. Overall optical sectioning and signal collection performance [20] is slightly superior to that of a confocal microscope employing a pinhole. In or-

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

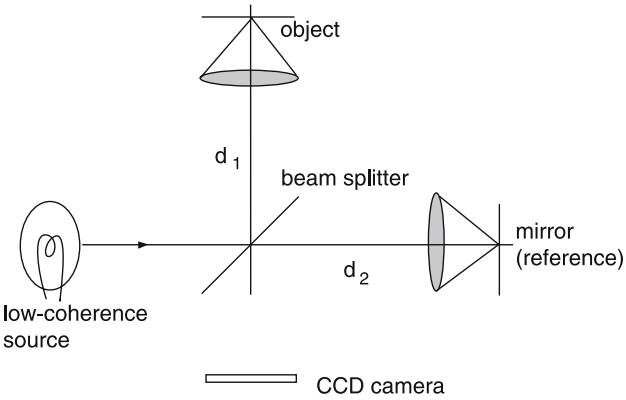


Fig. 13.1. The principle of the coherence probe microscope (CPM)

der to extract the height information from the interferogram, a complete axial scan through the sample must be recorded. An alternative method is based on phase shifting, in which the visibility at a particular axial position can be determined by three or more measurements for different values of reference beam phase. This avoids the necessity to scan through the complete depth of the image, and is particularly advantageous with thick structures. Phase shifting is traditionally performed with laser illumination, and with a broad-band source a problem is encountered in that the phase shift of the different spectral components, when reflected from a mirror, varies. This problem has been overcome by performing the phase shifting by use of geometric phase shifting (GPS) [21,22], in which polarisation components are used to effect an achromatic phase shift [23].

Although the principle of OCT is identical to that of CPM, the embodiment is usually quite different [11]. In OCT the signal, scattered from bulk tissue, is very weak. Hence the source is often replaced by a super-luminescent laser diode, which is very bright but exhibits a smaller spectral bandwidth than a white-light source. The source is used in a scanning geometry, in which the sample is illuminated point-by-point, Fig. 13.2, so that considerable time is necessary to build up a full 3-D image. Therefore, usually x-z cross-sections are recorded. In OCT the numerical aperture (NA) of the objective lens is small and its depth of focus large, so z scanning can be achieved by translation of the reference beam mirror. Instead of phase-shifting, the alternative techniques of heterodyning is used, which can result in shot-noise limited detection performance. In OCT optical sectioning results from the limited coherence length of the incident radiation, but if the NA of the objective lens is large, an additional optical sectioning effect similar to that in confocal imaging results [24]. The resulting instrument is called the optical coherence microscope (OCM). In OCM, because the depth of focus of the objective is small, the sample has to be scanned relative to the objective lens to image different z positions, so scanning is usually performed in x-y sections.

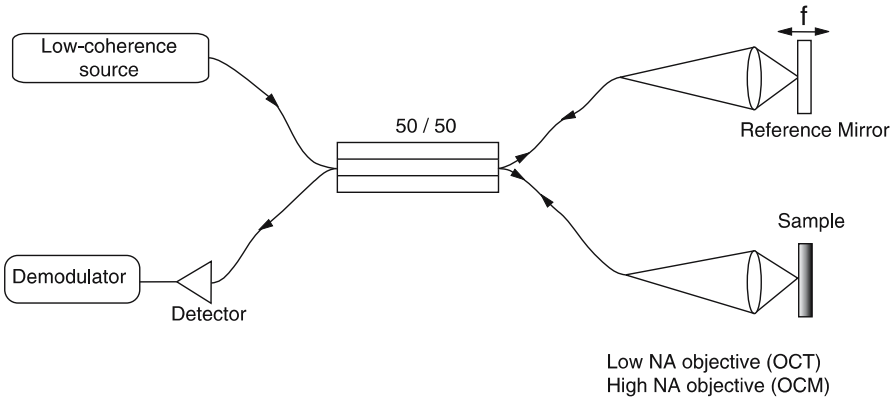


Fig. 13.2. The principle of optical coherence tomography (OCT)

Both OCT and OCM are usually implemented using single mode optical fibres for illumination and collection. A scanning microscope using single mode fibres behaves as a confocal microscope [25–27], and the geometry of the fibre spot influences the imaging performance of the system [28–32]. Confocal interference microscopes using monochromatic light and a pinhole in front of the detector were constructed long before the proposal of OCT [33–35]. Similarly single-mode optical fibre implementations of the confocal interference microscope with a monochromatic source have also been reported [36–38]. In these systems, a confocal optical sectioning effect is present. This optical sectioning behaviour is similar to, but exhibits some differences from, the correlation effect in the CPM. Either low-coherence light or ultra short pulses [39] can be used to measure internal structure in biological specimens. An optical signal that is transmitted through or reflected from a biological tissue contains time-of flight information, which in turn yields spatial information about tissue microstructure. In contrast to time domain techniques, LCI can be performed with continuous-wave light, avoiding an expensive laser and cumbersome systems. Summarizing, we distinguish between:

(a) CPM uses a medium/high NA, giving sectioning from a combination of a coherence and a correlation effect. Complete *x*-*y* images are recorded using a CCD detector.
(b) OCT uses a low NA, giving sectioning from coherence only. The low NA is used in OCT to provide a long depth of focus that enables $x - z$ cross-sectional imaging using coherence gate scanning.
(c) OCM uses a high NA, giving sectioning from a combination of a coherence and a confocal effect. The high NA is used in OCM to create a small focal volume in the sample that combines with the coherence gate, resulting in high resolution with rejection of out-of-focus or multiply scattered light.

13.2 Geometry of the Interference Microscope

There are several interference microscope available. Fig. 13.3 shows the Linnik, Mirau, Michelson and confocal interferometers, each with various advantages and disadvantages. For high magnification it is often advantageous to use a Linnik interference microscope [1], in which two matched microscope objectives with high NA are used. Using matched objectives, spherical and chromatic aberration can be removed. The main disadvantage of this type of microscope is that it is very sensitive to vibration and air currents, because it involves a long beam path. However, these vibrations are avoided in OCT by employment of heterodyning techniques that reject low frequency variations in signal.

Mirau [4] and Michelson interference microscopes require only one microscope objective. The difference between these two types is the positioning of the beam splitter and the reference surface. In the Mirau system, the beam splitter and reference mirror are positioned between the objective and test surface, whereas in the Michelson type, a long working distance microscope objective is used to accommodate the beam splitter between the objective and the test surface. These two methods, particularly the Michelson geometry, are aperture limited compared to the Linnik and confocal systems and therefore they are used for low magnifications.

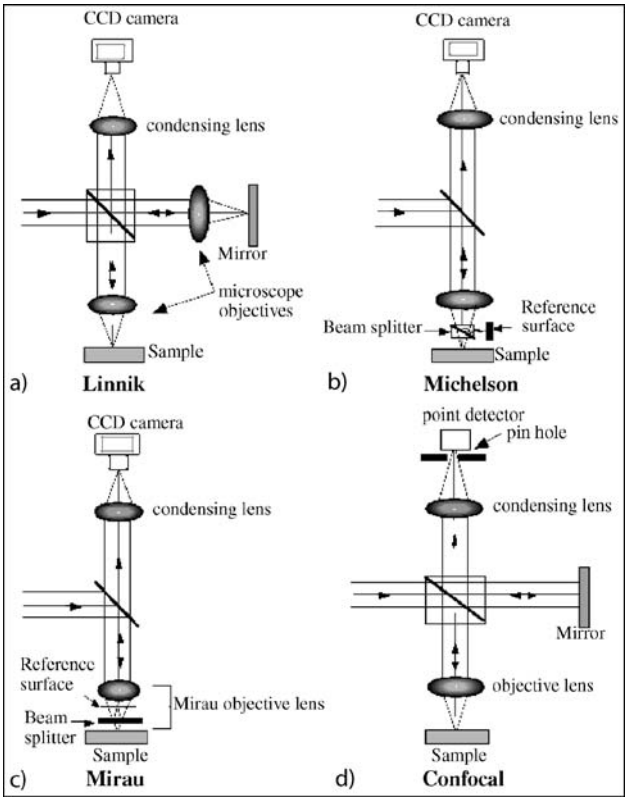


Fig. 13.3. Different types of interference microscope

On the other hand the concept of the confocal interference microscope [33] is quite different from the other methods, as shown in Fig. 13.3. This type is based on a point-by-point detection technique which requires a pinhole in front of a detector and because of this the wave-front distortion of the reference beam can be ignored. In practice of course the pinhole must have some finite size, and the wave front of the reference beam over the pinhole area does affect the overall performance. The pinhole can also be replaced by a single-mode fibre [36,38]. In this case the fibre acts as a coherent detector, so that only the component of the object and reference beams that match the mode profile are detected. In this case, aberrations on the reference beam only affect the strength of the reference beam.

13.3 Principle of Low-Coherence Interferometry

A schematic diagram of CPM is shown in Fig. 13.1, in which white, incoherent light is used as illumination. The sample is placed in one interferometer arm, and other arm provides a reference beam. The reflections from the sample are combined with the reflection from the mirror. The electric field of the light arriving at the camera, $U(t)$, is the superposition of the light traversing the two arms of the interferometer represented by

$$U(t) = U_s(t) + U_r(t+\tau) \,, \tag{13.1}$$

where U_s and U_r are the optical amplitude of signal and reference beam. The quantity τ is the time delay due to the difference in the length of the optical paths traversed by the beams. Normally any detector system for optical frequencies actually measures the time averaged intensity, which is the square of the electric field amplitude U, and expressed by

$$I_d(\tau) = \langle |U|^2 \rangle = \left\langle \, [U_s^*(t) + U_r(t+\tau)][U_s^*(t) + U_r^*(t+\tau)] \, \right\rangle \,, \tag{13.2}$$

where $\langle\rangle$ denotes the time average. Thus

$$I_d(\tau) = I_s + I_r + 2(I_s I_r)^{1/2} \Re\{\gamma(\tau)\} \,, \tag{13.3}$$

where $\gamma(\tau)$ is the normalized form of the mutual coherence function, the complex degree of coherence, that can be represented by

$$\gamma(\tau) = \frac{\langle U_s(\tau) U_r(t+\tau) \rangle}{(I_s + I_r)^{1/2}} \,. \tag{13.4}$$

Equation (13.3) is the generalized interference law for partially coherent light.

In general, the complex degree of coherence includes both temporal and spatial coherence effects. For a Michelson interferometer which is an amplitude splitting interferometer, for a single point on the spatially incoherent source, spatial coherence (e.g. as in the Young interferometer) can be neglected, so that the mutual coherence reduces to self coherence, or temporal coherence, which has the same behaviour for

CPM and OCT. When the complete incoherent source is considered, partial spatial coherence must also be considered, and then CPM and OCT behave differently. Confining our attention to temporal coherence for the present, (13.3) can be written

$$I_d(\tau) = I_s + I_r + 2(I_s I_r)^{1/2} \Re\{\gamma_{tc}(\tau)\} , \tag{13.5}$$

where $\Re\{\gamma_{tc}(\tau)\}$ is the real part of the complex degree of temporal coherence of the light source which is the normalized form of the self-coherence function $\gamma_{11}(\tau)$,

$$\gamma_{11}(\tau) = \frac{\Gamma_{11}(\tau)}{\Gamma_{11}(0)} , \tag{13.6}$$

where

$$\Gamma_{11}(\tau) = \langle U(t+\tau) U^*(\tau) \rangle . \tag{13.7}$$

Here the subscript 11 corresponds to a single point on the incoherent source. The normalized complex degree of coherence of the light source is given by the Fourier transform of the power spectrum of the light source.

Now for a polychromatic light source with a Gaussian power spectral density

$$G(f) = G_0 \exp\left[-\left(\frac{f - \bar{f}}{\Delta f} \right)^2 \right] , \tag{13.8}$$

where G_0 is constant, Δf represents the spectral width of the source, and $\bar{f}$ is the mean frequency. Thus

$$\gamma_{11}(\tau) = G_0 \int_{-\infty}^{\infty} \exp\left[-\left(\frac{f - \bar{f}}{\Delta f} \right)^2 \right] \exp(-\mathrm{i}2\pi f \tau) df , \tag{13.9}$$

or

$$\gamma_{11}(\tau) = G_0 \pi \Delta f \exp\left[-(\pi \tau \Delta f)^2 \right] \exp(-i2\pi \bar{f} \tau) . \tag{13.10}$$

But we know that $\gamma_{11}(0) = 1$, so we have

$$\gamma_{tc}(\tau) = \exp\left[-(\pi \tau \Delta f)^2 \right] \exp(-\mathrm{i}2\pi \bar{f} \tau) . \tag{13.11}$$

Equation (13.3) can be therefore be written as

$$I(\tau) = I_s + I_r + 2(I_s I_r)^{1/2} \exp\left[-(\pi \tau \Delta f)^2 \right] \cos(2\pi \bar{f} \tau) , \tag{13.12}$$

or

$$I_d(\tau) = I_0 \left[1 + V(\tau) \cos(2\pi \bar{f} \tau) \right] , \tag{13.13}$$

where I_0 is the background intensity, V is the fringe contrast function or envelope of the observed fringe pattern, given by

$$V(\tau) = \frac{2(I_s I_r)^{1/2}}{I_s + I_r} \exp\left(-(\pi \tau \Delta f)^2 \right) . \tag{13.14}$$

If we consider white light with wavelength range from 400 nm to 700 nm, the coherence time is then roughly 6.6 fs and hence the spectral bandwidth is about 1.5×10^{14} radians s^{-1}, giving a coherence length of about 1–2 μm. In OCT, the coherence time of the source is roughly 50 fs, and the coherence length typically 10–15 μm.

13.4 Analysis of White-Light Interference Fringes

Figure 13.4 shows the variations in intensity at a given point in the image as an object consisting of a surface is scanned along the depth axis (z axis). We assume that the origin of coordinates is taken at the point on the z axis at which the two optical paths are equal, and that the test surface is moved along the z axis in a series of steps of size Δz. With a broad-band spatially incoherent source, the intensity at any point (x, y) in the image plane corresponding to a point on the object whose height is h can be written as

$$I_d(\tau) = I_s + I_r + 2(I_s I_r)^{1/2} \gamma\left(\frac{p}{c}\right) \cos\left[\left(\frac{2\pi}{\bar{\lambda}}\right) p + \phi_0\right] , \tag{13.15}$$

where I_s and I_r are the intensities of the two beams acting independently, $\gamma(p/c)$ is the complex degree of coherence (corresponding to visibility of the envelope of the interference fringes), and $\cos\left[(2\pi/\lambda)p + \phi_0\right]$ is a cosinusoidal modulation. In (13.15), $\bar{\lambda}$ corresponds to the mean wavelength of the source, $p = 2(z - h)$ is the difference in the lengths of the optical paths traversed by the two beams, and ϕ_0 is the difference in the phase shifts due to reflection at the beam-splitter and the mirrors.

Each of these interference patterns in z for each pixel (see Fig. 13.4) can be processed to obtain the envelope of the intensity variations (the visibility function). We can therefore determine the peak amplitude of the intensity variations and the location of this peak along the scanning axis (Fig. 13.5). The values of the peak amplitude give an autofocus image of the test object (analogous to the similar technique in confocal microscopy [40], while the values of the location of this peak along the scanning axis yield the height of the surface at the corresponding points [40].

A major objective of the LCI is to find the height of a surface from the location of the peak fringe visibility by using variety of algorithms that can be categorized in two main groups:

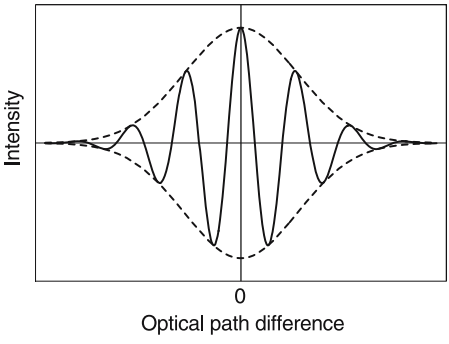


Fig. 13.4. The output from a white-light interferometer as a function of the position of one of the end mirrors along the z axis. The *dashed lines* represent the fringe envelope

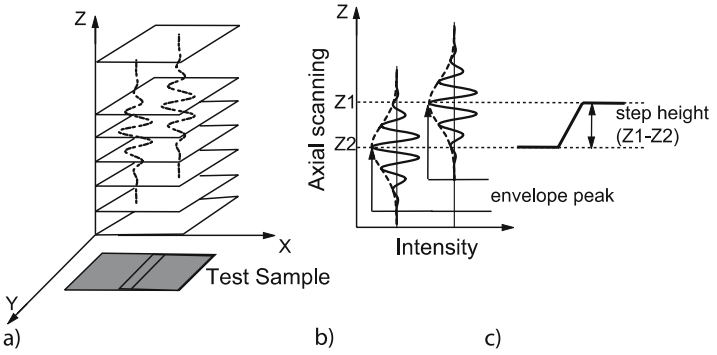


Fig. 13.5. A representation of different stages in white-light surface profiling: (**a**) A series of interferogram images is obtained by moving the sample along the z axis, (**b**) Processing of two sample pixels in the interferogram uses peak detection. (**c**) The location of the peak along the z axis corresponds to the height of the surface, allowing reconstruction of the surface profile

13.4.1 Digital Filtering Algorithms

The method most widely used to recover the fringe visibility function from the sampled data has been digital filtering in the frequency domain [4]. This procedure involves two discrete Fourier transforms (forward and inverse) along the z direction for each pixel in the sample. In order to recover the interference fringes the step size along the z axis, Δz often corresponds to a change in the optical path difference, p of a quarter of the shortest wavelength or less. Typically, a step Δz around 50 nm is used. Consequently this procedure requires a large amount of memory and processing time.

These requirements can be reduced to some extent by modified sampling and processing techniques. For example, according to the sampling theorem, the signal need be sampled only at a rate of twice the bandwidth, rather than for the highest frequency in the fringe pattern. This has been called sub-Nyquist sampling [7]. Other approaches include using a Hilbert transform [5], communication theory [41], or nonlinear filter theory [42].

13.4.2 Phase Shift Algorithms

A more direct approach, which is computationally much less intensive, involves shifting the phase of the reference wave by three or more known amounts for each position along the z axis and recording the corresponding values of the intensity. These intensity values can then be used to evaluate the fringe contrast directly at that position [6].

However, the usual method for introducing these phase shifts has been by moving the reference mirror in the interferometer along the axis by means of a piezoelectric translator (PZT) so as to change the optical path difference between the

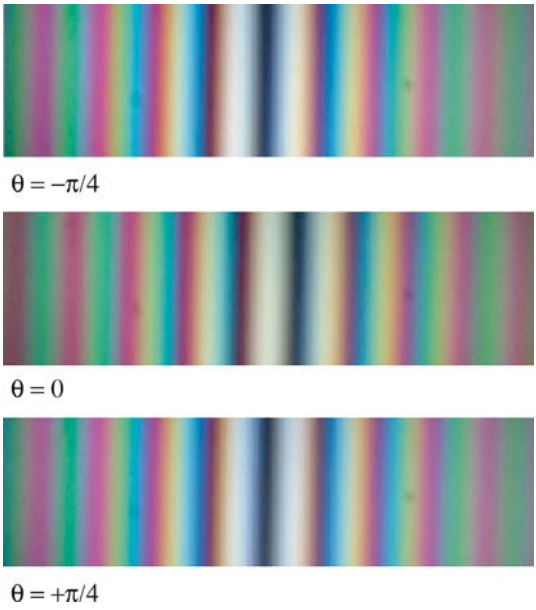


Fig. 13.6. White-light fringes for different angular settings of the geometric phase shifter, showing that the fringes move relative a stationary envelope

beams. This procedure has the drawback that the dynamic phase introduced by a change in the optical path p varies inversely with the wavelength. In addition, since the value of p is changing, the maximum value of the envelope function (fringe visibility) $\gamma(p/c)$ is not the same at a given position of the object. Both these factors can lead to systematic errors in the values of fringe visibility obtained. These problems can be avoided by using a phase shifter that generates the required phase shifts without any change in the optical path difference. The only phase shifter that can satisfy this condition is a geometric phase shifter (GPS) [22], an achromatic phase-shifter operating on the principle of the geometric phase [23], which can be achieved by using a polarising beam splitter, a quarter wave plate and a rotating polariser [43]. Figure 13.6 shows how the white light fringes move relative to a stationary envelope as the geometric phase is altered. For a given point on the envelope, the fringe pattern changes without bound as the geometric phase is changed. Thus for this variation the bandwidth is effectively zero, and only three discrete measurements are needed to recover the amplitude and phase of the fringe. This is then repeated at different points on the envelope, but to recover the envelope the distance between these points is fixed by the bandwidth of the envelope rather than the fringes.

A variety of different phase shifting algorithms can be used to extract the envelope, based on measurements for three or more values of phase step. Five step algorithms appear to be good for contrast or visibility measurement [42,44]. Five measurements are made of the intensity as follows:

$$
\begin{aligned}
I_1 &= I_0[1 + V\cos(\phi - \pi)] = I_0[1 - V\cos\phi] \,, \\
I_2 &= I_0[1 + V\cos(\phi - \pi/2)] = I_0[1 + V\sin\phi] \,, \\
I_3 &= I_0[1 + V\cos\phi] \,, \\
I_4 &= I_0[1 + V\cos(\phi + \pi/2)] = I_0[1 - V\sin\phi] \,, \\
I_5 &= I_0[1 + V\cos(\phi - \pi)] = I_0[1 - V\cos\phi] \,.
\end{aligned} \tag{13.16}
$$

Since the additional phase differences introduced by the GPS are the same for all wavelengths, the visibility of the interference fringes at any given point in the field can be expressed as [42]

$$
V^2 = \frac{1}{4\cos^4\delta}\left[(I_2 - I_4)^2 - (I_1 - I_3)(I_3 - I_5)\right] \,, \tag{13.17}
$$

where δ is the error in the phase step. This is a very efficient algorithm, which is also error correcting, so that if the phase step is not $\pi/2$ it still gives the visibility correctly, but scaled by a factor that depends on the phase step. This is satisfactory for determination of the peak position, but as the phase step can change if the spectral distribution is altered on reflection it can result in errors in the autofocus image. However, the scaling factor can be determined from the relationship

$$
\sin\delta = \frac{I_5 - I_1}{2(I_4 - I_2)} \,, \tag{13.18}
$$

and the intensity of the autofocus image corrected accordingly.

13.5 Spatial Coherence Effects

In a conventional interference microscope, nearly coherent illumination, produced by stopping down the condenser aperture, is used in order to give strong interference fringes [45]. However, in an interference microscope with digital storage, the interference term of the image can be extracted from the non-interference background terms by digital processing. In this case the relative strength of the fringes is not so important. If the aperture of the condenser is opened up until it is equal to that of the objective, as in a CPM, the relative strength of the fringes decreases. At the same time, a correlation effect is also introduced: only light which has a phase front which matches that of the reference beam results in an appreciable interference signal [16,17]. The result is an optical sectioning effect additional to that from the low-coherence source. The visibility of the fringes $V(p/c)$ is given by the product of the envelopes resulting from the low-coherence and the correlation effects [1]. The strength of the optical sectioning that results from the correlation effect increases rapidly as the numerical aperture (NA) of the system is increased. Typically, the envelope from the low-coherence effect has a width of about 5 μm for a broad-band source, while that from the correlation effect can be about 1 μm for high NA lenses.

In OCT, the system behaves as a confocal system, which can be regarded as a special case of a scanning interference microscope [46]. Using an optical fibre implementation, the mode profile of the fibre selects, for both the object beam and the reference beam, the overlap integral of the beam profile.

13.6 Experimental Setup

A schematic diagram of our LCI microscope [47] is depicted in Fig. 13.7. The optical system can be split into two parts: the illumination system, and the interferometer.

13.6.1 The Illumination System

A tungsten halogen lamp (12 V, 100 W) is used as a source. To illuminate the object uniformly, a Köhler illumination system is used, consisting of lenses L1-L4 (Fig. 13.7) together with a microscope objective. The system incorporates both aperture and field stops. The aperture stop is placed in a plane where the light source is brought to a focus, as shown by thick lines in Fig. 13.8. So by adjusting the aperture stop, the angle of illumination, and therefore the effective NA of the condenser, can be controlled. The field stop is placed in a plane where a back-projected image of the sample is formed, shown by thin lines in Fig. 13.8. Adjusting the field stop limits the area of illumination at the object, thereby reducing the strength of scattered light.

This system allows separate control of both the illumination aperture stop and the field stop. Stopping down the illumination aperture allows the system to be operated as a conventional interference microscope, with high spatial coherence. A 3 mW He-Ne laser is also provided for alignment. Since the coherence length of the laser is much longer than that of the white-light source, it can be used for finding the interference fringes.

13.6.2 The Interferometer

The interferometer used in our system is based on the Linnik configuration (Fig. 13.3). The linearly polarised beam transmitted by the polariser is divided at the polarising beam splitter into two orthogonally polarised beams which are focused onto a reference mirror and a test surface by two identical infinity tube-length $40\times$ microscope objectives with numerical aperture 0.75. After reflection at reference mirror and test surface these beams return along their original paths to a second beam-splitter which sends them through a second polariser to the CCD array camera.

The phase difference between the beams is varied by a geometric phase-shifter (GPS) consisting of a quarter-wave plate (QWP) with its axis fixed at an azimuth of 45°, and a polariser which can be rotated by known amounts [21,22]. The rotation of the polariser is controlled by a computer via a stepper-motor. In this case, if the polariser is set with its axis at an angle θ to the axis of the QWP, the linearly polarised beam reflected from the reference mirror and the orthogonally polarised beam from the test surface acquire a geometric phase shift equal to 2θ. This phase difference is very nearly independent of the wavelength. Achromatic phase-shifting based on a rotating polariser has better performance than a rotating half-wave plate, placed

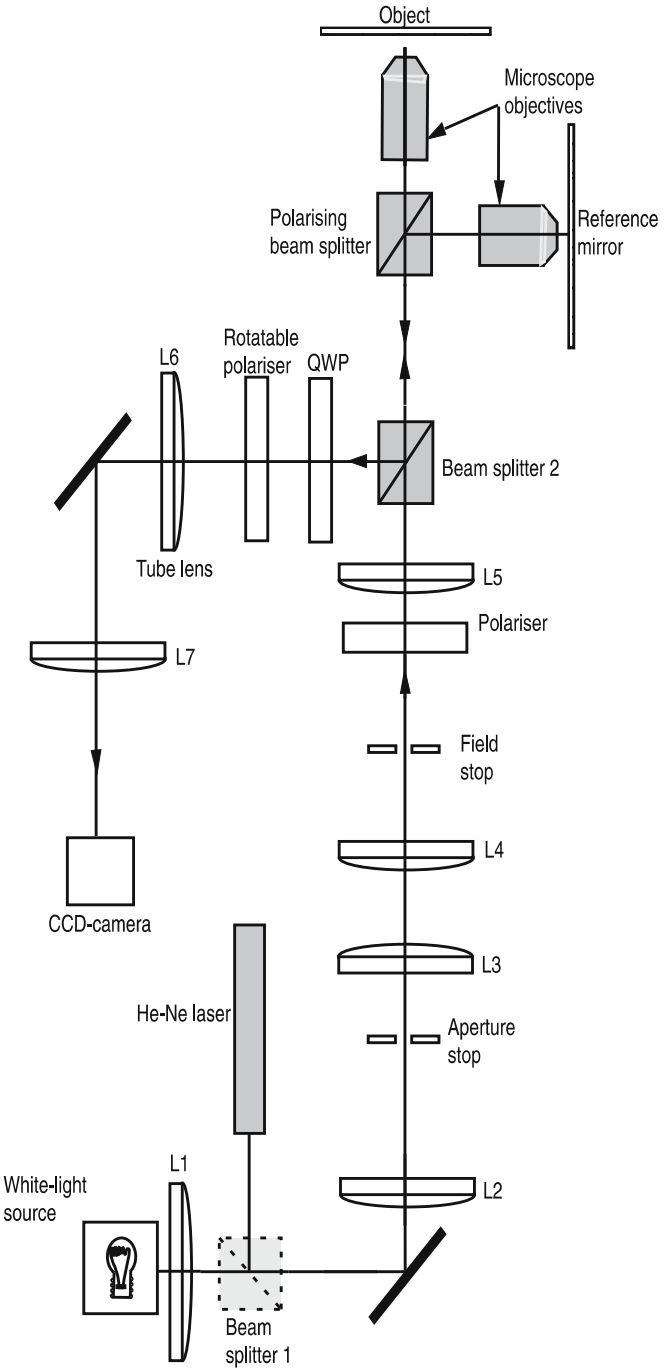


Fig. 13.7. Experimental arrangement of a white-light interference microscope using geometric phase shifter

between two quarter-wave plates, used in white-light interferometry to measure the fringe contrast function [21,43].

Measurements and data analysis were automated using a graphic code LabView program. The hardware was composed of a Macintosh Quadra 900 with a Neotech Frame Grabber and a digital-to-analog card. All the interference images obtained by the CCD video camera were transferred to the computer via a Neotech Frame Grabber which was used to convert the analogue video signal to digital images. The CCD camera has a pixel resolution of 768×512.

13.7 Experimental Results

A test object of an integrated circuit chip was scanned along the z axis by means of the PZT in steps of a magnitude $\Delta z = 0.66\,\mu\text{m}$ over a range of $5\,\mu\text{m}$, centred approximately on the zero-order white light fringe. In practice a single grabbed image taken at TV rates (1/25 seconds) exhibits significant noise, which can be reduced by averaging. Each grabbed image is thus in fact the average of N grabbed images, which reduces the random electronic noise of the capture and improving significantly the resolution of the system. In practice ten averages is a good compromise between speed and noise. Fig. 13.9 shows the effect of averaging frames on the random electronic noise.

A typical image of the white-light fringes is shown in Fig. 13.10. The surface height can be extracted by finding, for each pixel, the peak of the visibility variation. This can be done by various algorithms, such as evaluation of the first moment or a parabolic fit. We obtained good results by fitting a parabola through three points. Figures 13.11 and 13.12 show examples of the surface profile of a CCD chip obtained by our white-light interference microscope using a pair of 10×0.25 NA and 40×0.75 NA microscope objectives, respectively. Apart from the surface profile, we can simultaneously obtain an autofocus image, similar to a confocal autofocus image (maximum projection). In this, for every pixel, an image is formed from the maximum value of the visibility curve. This has the effect of bringing to focus all the different planes of the image, showing the reflectivity of the sample at any pixel. A surface profile of an integrated circuit, and the corresponding autofocus image, using a pair of 10×0.25 NA microscope objectives is shown in Fig. 13.13.

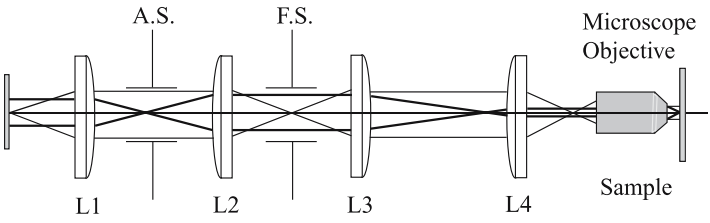


Fig. 13.8. The Köhler illumination system

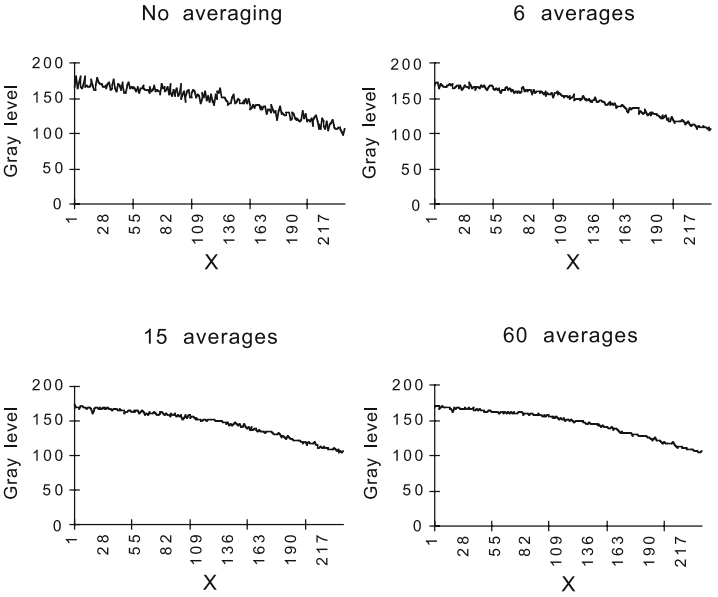


Fig. 13.9. Effect of averaging N images on the random noise

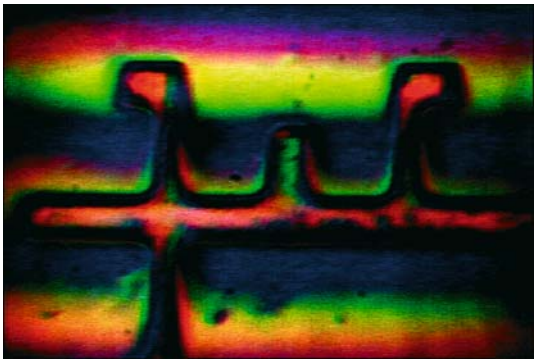


Fig. 13.10. A measured two-dimensional interference image

Fig. 13.11. A surface profile of an integrated circuit measured with our system using NA microscope objectives. The dimensions of the images are $170 \times 100\,\mu m^2$ with $1\,\mu m$ height

13.8 Discussion and Conclusion

We have demonstrated the successful use of a GPS that is close to achromatic in nature, for white-light interferometric surface profiling on a microscopic scale. Our

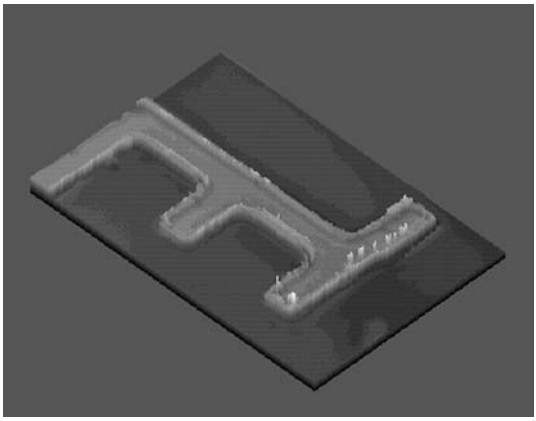


Fig. 13.12. A surface profile of an integrated circuit measured with our system using NA microscope objectives. The dimensions of the images are $25 \times 43\ \mu m^2$ with $1\ \mu m$ height

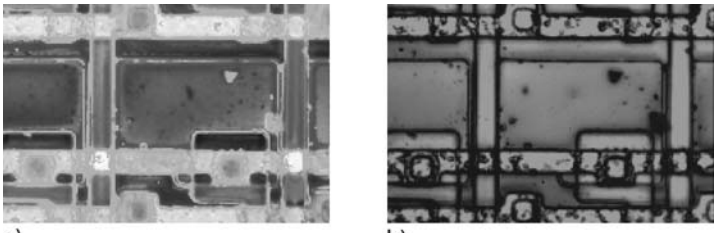


Fig. 13.13. (**a**) Surface profile, visualized as grey levels, of an integrated circuit measured with the system using $\times 10$, 0.25 NA microscope objectives. (**b**) The corresponding autofocus (peak projection) image. The dimensions of the image are $100g \times 70\ \mu m^2$ with $1\ \mu m$ height

system incorporate Köhler illumination, with control over illumination aperture and field diaphragms, and high numerical aperture microscope objectives. With high aperture optics an additional optical sectioning mechanism is provided by the spatial correlation effect. This property is not present in conventional laser interferometers because they use a low aperture condenser system. Thus separate control of the illumination (condenser) and imaging apertures allows us to investigate the properties of these different sectioning mechanisms. 3-D images of a variety of specimens including integrated circuit chips and thin film structures have been obtained by extracting the visibility envelope of the interferograms. The range of surface heights that can be profiled with this technique is limited only by the characteristics of the PZT used to translate the test specimen along the z axis and the available computer memory. However, since the steps between height settings at which data have to be taken can correspond to changes in the optical path difference of the order of a wavelength or more, a much smaller number of steps are required to cover a large range of depths.

If both fringe and envelope data is available, further information about the sample can be extracted, for example on the material properties of the sample. By knowing the fringe information, the phase of the signal can be recovered, the imaging

system behaving as a coherent imaging system. Alteration of the condenser aperture alters the properties of this coherent system, the use of matched condenser and imaging apertures resulting in three-dimensional imaging performance similar to that in a confocal reflection microscope. However, the low temporal coherence of the illumination results in improved rejection from out-of-focus regions and improved three-dimensional imaging of thick objects. This could result in important applications in the biomedical area, similar to the current uses of optical coherence tomography, with higher resolution albeit for reduced imaging depth.

References

1. M. Davidson, K. Kaufman, I. Mazor and F. Cohen: P. Soc. Photo–Opt. Inst. **775**, 233 (1987)
2. B. S. Lee and T. C. Strand: Appl. Optics **29**, 3784 (1990)
3. B. L. Danielson and C. Y. Boisrobert: Appl. Optics **30**, 2975 (1991)
4. S. S. C. Chim and G. S. Kino: Opt. Lett. **15**, 579 (1990)
5. S. S. C. Chim and G. S. Kino: Appl. Optics **31**, 2550 (1992)
6. T. Dresel, G. Häusler, and H. Venzke: Appl. Optics **31**, 919 (1992)
7. P. de Groot and L. Deck: Opt. Lett. **18**, 1462 (1993)
8. L. Deck and P. de Groot: Appl. Optics **33**, 7334 (1994)
9. P. Sandoz and G. Tribillon: J. Mod. Optics **40**, 1691 (1993)
10. P. Sandoz: J. Mod. Optics **43**, 1545 (1996)
11. D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M.R. Lee, T. Flotte, K. Gregory, C.A. Puliafito and Fujimoto, J. G.: Science **254**, 1178 (1991)
12. J. Schmitt, A. Knüttel and R. F. Bonner: Appl. Optics **32**, 6032 (1993)
13. S. Chen, A. W. Palmer, K. T. V. Grattan and B. T. Meggitt: Appl. Optics **31**, 6003 (1992)
14. Y. J. Rao, Y. N. Ning and D. A. Jackson: Opt. Lett. **18**, 462 (1993)
15. M. V. Plissi, A. L. Rogers, D. J. Brassington and M. G. F. Wilson: Appl. Optics **34**, 4735 (1995)
16. V. J. Corcoran: J. Appl. Phys. **36**, 1819 (1965)
17. A. E. Siegman: Appl. Optics **5**, 1588 (1966)
18. Y. Fujii, and H. Takimoto: Opt. Commun. **18**, 45 (1976)
19. T. Sawatari: Appl. Optics **12**, 2768 (1973)
20. M. Kempe and W. Rudolf: J. Opt. Soc. Am. A **13**, 46 (1996)
21. P. Hariharan: J. Mod. Optics **4**, 2061 (1993)
22. M. Roy and P. Hariharan: P. Soc. Photo–Opt. Inst. **2544**, 64 (1995)
23. M. V. Berry: J. Mod. Optics **34**, 1401 (1987)
24. J. A. Izatt, M. R. Hee, G. M. Owen, E. A. Swanson, J.G. Fujimoto: Opt. Lett. **19**, 590 (1994)
25. Y. Fujii and Y. Yamazaki: J. Microsc. **158**, 145 (1990)
26. L. Giniunas, R. Juskaitis and S. V. Shatalin: Electron. Lett. **27**, 724 (1991)
27. T. Dabbs and M. Glass: Appl. Optics **31**, 705 (1992)
28. M. Gu, C. J. R. Sheppard and X. Gan: J. Opt. Soc. Am. A **8**, 1755 (1991)
29. M. Gu and C. J. R. Sheppard: J. Mod. Optics **38**, 1621 (1991)
30. R. Gauderon and C. J. R. Sheppard: J. Mod. Optics **45**, 529 (1998)
31. M. D. Sharma and C. J. R. Sheppard: Bioimaging **6**, 98 (1998)
32. M. D. Sharma and C. J. R. Sheppard: J. Mod. Optics **46**, 605 (1999)

33. D. K. Hamilton and C. J. R. Sheppard: Opt. Acta **29**, 1573 (1982)
34. D. K. Hamilton and H. J. Matthews: Optik **71**, 31 (1985)
35. H. J. Matthews, D. K. Hamilton and C. J. R. Sheppard: Appl. Optics **25**, 2372 (1986)
36. M. Gu and C. J. R. Sheppard: Opt. Commun. **100**, 79 (1993)
37. M. Gu and C. J. R. Sheppard: Micron **24**, 557 (1993)
38. H. Zhou, M. Gu and C. J. R. Sheppard: Optik **103**, 45 (1996)
39. J. G. Fujimoto, S. De Silvestri, E. P. Ippen, C. A. Puliafito, et al.: Opt. Lett. **11**, 150 (1986)
40. I. J. Cox and C. J. R. Sheppard: Image and Vision Computing **1**, 52 (1983)
41. P. J. Caber: Appl. Optics **32**, 3438 (1993)
42. K. Larkin: J. Opt. Soc. Am. A **13**, 832 (1996)
43. H. S. Helen, M. P. Kothiyal and R. S. Sirohi: Opt. Commun. **154**, 249 (1998)
44. P. Hariharan, B. F. Oreb and T. Eiju: Appl. Optics **26**, 2504 (1987)
45. D. Gale, M. I. Pether and J. C. Dainty: Appl. Optics **35**, 131 (1996)
46. C. J. R. Sheppard and T. Wilson: P. Roy. Soc. Lond. A **295**, 513 (1980)
47. M. Roy and C. J. R. Sheppard: Optics and Lasers in Engineering **37**, 631 (2002)

14 Surface Plasmon and Surface Wave Microscopy

Michael G. Somekh

Summary. Surface plasmons are electromagnetic surface waves confined to the region between two dielectrics and a conductor. These waves have unique features, since the field at the interface between the conductor and one of the dielectrics becomes very large with the result that surface plasmons are especially sensitive to the properties of any material or discontinuities at or close to this interface. This makes microscopy techniques based on surface plasmons very attractive for studying surfaces and thin films, such as, for instance, cell membranes and biological monolayers. The physical properties of surface plasmons also mean that the image formation when using surface plasmons is signficantly different from the situation that occurs when unconfined light waves are involved. This chapter will discuss methods both old and new, which utilise the unique properties of surface plasmons-and indeed other surface wave modes- which may be harnessed to make effective and high resolution microscope systems for imaging surface properties.

Abstracting terms: Surface plasmon, surface wave, interface, scanning microscopy, interference microscopy, widefield microscopy, label-free detection, fluorescence.

14.1 Introduction

Surface plasmons (SPs)-or surface plasmon polaritons- are guided electromagnetic surface waves (Raether, 1988). In their simplest form these are bound in a metallic or conducting layer sandwiched between two dielectrics. The confinement of these waves makes them particularly attractive as label-free chemical and biological sensors being especially sensitive probes of interfacial properties, Flanagan and Pantell (1984), Green *et al.* (1997). On the other hand, the fact that SPs have characteristics distinct from waves propagating in free space means that the optimum way to perform high resolution microscopy can differ significantly from other microscopy techniques.

The purpose of this chapter is not to provide an exhaustive review of SPs or indeed even SP microscopy, but rather to examine approaches to obtaining high lateral resolution and sensitivity with SPs. The emphasis will be how the particular characteristics of SPs affect the way microscope systems perform. Most of the observations and conclusions in this chapter will also be applicable not only to SP microscopy, but more generally to microscopies exploiting surface waves. Near field methods for SP microscopy (Konopsky *et al.*, 2000, Bozhevolnyi, 1996, Hecht *et al.* 1996), can, of course, give very high lateral resolution, but will not be discussed in detail, since these are not readily compatible with biological microscopes and also involve

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

inconvenient probing geometries as we will discuss later. We will briefly mention one technique (which is an intriguing combination of near field and far field microscopy) that offers potentially spectacular resolution, albeit with a rather unusual configuration (Smolyaninov *et al.*, 2005a and 2005b).

SP microscopy in the far field with lateral resolution similar to that of good quality optical microscopes is a relatively young field; for this reason the reader will not find an exhaustive range of applications, however, the techniques are now beginning to prove themselves.

This article is organised along the following lines. In Sect. 14.2 we will describe the underlying physics necessary to understand the key issues that make SP microscopy unique. In particular, we discuss excitation and detection of SPs in the context of the variation of reflectance functions with angle of incidence. We also give a brief discussion of the field distributions associated with these waves. Section 14.3 discusses some of the underlying mechanisms of surface wave generation and how this impacts on contrast in surface wave microscopy. Section 14.4 reviews SP microscopy using the prism based Kretschmann configuration. Section 14.5 discusses some of the necessary and desirable features required of objective lenses when they are used for SP microscopy. Section 14.6 considers objective based microscopy that does not rely on interferometry, whereas Sect. 14.7 considers the use of SP interferometry. Section 14.8 discusses the long term role of SP microscopy and future areas where the techniques are likely to have an impact.

14.2 Overview of SP and Surface Wave Properties

The simplest system on which SPs exist is between the interface of a relatively thin (in terms of the free space wavelength) conducting layer located between two dielectric layers. This situation is depicted schematically in Fig. 14.1a. A condition for the existence of SPs is that the real part of the dielectric permittivity, ε, of the conducting layer is negative. Physically SPs (Raether, 1988) may be thought of as collective oscillations of charge carriers in the conducting layer, with the result that energy is transferred parallel to the interface. In practice noble metals such as gold and silver are excellent metals to support the propagation of SPs and indeed the great majority of SP studies in the literature use these metals. Aluminium also supports SP propagation although the attenuation with this metal is very severe, since the positive imaginary part of the dielectric permittivity is relatively large.

A key property of SPs is that the $k-$vector is greater than the free space k-vector. This means that SPs cannot be excited from free space unless evanescent waves or a structured surface such as a grating is used. In order to overcome this problem light from a region of high refractive index can be used to illuminate the sample.

The variation of reflection coefficient with incident angle from a thin metal film (Azzam and Bashara, 1986) gives considerable insight into the behaviour of SPs. Fig. 14.2 shows the modulus of the amplitude reflection coefficient for 633 nm incident wavelength, for the system shown in Fig. 14.1a, when $n_1 = 1.52$,

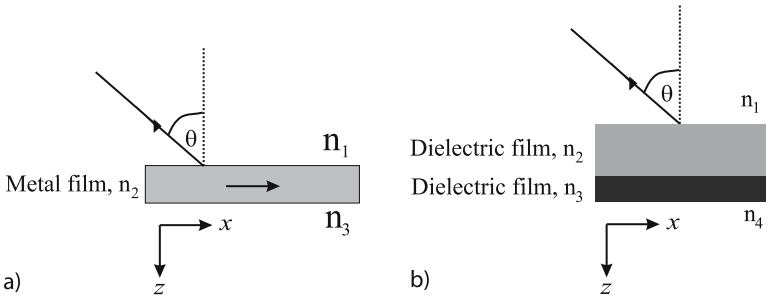


Fig. 14.1. a Simple three-layer system allowing the excitation of surface plasmons, consisting of a thin metal layer sandwiched between two dielectrics. Excitation occurs from the layer of higher refractive index. **b** A fourlayer system supporting surface waves, generally the layer n_2 has a lower refractive index than n_1 and n_3

$n_2 = \varepsilon^{1/2} = (-12.33 + 1.21\mathrm{i})^{1/2} = 0.17 + 3.52\mathrm{i}$, and $n_3 = 1$. These values correspond to the refractive index of glass or coupling oil, gold and air. The thickness of the gold layer in the simulations is 43.5 nm. We see that for *s*-incident polarisation there is not a great deal of structure in the modulus of the reflectivity. On the other hand, for *p*-incident polarisation there is a sharp dip at an incident angle, θ_p close to 43.6 degrees; this corresponds to excitation of SPs, which can only be excited for *p*-incident polarisation. The reduction of reflection coefficient relates to the fact that when SPs are excited some of the energy is dissipated in ohmic losses in the film. Excitation of SPs at an angle of approximately 44 degrees is indicates that the *k*-vector of the SPs is greater than the free space *k*-vector. From the position of the dip we can see that the *k*-vector of the SP is $n_1 \sin\theta_p \approx 1.05$ times greater than the free space propagation constant.

The modulus of the reflection coefficient for *p*-polarised light shown in Fig. 14.2 is indicative of the excitation of SPs when losses are present. If, however, the refractive index of the metal film is entirely imaginary the dip disappears, although SPs are still excited. The dip in the modulus is thus not an essential manifestation of the SP excitation. The phase of the reflection coefficient, in fact, gives real evidence of the excitation of surface waves, we will show, moreover, that it is the phase of the reflection (and transmission) coefficient, which allows one to exploit new imaging modes. Figure 14.3 shows the phase of the reflection coefficient for both *s*-polarised and *p*-polarised light. The phase behaviour of the *s*-polarised light, like the amplitude, shows little structure, but the phase variation of the *p*-polarised light shows a strong feature close to the angle θ_p, around this angle the phase changes through close to 2π radians. This is a more fundamental feature of SP excitation than the dip in the modulus of the amplitude reflection coefficient.

Figures 14.4 and 14.5 show the amplitude and phase of the reflection coefficient versus angle for a dielectric layer guiding structure defined by the figure caption. Note that above the critical angle the modulus of the reflection coefficient is one, because there are no losses in the system, the phase plots show a rapid 2π radian phase

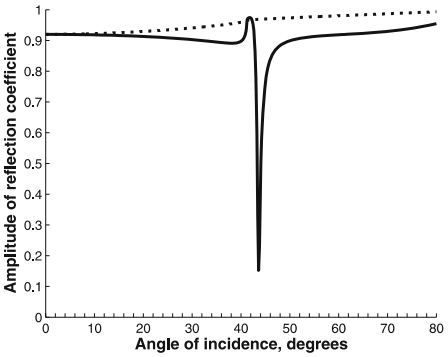


Fig. 14.2. Modulus of the reflection coefficient for a three material system consisting of a gold layer sandwiched between two semi-infinite layers of refractive indices. $n_1 = 1.52$ and $n_3 = 1$. For gold layer $n_2 = \varepsilon^{1/2} = (-12.33 + 1.21\mathrm{i})^{1/2}$ layer thickness = 43.5 nm. Optical wavelength 633 nm. *Solid line p*-incident polarisation, *fine dashed line s*-incident polarisation

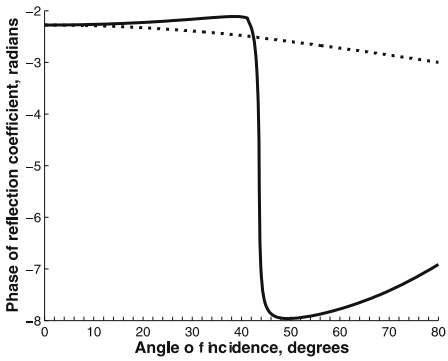


Fig. 14.3. Phase of reflection coefficient corresponding to case described in Fig. 14.2, note for the reflection coefficients the phase of the *p*-polarisation is shifted by π radians, so that they match at normal incidence. *Solid line p*-incident polarisation, *fine dashed line s*-incident polarisation

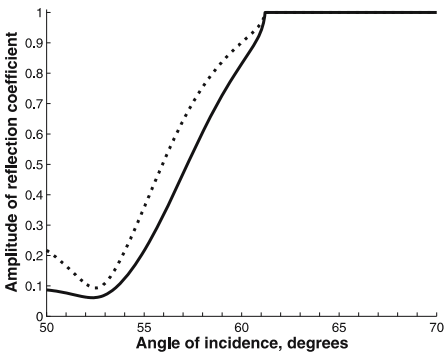


Fig. 14.4. Modulus of the reflection coefficient for a guided wave structure consisting of two dielelectic layers sandwiched between two semi-infinite layers of refractive indices. $n_1 = 1.52$ and $n_4 = 1.33$, layer $n_2 = 1.33$ layer thickness = 350 nm, layer $n_3 = 1.45$, layer thickness = 750 nm, optical wavelength 925 nm. *Solid line p*-incident polarisation, *fine dashed line s*-incident polarisation

shift corresponding to surface wave excitation, surface waves are produced for both *p*- and *s*- polarisations, albeit at slightly different angles of incidence. Physically the phase shift corresponds to the movement of energy across the interface. It is easy to see how this arises by multiplying the incident angular spectrum distribution with a linearly varying phase shift. On retransforming back into the spatial domain application of the Fourier shift theorem shows that the original field is displaced. It is easy to show that for constant reflectivity the gradient of the phase change around the critical angles is proportional to the propagation length of the surface waves.

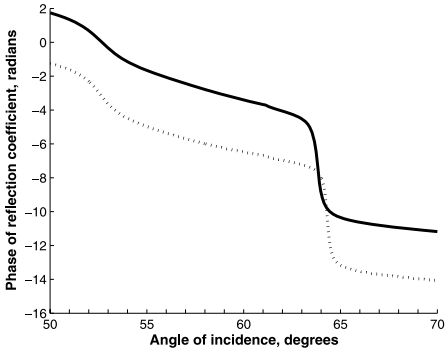


Fig. 14.5. Phase of reflection coefficient corresponding to case described in Fig. 14.4. *Solid line p*-incident polarisation, *fine dashed line s*-incident polarisation

This effect is essentially the Goos Hänchen shift (Born and Wolf, 1999) associated with light or the Schoch displacement (Brekhovskikh, 1960) that occurs with surface acoustic wave excitation in ultrasonics. The phase variation of the reflection coefficient around the angle for SP excitation contains much of the information concerning the propagation of the SPs. We will show later that the phase structure of both the reflection and transmission coefficients play a crucial role in developing high resolution SP microscopy methods. There is lateral displacement of energy at a critical angle on an unlayered sample but this is very small (except very close to the critical angle) compared to the shift around a pole in the reflection coefficient corresponding to a surface wave, where the rate of change of phase with incident angle is much greater.

The propagation of SPs is nicely demonstrated in the simple experiment depicted in Fig. 14.6a. This shows SPs excited in the metal by a weakly focused beam from a 633 nm HeNe laser source. This arrangement for excitation of SPs is the so-called Kretschmann configuration (Raether, 1988), is described in more detail at the end of this section. The SPs propagate in the metal film and leak back into reflected modes or are converted to heat by the absorption of the metal film. For a smooth uniform film there will be no 633 nm radiation propagating into the air since the light is evancescent in this medium. The gold film was then heated periodically with a frequency doubled YAG laser (532 nm) modulated at a frequency of several kHz. The effect of the periodic heating produced by the green laser was to alter the local optical properties of the metal layer and the surrounding dielectric. This perturbation caused the SPs to be converted into light waves propagating in the air. This occurs because the k-vector of the local perturbation can change that of the SP so that the resulting k-vector corresponds to a propagating mode. The 633 nm light scattered by the green laser was detected with the photodetector, after passing through an interference filter which blocked the 532 nm radiation. The distribution of scattered light corresponds closely to the SP distribution and is shown in Fig. 14.6b where see that there is a peak corresponding to the position of excitation and a 'hump' demonstrating the lateral propagation of the SPs. The dip between the two maxima arises from destructive interference between the directly reflected light and that converted to SPs (Velinov *et al.* 1999). The idea of interference between the two terms

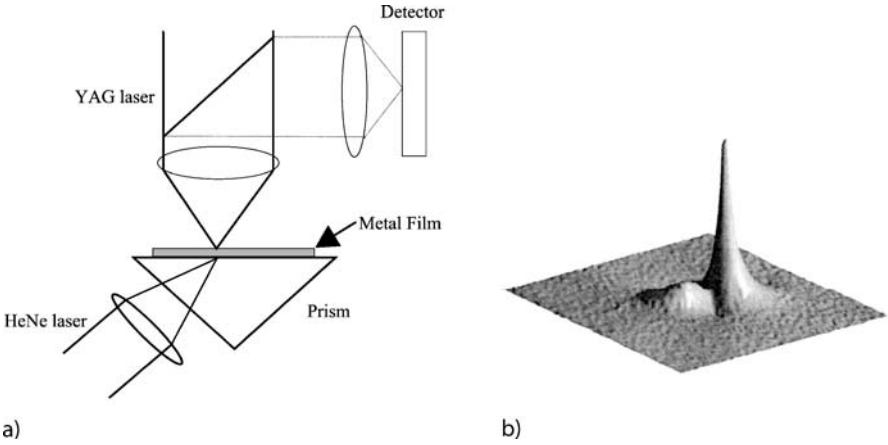


Fig. 14.6. a Schematic diagram of equipment used in photoscattering experiment, consisting of HeNe laser to generate surface plasmons and Nd doubled YAG laser to scatter them into propagating modes. **b** Experimental results obtained with equipment shown in **a**. Picture dimensions: 90 by 80 microns

is brought out explicitly in Sect. 14.3, where a simplified model to explain the form of the surface wave reflection coefficient is developed. When the polarisation of the incident 633 nm radiation was changed to be predominantly *s*-polarised very little signal was observed because of the field enhancement property discussed later in this section. Moreover, from the small amount of signal that could be seen there was no evidence of lateral displacement.

The extreme sensitivity of SPs arises from the large field enhancement at the interface between n_2 and n_3 (Fig. 14.1a). This is shown in Fig. 14.7a for the case of gold at incident angle of 43.6 degrees. Interestingly, the field strength increases with depth in the metallic film. This is due to the large 'reflected' evanescent field, which decays away from the source and grows towards it.[1] There is an extremely large field at the interface between the metal layer and the final dielectric, n_3. The discontinuity in the absolute magnitude of the field at this interface arises because it is the tangential component of the *E*-field that is conserved across the interface. For *s*-polarised incident light, on the other hand, the field decays in the metal so the field at the interface is rather small, the field is continuous across the interface since the *E*-field vector lies in the plane of the interface. If we again consider an 'ideal' material with the same negative permittivity as gold and zero imaginary part the field distribution of Fig. 14.7b results. This shows even greater *p*-field enhancement compared to gold and would be even more sensitive to surface properties. We

[1] This viewpoint does not contravene conservation of energy as the field corresponding to the SPs is evanescent in medium 3. An interesting extension of this concept has been proposed by Pendry (2000), where he predicts that it is possible, provided certain other conditions are met, to obtain a infinitely sharp focal spot by enhancing the evanescent wave components.

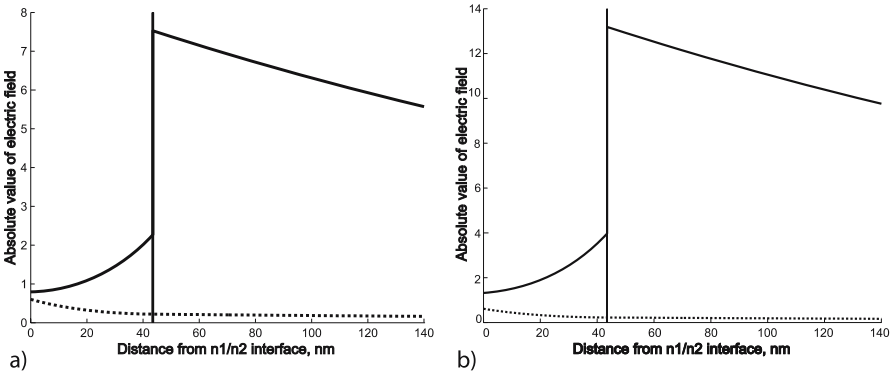


Fig. 14.7. a Modulus of the electric field distribution through the sample described in Fig. 14.2. Note the field increases increases throughout the sample. Angle of incidence from $n_1 = 43.6$ degrees. *Solid line p*-incident polarisation, *fine dashed line s*-incident polarisation. **b** Same case as Fig. 14.4a above except the metal layer is replaced with 'lossless' gold, that is $n_2 = \varepsilon^{1/2} = (-12.33)^{1/2}$

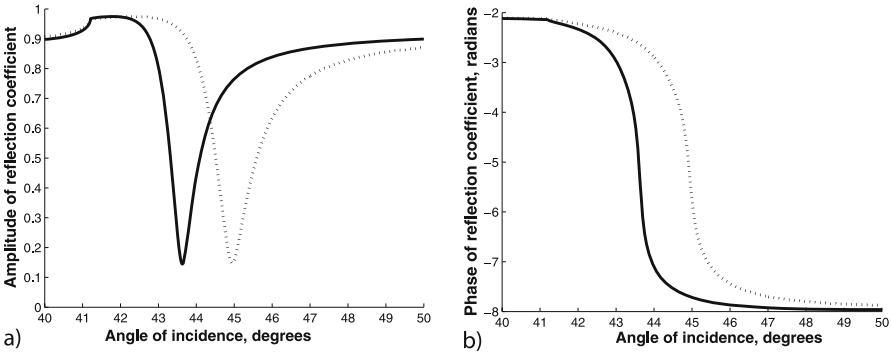


Fig. 14.8. a Comparison of modulus of reflection coefficient for bare gold layer as shown in Fig. 14.2 (*dashed line*) with same layer coated with a 10 nm layer of SiO_2 (*continuous line*). Figure 14.6b Phase plot corresponding to Fig. 14.5a. For *p*-incident polarisation

note that the *s*-fields for 'real gold' and 'ideal gold' differ by about 2% indicating a relatively weak interaction with the metal layer.

Figure 14.8a and b compares the amplitude and phase of the reflection coefficients for a single gold layer in contact with air and a similar structure with an additional intermediate layer corresponding to 10 nm of silicon dioxide, for clarity a relatively narrow range of incident angles is shown. We see that there is a considerable change in the position and shape of both the amplitude and phase features corresponding to SP excitation. In the case of *s*-incident polarisation the effect of the layer is much less noticeable. In Sect. 14.8 we will discuss differences in the sensitivity to interfacial condition compared to total internal reflection microscopy (Axelrod (1984), see also Chap. 8 in this book.

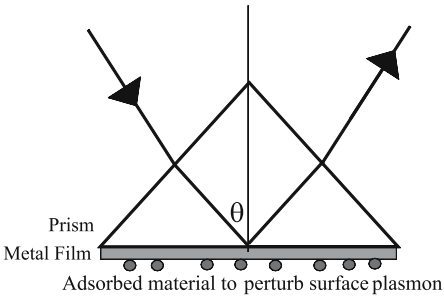


Fig. 14.9. Schematic diagram showing Kretschmann configuration used to excite surface plasmons on the hypotenuse of a prism

The most popular way to excite and detect SPs is to use the Kretschmann configuration, see Fig. 14.9. The metal film-usually gold or silver-is deposited on the hypotenuse of the prism and the light is incident along one of the other sides so it hits the metal film through the glass at an angle, θ, close to θ_p. The passage of the light through the glass increases the k-vector of the incident wave so that its component parallel to the surface matches that of the SP. The changes in the reflected light enable the properties of materials absorbed on the layer to be sensitively monitored. SPs find wide use as chemical and biosensors for application such as monitoring antibody-antigen reactions, cell attachment and toxicology (and indeed any areas where sensitivity to minute changes in the properties of an interface is required). Standard Kretschmann based SP techniques are thus extremely powerful, they do, however, lack the spatial resolution required to image highly localised regions on, for instance, a cellular level. For this reason there has been substantial work to develop SP microscopy techniques that combine high lateral resolution with good surface sensitivity.

14.3 Surface Wave Generation and Contrast Mechanisms in Surface Wave Microscopy

In this section a simple model is developed that describes some of the key features of SP and surface wave imaging; the concepts are essentially the same, so in this section the terms ‘surface wave’ and ‘surface plasmon’ are generally used interchangeably. We will develop a, somewhat heuristic, model which gives insight into the process of SP generation and explains key features in the reflection coefficient discussed in Sect. 14.2.

Consider a plane wave incident on a sample as shown in Fig. 14.10. The parallel lines represent wavefronts on the plane wave incident along the direction indicated by the dashed arrowed line. Consider that each line (indicated as a point in Fig. 14.10) on the wavefront gives two contributions to the reflected light. The first will be a direct reflection at the point of incidence marked on the figure ‘spec’ for ‘specular reflection’; the second contribution arises from incident light that is converted to SPs and reradiated back into the incident medium with a characteristic decay length, this is indicated as ‘plas’ on the figure, note that this contribution is

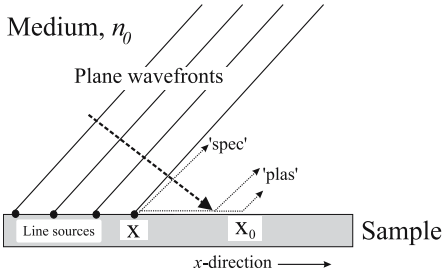


Fig. 14.10. Schematic diagram explaining application of the simplified Green's function model

distributed over the whole of the sample plane since the decay length defines only an average propagation length not a specific position of reradiation. The contribution from each point can be considered to be independent of the angle of incidence and in both directions, on the other hand, the contribution to the overall reflection coefficient is very sensitive to the angle of incidence since efficient excitation of surface waves relies on the contributions from each point adding in phase, which, of course, occurs when the component of the k-vector parallel to the interface matches that of the surface wave.

In the following discussion we use a simplified Green's function based to develop and approximate reflection coefficient[2]. It is, in fact, quite possible to develop a more accurate Green's function based on the true reflection coefficient calculated from the Fresnel equations, the loss in mathematical simplicity, however, obscures the simple essential physics, it is therefore more useful to stick with the simplification.

Consider a plane wave incident from a medium, n_0, whose k-vector along the plane of the sample is k_x, represented as $\exp \mathrm{i}k_\mathrm{x}x$. We consider the reflection coefficient corresponding to the specular contribution to be -1. This approximation can be easily refined but this value serves our present purposes. The plane waves are excited on the sample from a position denoted as x on Fig. 14.10 and propagate as a surface waves to point x_0 prior to reradiating. The propagation constant for the surface wave, k_p, is given by:

$$k_p = k_\gamma + \mathrm{i}(k_\mathrm{c} + k_\mathrm{a}) \tag{14.1}$$

This propagation constant consists of the real part of the propagation constant k_γ, and two terms which contribute to the imaginary part of k_p. k_c corresponds to the attenuation due to SPs coupling into propagating photons in the medium, n_0 and the k_a term corresponds to ohmic losses in the metal layer.

We can then write the contribution to the field at x_0 due to the plane wave incident on the sample on the left of x_0, $u_\mathrm{l}(x_0)$, as:

$$u_\mathrm{l}(x_0) = 2k_\mathrm{c} \int_{-\infty}^{x_0} \exp(\mathrm{i}k_\mathrm{x}x) \exp \mathrm{i}k_p(x_0 - x)\mathrm{d}x \tag{14.2a}$$

[2] The process can be reversed, that is the Green's function can be derived as the pole contribution from an approximate reflection coefficient.

Similarly, the contribution from the wave arriving to the right of x_0, $u_r(x_0)$ is given by

$$u_r(x_0) = 2k_c \int_{x_0}^{\infty} \exp(ik_x x) \exp -ik_p(x_0 - x)dx \tag{14.2b}$$

In physical terms this means that the contribution *from a line* at x_0 arises from the summation of line sources along the incident plane wave multiplied by the complex phase factor arising from propagation from point x to x_0. The factor of $2k_c$ is the factor that accounts for the efficiency of coupling and reradiation of the SP. This factor is analogous to the coupling factor used for surface acoustic wave excitation and may be calculated by expanding the reflection coefficient around the pole as a Laurent series and evaluating the field contribution by contour integration (Bertoni and Tamir, 1973).

Evaluating this term we obtain an expression for the field above the sample due to surface waves as:

$$u_p(x_0) = u_l(x_0) + u_r(x_0) = 2k_c \exp(ik_x x_0) r(k_x, k_p) \tag{14.3}$$

where $r(k_x, k_p) = \frac{i}{(k_p - k_x)} + \frac{i}{(k_p + k_x)}$

We can see that this term reaches a maximum value very close (exactly so if the first term only is considered) to $k_\gamma = k_x$ and decreases away from this value, as mentioned earlier this maximum arises from the sources adding in phase along the sample surface. Note that for waves as depicted in Fig. 14.10 the wave propagating from left to right greatly dominates the backward propagating wave. For plane wave excitation only those angles close to the phase matching condition need to be considered. From the foregoing arguments we see that the approximate reflection coefficient, $R(k_x)$, corresponding to surface wave excitation is given by:

$$R(k_x) = -1 + 2k_c r(k_x, k_p) \tag{14.4}$$

As mentioned earlier a more accurate expression for the reflection coefficient can be extracted using the reflection coefficient obtained from the Fresnel equations, using a mean value of reflection coefficient corresponding to the values away from the pole in the reflection coefficient.

Figure 14.11 shows an example of the idealized reflection coefficient close to the region where the surface waves are excited. We note that the dip in the reflection coefficient occurs very close to a normalized incident wave number of unity, where this quantity represents the ratio of the incident wave number to the real part of the wave number of the surface wave (and is therefore proportional to the sine of the incident angle). We also note the characteristic rapid phase change through 360 degrees. Figure 14.12 shows how this arises. The circle represents the locus of the reflectivity as the incident wave number changes. The points, a,b,c and d on Figs. 14.11 and 14.12 represent normalized incident wavenumbers of 0.98, 0.99, 1.005 and 1.015 respectively.

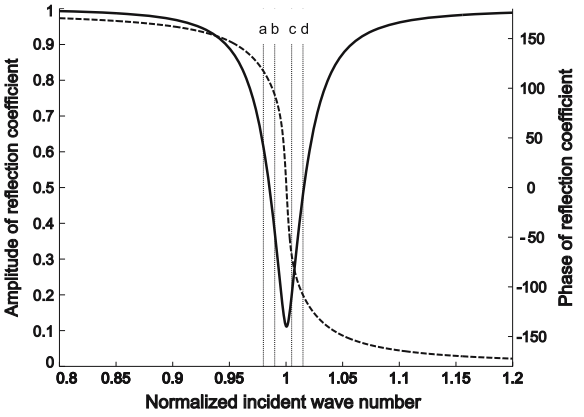


Fig. 14.11. Calculated amplitude (*solid*) and phase (*dotted*) reflection coefficient. For 'overcoupling' case, $k_a = 0.012k_\gamma$, $k_c = 0.015k_\gamma$

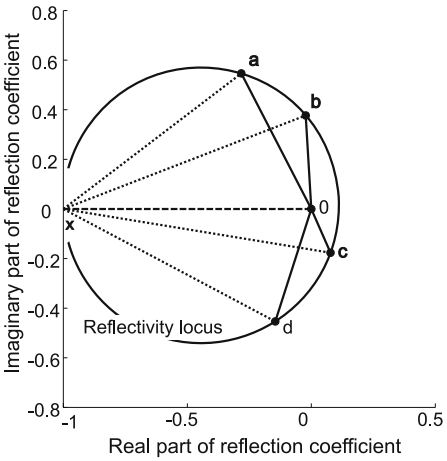


Fig. 14.12. Rotating phasor model showing formation of reflection coefficient. *Ox* is a 'constant' specular reflection, *xa*, *xb*, *xc*, *xd* are surface wave contributions that change with incident angle with a resultant reflection *Oa*, *Ob*, *Oc* and *Od* respectively. Points 'a', 'b', 'c' and 'd' correspond to angles marked on Fig. 14.11

We can see how the particular values of the reflection coefficient are formed from Fig. 14.12, at the smallest angle of incidence, there is a constant[3] non-surface wave reflection whose value is −1 (0x) on the figure this is added to a surface wave contribution 'xa', which add to give a resultant relectivity '0a'. We see that as the incident angle approaches unity the surface wave contribution increases but the modulus of the reflectivity decreases. The rapid phase change is also apparent as the incident angle changes from 'a' through to 'd'. In the example presented here the $k_c = 1.25k_a$, this corresponds to the case with SPs of 'overcoupling' where the metal layer is thinner than the value required for the reflectivity to go to zero. The reflectivity locus has a radius of unity when $k_c = k_a$, so that the reflectivity falls to zero at resonance, where the 'specular' reflectivity cancels the 'surface wave' contribution. The situation of 'undercoupling' is represented by the surface wave contribution

[3] For the true reflection coefficient this value is not a strict constant but, around the wave number of the surface wave it is very slowly changing, compared to the rapidly changing surface wave contribution.

being insufficiently large to cancel the specular reflection that is $k_c < k_a$. Another interesting case is when there is no absorption, that is when $k_a = 0$. In this case the reflectivity locus follows the unit circle which means the reflectivity always has a magnitude of one although the phase changes through 360 degrees as discussed above. In this case the surface wave vector reaches a maximum magnitude of 2. The analogy to this situation would apply to a metal with no losses, that is with negative real dielectric constant (as modeled in Fig. 14.7b). Again this confirms that the phase change is a more reliable indicator of surface wave activity than the modulus reflectivity.

We will now consider mechanisms that give image contrast in the SP microscope. Simply these can be divided into two:

1. A change in the local propagation constant for the surface wave due to a change in the local resonant conditions.
2. Scattering of a surface wave into a propagating wave which is collected by a microscope objective

The model discussed above provides a very convenient way to discuss the first mechanism. Let us consider the simplest inhomogeneous sample consisting of regions with different SP k-vectors, k_{p1} and k_{p2}, as shown in Fig. 14.13. This corresponds to the situation where the propagation of the SP is affected by, for instance, the presence of an overlayer.

Now consider a single plane wave with incident wavevector, k_x, close to k_γ, this ensures that excitation of a wave propagating from left to right dominates (see Fig. 14.13) the output field is therefore given by a specular reflection and three principal surface wave terms namely.

(i) a field generated in region 1 and reradiated in region 1,
(ii) a field generated in region 1 and reradiated in region 2,
(iii) a field generated in region 2 and reradiated in region 2.

In addition there is a field generated in region 1 which is reflected from the interface and reradiates back in region 1. Calculating the contribution for each region allows us to calculate the field arising for a plane wave incident on an inhomogeneous sample. (Zhang *et al.* 2006).

Figure 14.14 summarises the intensity responses we would expect from bright field SP microscopy with an incident single plane wave to excite surface waves from left to right. The solid curve in Fig. 14.14 shows a wave slightly off resonance in

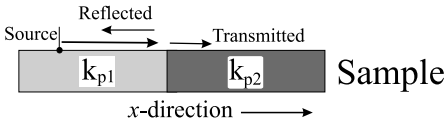


Fig. 14.13. Schematic diagram showing reflection and transmission of a source from a structured sample with different surface wave wavevectors

the left hand medium, as the wave passes into the region where the surface wave is in resonance the signal decreases with a characteristic decay length related to the strength of the coupling and the ohmic losses. We note that the wave does not decay to zero since the case was chosen to match a 45 nm layer of gold where k_c and k_a are not precisely equal as shown in the figurecaption. The oscillations from the interface are due to reflection forming standing waves which have not been observed in our experiments presumably because the interface is not sufficiently smooth, in prism based SP imaging the spatial bandwidth is insufficient to resolve these features. The dashed curve shows the situation where the wave is on resonance in the left hand region and passes to a region where the surface wave is off resonance with a consequent increase in signal level. Figure 14.15 shows exactly the same situation except that coupling and attenuation are increased, we see essentially the same phenomena except that the transitions are much more rapid corresponding to improved resolution. On the other hand, we note that the change in signal is now smaller because of the gradient of the phase change of the reflectance function is much slower. This argument above provides a formal framework consistent with the explanations of the tradeoff between resolution and sensitivity discussed throughout the literature of prism based surface wave microscopy see, for instance, (Berger *et al.* 1994).

Much of the literature for SP imaging has pointed out that that the lateral resolution in the direction of surface wave propagation is poorer than in the direction normal to the propagation direction. This is based on the idea that the first mechanism predominates along the direction of propagation and the scattering mechanism applies to normal to it. This is clearly something of a simplification since the scattering mechanism can operate both parallel and normal to the direction of propagation.

To fully understand the imaging behavior when there is excitation over a range of azimuthal angles, as is the case for the objective based microscopy described in Sect. 14.6 a Green's model for a point source is needed; this can be used to explain why the spatial resolution of bright field SP microscopy through a microscope objective is superior to that obtained with prism based excitation, however, for even objective based excitation the lateral resolution depends to some extent on the decay of the surface waves.

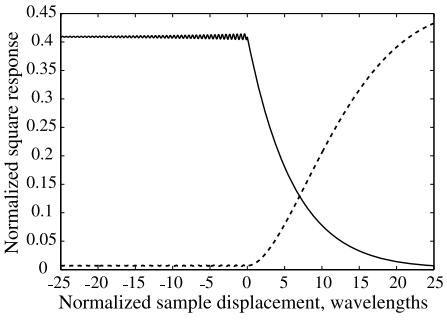


Fig. 14.14. Surface wave response across an interface on and off resonance for 'weak' coupling for a single incident angle k_x. One unit of wavelenth is the surface wave wavelength in the faster medium. Weak coupling corresponding to $k_a = 0.0055k_\gamma$, $k_c = 0.0065k_\gamma$ *Solid curve*: Left hand medium $k_\gamma = 1.01k_x$, right hand medium $k_\gamma = k_x$. *Dashed curve*: Left hand medium $k_\gamma = k_x$, right hand medium $k_\gamma = 1.01k_x$

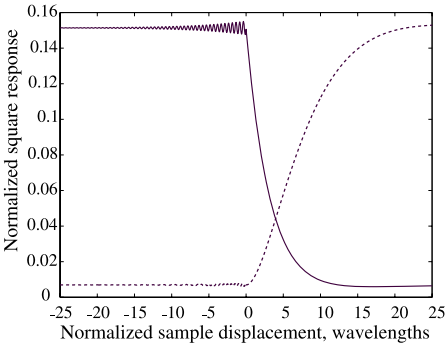


Fig. 14.15. As Fig. 14.14 except for 'strong' coupling corresponding to $k_a = 0.011k_\gamma$, $k_c = 0.013k_\gamma$ *Solid curve*: Left hand medium $k_\gamma = k_x$, right hand medium $k_\gamma = 1.01k_x$. *Dashed curve*: Left hand medium $k_\gamma = k_x$, right hand medium $k_\gamma = 1.01k_x$

The scattering mechanism applies to, for instance, point objects and is related to the scattering of the SPs into propagating light. Provided the scattering is relatively weak so that it does not perturb the propagating plasmon field greatly the resolution obtained is dependent on the detection optics. A large field enhancement associated with surface waves and plasmons will increase the extent of the scattering and thus increase the image contrast. For surface wave fluorescence as discussed in Subsect. 14.6.3, the situation is similar in that the fluorphore can be regarded as local inelastic scatterer, so that the resolution is derived from the detection optics and the wavelength of the SPs rather than their decay length.

Finally, it is worth discussing the desirable properties of the surface waves for optimum imaging performance. There is very clear tradeoff between the lateral resolution and contrast when mechanism 2 prevails, as discussed many times in the literature and exemplified by application of the Green's function model in Figs. 14.14 and 14.15. The model shows how the width of the dip affects the extent of the contrast and how this is intimately linked to the lateral resolution, there is, however, one other factor that needs to be taken into account. Namely, how much does the dip move when the surface is modified by a given structure, say, a protein monolayer as discussed in Sects. 14.6 and 14.7, in other words, what is the change in k_p for a given change in the surface. In this case SPs are superior to most other surface wave modes, where even if the dip is very narrow the change in k_p is very small when a small dielectric overlay is deposited, so that the resulting change in signal is small. On the other hand, surface wave structures, as shown in Fig. 14.1b, are often associated with even larger field enhancements compared to SPs so that if the scattering mechanism predominates these surface structures way well be preferred to surface waves. Indeed, the fact that the resonance does not change substantially may be a real advantage in this situation since the waves will be strongly excited even when the surface properties change slightly.

14.4 Surface Plasmon Microscopy – Kretschmann Prism Based Methods

In the mid-nineteen eighties the idea of combining the sensitivity of SPs with good lateral resolution took hold. The earliest work in the field was that of Yeatman and Ash (1987) in the UK and Rothenhausler and Knoll (1988) in Germany. Figure 14.16 combines two wide field imaging modes based on the Kretschmann configuration. The sample is illuminated with a plane beam of light at an incident angle close to θ_p. In the 'bright field' configuration the light reflected from the sample is imaged onto the light sensitive detector where the image is recorded. Local variations in the SP propagation and local scattering will change the intensity of the reflected light allowing an SP image to be formed. Detection can also occur on the far side of the prism as also shown in Fig. 14.16; this effectively produces a 'dark field' configuration. In this case discontinuities in the sample scatter the SPs into propagating waves on the far side of the prism, rather like the pump beam in the experiment described in Sect. 14.2. This scattered light is then imaged onto a light sensitive detector. Clearly, in this case a continuous uniform film will not scatter any light so the image will appear dark.

Figure 14.17 shows an alternative scanning configuration. In this case a weakly focused beam is incident on the sample and the reflectivity is monitored as function of scan position. In fact the imaging properties of this system are similar to that the 'bright field' wide field system shown in Fig. 14.16. This scanning configuration illustrates the trade-offs implicit in bright field Kretschmann based SP imaging. In order to improve the lateral resolution one needs to focus the beam as tightly as possible, using a large numerical aperture. On the other hand, SPs are only excited over a narrow range of angles so that the numerical aperture must be limited to this range. A sharp phase change corresponds to a large decay length, which, in turn,

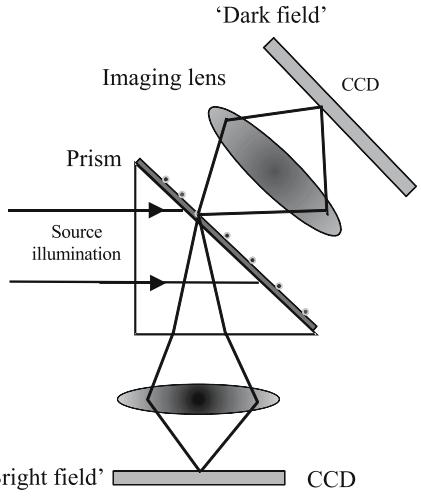


Fig. 14.16. Schematic diagram showing wide field 'bright field' and 'dark field' microscope systems based on the Kretschmann configuration

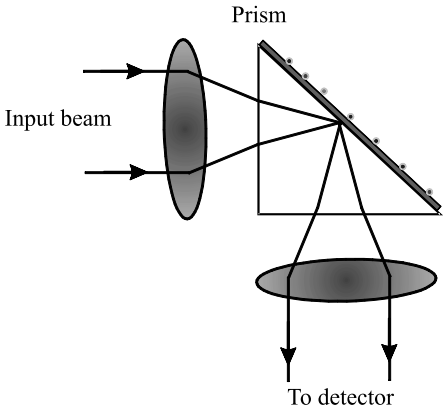


Fig. 14.17. Schematic diagram showing scanning system based on the Kretschmann configuration

restricts the maximum useful numerical aperture of the exciting beam and hence the lateral resolution. Yeatman (1996) has shown how improving the lateral resolution of the system results in reduction of sensitivity. Similarly, the 'bright field' wide field system has lateral resolution limited by the decay length of the SPs.

The lateral resolution of the prism based SP microscope has been studied in detail by Berger *et al.* (1994). When the observed contrast arises from the change in the k-vector of the SPs in different regions of the sample, it was demonstrated experimentally that the propagation distance or decay length of the SP determined the lateral resolution that could be obtained in a test structure, giving essentially the same conclusions as those summarised in Figs. 14.14 and 14.15. Using the fact that the decay length of the SPs depends on the excitation wavelength, Berger *et al.* (1994) were able to show that the lateral resolution could be improved by using a shorter wavelength with a correspondingly smaller decay length. The best resolution obtained was approximately two microns obtained at 531 nm, unfortunately when the decay length is this small the sensitivity to the surface properties is greatly reduced. A similar approach to improving the lateral resolution has been presented by Geibel *et al.* (1999), where a coating of aluminium was used, which has a relatively large imaginary part of the dielectric constant with a consequent reduction in decay length. Once again the improvement in lateral resolution is obtained with a reduction in sensitivity to surface properties. Despite this, useful images monitoring cell attachment and motility were obtained.

Figure 14.16 shows the 'dark field' system where SP imaging is obtained from the *scattering* on the far side of the prism. In this case if the scattering of the particle is imaged onto the CCD with a sufficiently high numerical aperture system then it possible to obtain lateral resolution that exceeds that of the decay length of the SPs. Improved resolution using this approach does not appear to have been demonstrated in practical situations. This approach has the considerable disadvantage since it is necessary to probe on the far side of the sample, which means that the detection optics restricts access to the sample. 'Dark-field' SP microscopy can, however, be obtained in reflection by blocking the zero order reflection in the Fourier plane as

described by Grigorenko *et al.* (2000), this will allow only scattered light to be detected. We discuss how analogous dark field techniques can be applied to objective based SP microscopy later in this chapter.

The difficulty with Kretschmann based prism configurations arises from two principal factors:

1. the prism means the system is not readily compatible with a conventional microscope configuration, for instance, imaging through the prism is difficult, especially when the plasmon is excited close to grazing incidence.
2. the lateral resolution is poor compared to that achievable with optical microscopy, so that the range of applications will be restricted. In particular, applications in extremely important fields, such as cell biology, will be severely limited. For these reasons it is necessary to develop techniques capable of overcoming these limitations.

There has been recent interest in using the phase of the reflectance function in Kretschmann based SP imaging systems. The variation of the phase of the reflectance function around θ_p provides an alternative quantity in addition to the amplitude of the reflection. A discussion of the use phase has been discussed by Grigorenko *et al.* (1999). They have shown that as the metal thickness is changed so that the minimum of the reflectivity is zero (corresponding to a thickness of approximately 47.7 nm of gold using the parameters given in Fig. 14.2) the gradient of the phase variation with incident angle increases without bound around θ_p. This effect is easily demonstrated in the framework presented Fig. 14.12, when $k_a = k_c$. This means that close to these conditions the phase of the reflectance function can be more sensitive to small changes compared to the amplitude. This sensitivity is borne out experimentally in the work of Notcovich *et al.* (2000), who have imaged gas flows corresponding to a refractive index change of 10^{-6}.

14.5 Objective Lenses for Surface Plasmon Microscopy

One of the main themes of this chapter is obtaining high resolution surface wave microscopy with objective based microscope configurations. We will firstly review the relationship between the light incident on a microscope objective and the polarisation state and incident angles that illuminate the sample. Consider an aplanatic objective with a source of illumination in the back focal plane. Each point incident on this plane maps to an incident plane wave propagating in a direction whose k-vector parallel to the plane of incidence is proportional to the radial position of the source in the back focal plane. This is shown schematically in Fig. 14.18, which relates the k-vector incident on the sample to the back focal plane distribution. If the polarisation of the beam incident on the back focal plane is linear, the polarisation state of the incident radiation varies from pure p-polarised to pure s-polarised. In general, the polarisation is a combination of the two states, which can be determined by resolving with respect to the azimuthal angle, ϕ, as shown in Fig. 14.18.

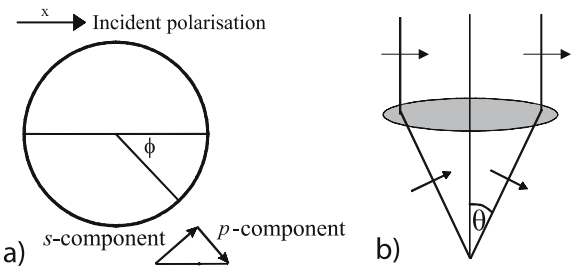


Fig. 14.18. a View of the back focal plane showing the relationship between azimuthal angle and polarisation state. **b** View in the plane of incidence showing relation between radial position and incident angle. Also shows rotation of direction of *E*-field vector passing through the objective

For an oil immersion lens with a numerical aperture greater than about 1.1 a fully illuminated objective lens will generate plane waves whose incident angle and polarisation state can excite SPs on the sample provided the final dielectric has an index close to unity. Since the *k*-vector of SPs with air backing is only a few percent greater than that in free space, even relatively inexpensive oil objectives have sufficient numerical aperture for imaging, since an NA of 1.25 which is very commonly available in low end objectives can be used very satisfactorily, especially for scanning applications at single wavelengths. Figure 14.19 shows a back focal plane distribution calculated for a gold sample for horizontally linear polarised light incident on an objective with NA = 1.25 (several experimental distributions will be presented in Sect. 14.6). Along the horizontal direction the dip corresponding to SP propagation is observed, the proportion of *p*-polarized to *s*-polarized light decreases as the azimuthal angle changes, so the strength of the dip diminishes and vanished completely in the vertical direction where the light is entirely *s*-incident. Clearly radial polarisation will give a strong dip at all azimuthal angles.

The goal of SP imaging is generally directed towards biological imaging in aqueous materials where the refractive index of the sample is, at least, 1.33 and

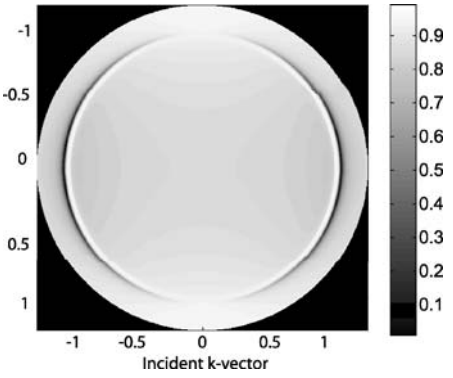


Fig. 14.19. Theoretical back focal plane distribution for a 1.25 NA for gold air system described in Fig. 14.2. Note horizontal direction represents *p*-polarisation where the dip is most pronounced and the vertical direction is *s*-polarisation where no dip is visible

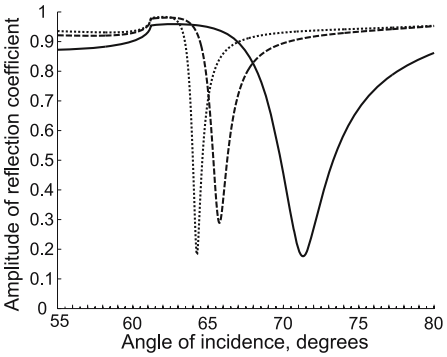


Fig. 14.20. Reflection coefficient for 43 nm thick gold layer with water backing, for *p*-polarisation at three wavelengths 633 nm (*solid*), 800 nm (*dashed*) and 925 nm (*dotted*). Refractive index values taken from Johnson and Christy (1972)

usually a few percent more. For total internal reflection microscopy a numerical aperture greater than the refractive index of the final medium is generally satisfactory, although there are advantages in exceeding this by significant extent to reduce background signal from propagating waves, so a high quality immersion objective with numerical aperture of 1.4 or 1.45 such as the Zeiss Plan Fluar 100x is a good choice offering high numerical aperture, conventional immersion oil ($n = 1.518$) and inexpensive consumables. For water based SP microscopy the situation is rather more problematic. This can be seen in Fig. 14.20, which shows the reflection coefficient for *p*-incident light from 43 nm thick gold surface with water backing for three different wavelengths: 633 nm, 800 nm and 925 nm. We can see that the angle of incidence where SPs are excited at 633 nm is approximately 71.3 degrees which means that a numerical aperture of approximately 1.44 is necessary to excite SPs.

Even a 1.45 NA objective does not have sufficient numerical aperture for SP imaging in water at visible wavelengths.[4] The solution is therefore either to use surface waves which are excited at smaller angles of incidence, such as SPs in gold at longer wavelengths (note also the decreasing width of the resonance at long wavelength denoting increasing propagation length). Another example of surface wave structures that excite surface waves at relatively small angles of incidence is shown in Fig. 14.5. The other approach is to use objectives with exceptionally high numerical apertures such as the Olympus 100x Apo with a numerical aperture of 1.65. To obtain such a numerical aperture high index immersion fluid and coverslips are required. The immersion fluid used is diiodomethane which has an index of 1.78 and this is about the highest index one can obtain before the fluids become seriously toxic. The coverslips used with this objective are sapphire. Although the objective appears to function well the fluid and the coverslips make use of this objective less than ideal. The fluid is somewhat volatile using dissolved sulphur to increase the refractive index, the fluid therefore changes as the experiment proceeds making long term quantitative experiments inconvenient. The other problem lies in the price of

[4] Nikon have introduced an objective with 1.49 NA, which uses conventional coverslips and immersion fluid, this *may* have sufficient numerical aperture for SP imaging.

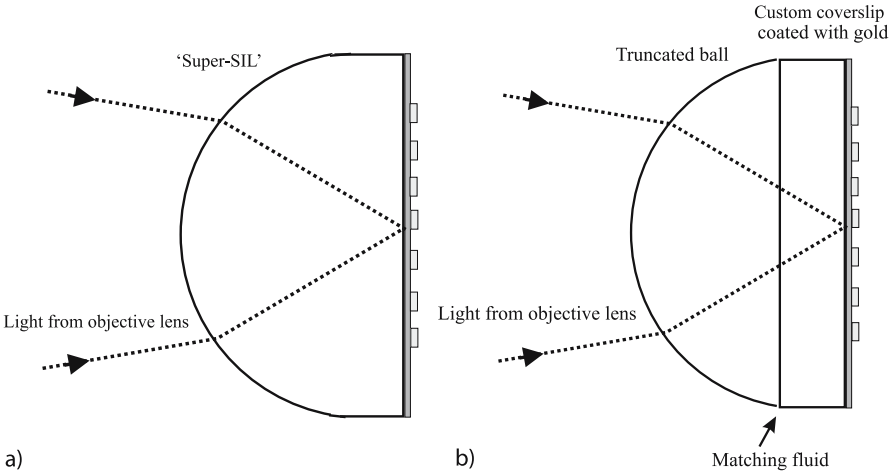


Fig. 14.21. a Schematic diagram of solid immersion (SIL) objective. **b** As above but split ball to allow sample movement

the coverslips that is nothing less than exhorbitant; it is possible to obtain well over 1000 conventional coverslips for the price of one sapphire version!

High numerical aperture lenses require a high index immersion media, since solids are available with much refractive higher indexes than liquids, the use of solid immersion lens (SILs) (Mansfield and Kino, 1990 and Terris *et al.* 1994) seems a natural means to achieve the high numerical aperture necessary for SP microscopy. Figure 14.21a shows a solid immersion SIL system. The system here is sometimes referred to as the 'super-SIL' configuration because the ball acts to both increase the incident angle and reduce the wavelength by virtue of the higher refractive index. For this configuration the ball is larger than a hemisphere by the radius divided by the refractive index. Hemispherical balls can be used where the light from the objective hits the ball at normal incidence in this case the increase in numerical aperture only arises from the reduction in wavelength of the illuminating light. A detailed study of the aberrations associated with the two configurations also shows that the 'super-SIL' configuration is greatly superior in terms of off-axis aberration, which is particularly important for widefield imaging (Zhang, 2006).

From the foregoing discussion we can see that the effective refractive index, NA_{eff}, of the 'super-SIL' configuration is given by:

$$NA_{eff} = n^2 NA \tag{14.5}$$

where n is the refractive index of the ball and NA is the numerical aperture of the illuminating lens.

The larger increase in numerical aperture as well as the improved aberration performance mean that 'super-SIL' configuration was the preferred choice in our work. In our studies we used a ball made of S-LaH79 glass ($n = 1.996$ at $\lambda =$

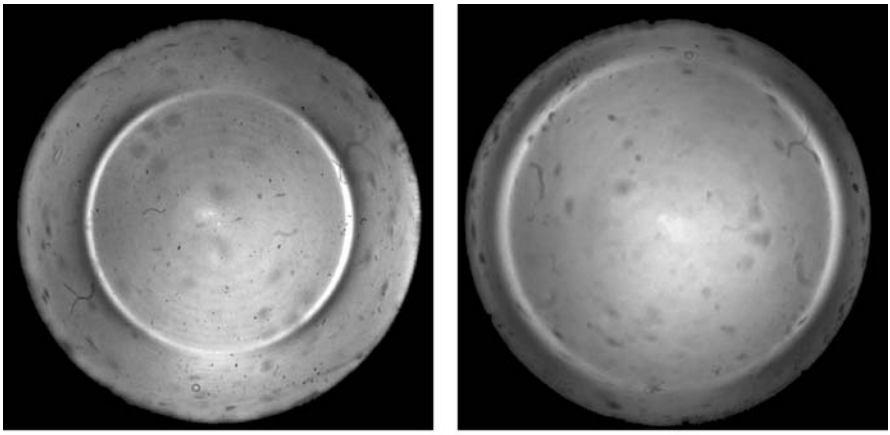


Fig. 14.22. Back focal plane distributions using split SIL structure above. **a** Air-backing. **b** Water backing

633 nm) with an objective of NA = 0.42 (large working distance Mitutoyu), in principle, this will give an $NA_{eff} = 1.67$.

The SIL configuration provides a very cost-effective way of achieving extremely high numerical apertures and numerical apertures greater than 2 are readily achievable with the combination of high numerical objectives and high index balls. Imaging was initially demonstrated using a single in tact ball with the gold and sample deposited directly on the ball (Zhang *et al.* 2004). To make the system truly practical, however, we need to be able to move the sample relative to the optical axis, otherwise our effective viewing area is highly restricted. Our approach to this problem was to split the ball into two sections (Zhang, 2006 and Zhang *et al.* 2006), consisting of a truncated ball and a coverslip (shown in Fig. 14.21b). The total thickness of the truncated ball and the coverslip was equal to the required value for the 'super-SIL'. In order to couple the truncated ball and the coverslip a matching fluid is needed, as mentioned earlier, there are no non-toxic liquids available to match the very high index glasses. We therefore used the thin layer of diiodomethane, which proved satisfactory, despite the refractive index mismatch, as shown in the back focal plane distributions with both water and air backing as shown in Fig. 14.22.

From the positions of the SP dips in both air and water we can see that the numerical aperture of the SIL ball combination is approximately 1.61, somewhat less than the theoretical value referred to above. The discrepancy is mainly due to incomplete illumination of the back focal plane of the objective. Smaller reductions in effective numerical aperture are introduced by reflection losses at the curved surface of the ball and at the interface between the ball and the coverslip. The latter effect is rather small although when conventional matching oil (n=1.518) between the ball and the coverslip was used the numerical aperture was very severely reduced. It is important that the light in the 'matching' fluid does not become evanescent; that is

to say that the refractive index of the fluid should be greater than the desired effective NA. As we mentioned earlier it should be relatively easy to obtain NAs greater than 2 with single in-tact ball, but it may be difficult to get much more than 1.7 with the split configuration described above.

In summary SP imaging in air is achieved with even modest immersion objectives, whereas water imaging in the visible is best performed with special coverslips or solid immersion objectives. Images presented in this chapter will use both fluid immersion and solid immersion approaches.

14.6 Objective Lens Based Surface Plasmon Microscopy: Non Interferometric Methods

In Sect. 14.5 we discussed how a high index objective lens can be used to excite SPs on a sample surface. This section and Sect. 14.7 will discuss objective based SP and surface wave microscopy techniques where the advantages compared to prism excitation will become apparent.

There have been several attempts to use objective lens for high resolution SP and surface wave microscopy; both widefield and scanning configurations have been presented. For convenience we divide our discussion into non-interferometric methods (this section) and interferometric methods Sect. 14.7. The interferometric methods have been less widely reported, however, we feel that they warrant a section because of the important ideas they illustrate and the fact the these methods have yielded the best resolution so far obtained with far field objective lens based methods.

14.6.1 Scanning Methods

An oil immersion objective lens was used by Kano *et al.* (1998) to produced localized generation of SPs. In addition, they demonstrated a system somewhat akin to the 'dark-field' system shown in Fig. 14.16, whereby SPs were excited on the sample through an oil immersion objective (NA = 1.3). The presence of local scatterers was detected on the far side of the sample by scattering into waves propagating away from the source. The rescattered SPs were collected with a dry objective on the far side of the sample. A simplified schematic of this system is shown in Fig. 14.23. This experiment demonstrated the potential of excitation using an objective lens as a means of exciting SPs and showed that resolution comparable to the spot size is obtainable. On the other hand, the arrangement described is not practical for most imaging applications on account of the fact that like the prism based 'dark-field' arrangement detection takes place on the far side of the sample.

Measuring the distribution of reflected light in the back focal plane is the basis of the technique of back focal plane ellipsometry, where the sample reflectivity as a function of incident angle and polarisation state may be monitored. The advantage of this technique is, of course, that the focal spot is confined to a small submicron

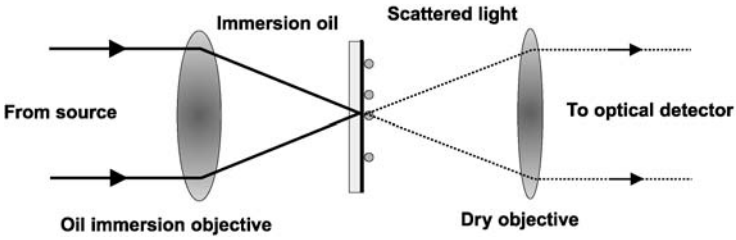


Fig. 14.23. Simplified schematic of 'dark-field' system of Kano *et al.* (1998) using fluid immersion excitation of surface plasmons

area allowing the properties to be measured in a highly localised region. This technique has been used with dry objectives by several authors to measure film thickness in semiconductors (Fanton *et al.* 1993), to compensate for material variations in sample properties when measuring surface profile (See *et al.*, 1996) and for extracting ellipsometric properties over a local region (Shatalin *et al.* 1995). This concept has also been extended by Kano and Knoll (1998) using an oil immersion objective to measure the local thickness of Langmuir–Blodgett films. In essence they measured a distribution such as the one shown in Fig. 14.19 and followed the position of the dip as the objective was focused on different regions of the sample, thus obtaining a local value for θ_p. Local measurements corresponding to 4 Langmuir–Blodgett monolayers were detected reasonably easily with this technique.

These authors subsequently extended this approach to make a scanning microscope configuration (Kano and Knoll, 2000) where the position of the ring is monitored as a function of scan position, thus allowing microscopic imaging. The lateral resolution obtained with this method appears to be approximately 1.5 microns. The difficulty in achieving the better lateral resolution may, in part, arise from the fact that the experimental distributions in the back focal plane are rather prone to the presence of interference artefacts, as well as spreading arising from SP propagation.

The fact that an objective lens can view a range of azimuthal angles has been exploited in Tanaka and Yamamoto (2003). In this technique a laser was used to illuminate the back focal plane of an objective lens, thus resulting in excitation of surface waves as discussed earlier. Four detectors were placed in the back focal plane of the objective oriented so that they would detect scattering both along and normal to the direction of propagation; different resolution and contrast was obtained by each detector depending on the orientation of the detector relative to the feature. Imaging was achieved by sample scanning. Although reasonable lateral resolution ($\approx 1.6\,\mu$m) was achieved, this method is somewhat cumbersome and it could be argued that there are easier ways to exploit the fact that the objective lens can image a range of azimuthal angles.

14.6.2 Wide Field SP and Surface Wave Microscopy

A conventional widefield reflection optical microscope with a suitable objective lens will excite SPs. Consider the system shown in Fig. 14.24, ignoring for the moment

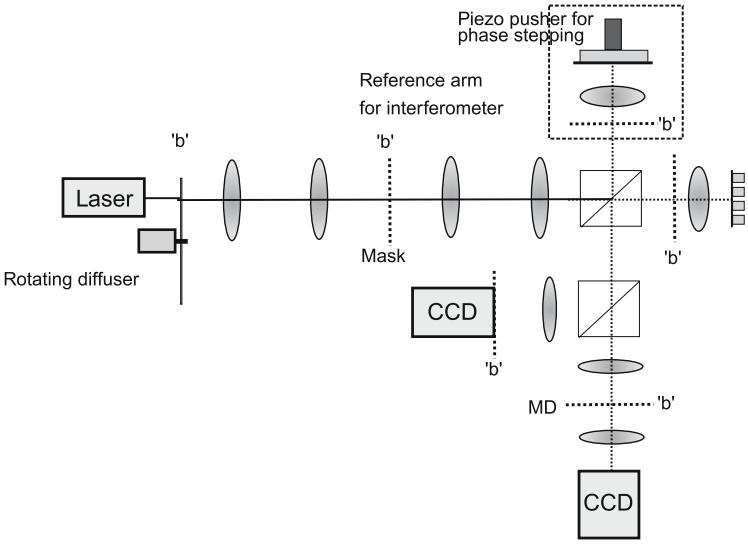


Fig. 14.24. Schematic diagram of widefield microscope systems, showing additional arm for interferometric operation

the mask and the reference channel which will be discussed in Sect. 14.7. The spatial coherence of the laser source is destroyed by the rotating diffuser, so that we have a monochromatic spatially incoherent light source conjugate with the back focal plane of the objective. Planes conjugate with the back focal plane are marked 'b' on Fig. 14.24. Two CCD cameras are used in the detection arm, one to record the back focal plane distribution and the other to record the image.

Figure 14.25a shows a back focal plane distribution from a 45 nm thick gold layer taken with the system using a Zeiss Plan Fluar 100x objective with an NA of 1.45. We see the characteristic dips arising from the excitation of SPs, note, however, that the distributions obtained with the widefield system differ from those obtained with point scanning methods such as that of Tanaka and Yamamoto (2003), because the back focal plane distribution is now recording information from an extended area on the sample surface rather than a single diffraction limited spot. The use of an extended incoherent source has removed the speckle arefacts visible in their back focal plane distributions. It is indeed very difficult to remove such artefacts when a spatially coherent source is used. Figure 14.25b and c show the back focal plane distributions obtained from the grating sample shown diagrammatically in Fig. 14.25d. Note that several periods ($\approx 7 - 10$) of the grating are illuminated. Figure 14.25b is taken with the grating vector oriented parallel to the direction of maximum p-polarisation (that is across the grating). We see that there is character-

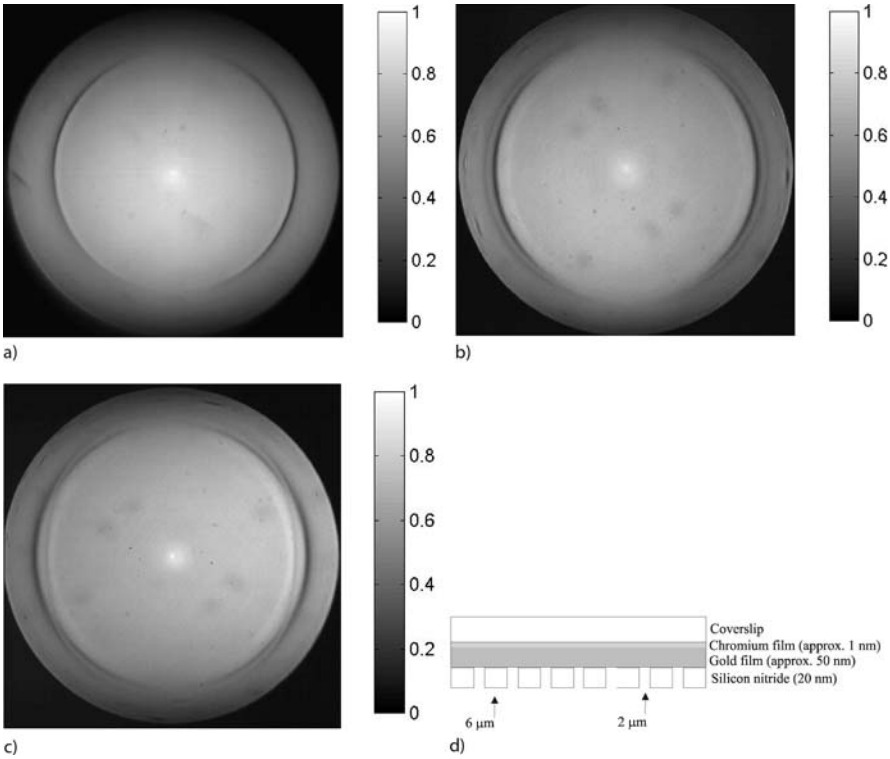


Fig. 14.25. Back focal plane distributions with air backing obtained with widefield system of Fig. 14.24. **a** Plane gold sample. **b** Grating sample of Fig. 14.25d oriented with grating vector parallel to incident polarizaton in the back focal plane. **c** Grating sample of Fig. 14.25d oriented with grating vector perpendicular to incident polarizaton in the back focal plane. **d** Schematic diagram of grating sample

istic dip in the back focal plane corresponding to excitation of the surface wave[5]. We can also observe a second dark crescent corresponding to a diffracted order. Measurements show (Stabler *et al.* 2004) that the diffraction crescent corresponds to the expected position to within a few percent. An important point to note from this figure is that the diffraction effects are only observed around the SP resonance, indicating much stronger potential image contrast from SPs compared to the background. Figure 14.25b shows the back focal plane distribution that occurs when the grating vector is perpendicular to the maximum *p*-polarisation. The pattern is more complicated but again shows strong diffraction around the plasmon angle.

The back focal plane distributions are therefore strongly indicative that SP imaging can be achieved with a wide field microscope configuration. If one simply tries

[5] In fact measurement of the position of the dip shows that it corresponds to the position expected for gold layer with a uniform 15 nm coating of silicon nitride. This is the mean coverage of the layer (6 μm of 20 nm thick gold, and 2 mm bare).

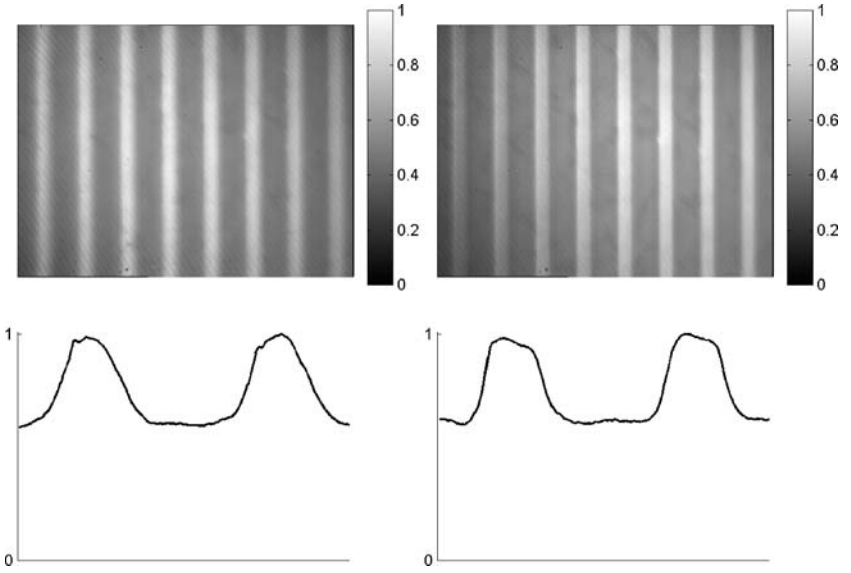


Fig. 14.26. Images of grating sample shown in Fig. 14.25d. **a** Image with grating vector parallel to incident polarisation direction. **b** Image with grating vector perpendicular to incident polarisation direction. Image width 65 microns

to image, however, there is virtually no contrast. This arises from the fact that there is a great deal of background from other angles of incidence which do not contribute to the image contrast. The obvious solution is to use an annular mask conjugate with the back focal plane which allows the only light incident at or close to the angle for excitation of surface waves to illuminate the sample. This blocks light that does not contribute to the image contrast. In our experiments we have used a physical mask and also a spatial light modulator, the former has excellent contrast, whereas the latter allows for the preferred angles to be varied in a very convenient manner. Note that in these experiments no mask was inserted in the detection arm so that the full aperture was available to detect scattered light. The mask used allowed light incident between 45 degrees and 52 degrees to illuminate the sample. The images of Fig. 14.26 show the grating structure when the grating vector lies parallel and normal to the principal direction of SP propagation; that is the direction of maximum *p*-polarisation. From the height of the transition we estimate that the resolution along the direction of SP propagation is approximately 1.3 μm and 0.93 μm normal to this direction. The lateral resolution is clearly still somewhat dependent on the propagation direction, but the range of azimuthal angles has reduced this effect. The microscope objective, rather like the, multiple detectors described by Tanaka and Yamamoto, acts to collect light scattered in all directions but in rather more convenient manner.

The contrast in the widefield SP images can be easily controlled by varying the incident angles on the sample surface. This is most conveniently done with a spatial

light modulator, however, in order to ensure that the unwanted light was properly filtered a mechanical mask was used. In these experiments the annular mask was replaced with a narrow slit, whose width is approximately 45 μm, thus covering a range of angles of approximately 0.5 degrees in the back focal plane. The images obtained are shown in Fig. 14.27 are stamped protein gratings of thickness 5 nm and width 5 μm. When the incident angle is less than the SP resonance angle, θ_p, the protein grating which has a higher refractive index than the backing medium (in this case air) appears bright compared to the uncoated region, as expected from the different reflection coefficients for the coated and uncoated material (Fig. 14.27a). Figure 14.27b shows the situation where the mask is placed so the predominant incident angles are greater than θ_p, in this case we see that the coated region appears dark, albeit with much poorer contrast, again this is what we would expect as the gradient of the reflectivity curve is less steep above θ_p than below it. Figure 14.27c shows a case of deliberate misalignment where the angle of incidence varies across the image. This was achieved by misaligning the system so that the source was not positioned at the true back focal plane. This means that the sample is illuminated with a slightly diverging beam, so that the angle of incidence varies across the sample. We can see that there is a division across the image, where the image of the grating disappears. On either side of this region the image contrast is inverted. These experiments were taken with water as the backing medium. This demonstration is in some ways easier with water backing because the width of the resonance is wider.

The images presented in Fig. 14.27 show that imaging with water backing is indeed possible with widefield SP microscopy. Figure 14.22 shows back focal plane distributions obtained the 'split' SIL objective, the relative weakness of the grating structure means that the sidebands are not visible as in Fig. 14.25a and b where the grating structure is very strong. The images corresponding to the back focal plane distributions are shown in Fig. 14.28. Figure 14.28a shows the protein grating structure with good contrast and resolution, and the second figure shows that good contrast can be obtained with water backing (Fig. 14.28b). The extremely high numerical aperture of the SIL arrangement or the 1.65 NA Olympus is necessary for water imaging as no really satisfactory SP images[6] with water backing have been obtained with even a 1.45 NA objective. Imaging in aqueous media is a crucial stage in developing practical SP microscopy and work is presently underway to apply SP imaging to study binding events in real time. It is worth pointing out the aqueous imaging with the prism based Kretschmann configuration is extremely challenging because oblique illumination introduces very severe aberrations.

Dark field imaging (Zhang *et al.* 2006) may be achieved by modifying the bright field system by using a second mask in the detection arm, marked 'MD' in Fig. 14.24. This ensures that only diffracted and scattered light, is recorded in the imaging system. The dark field operation ensures that only a scattering contrast mechanism operates. Figure 14.29a and b show bright and dark field images of the same region respectively of stripes of a 5 nm thick protein layer deposited on a gold coverslip and imaged with solid immersion lens system described in Sect. 14.5. The

[6] Unless additional coatings are used to reduce the angle for surface wave excitation.

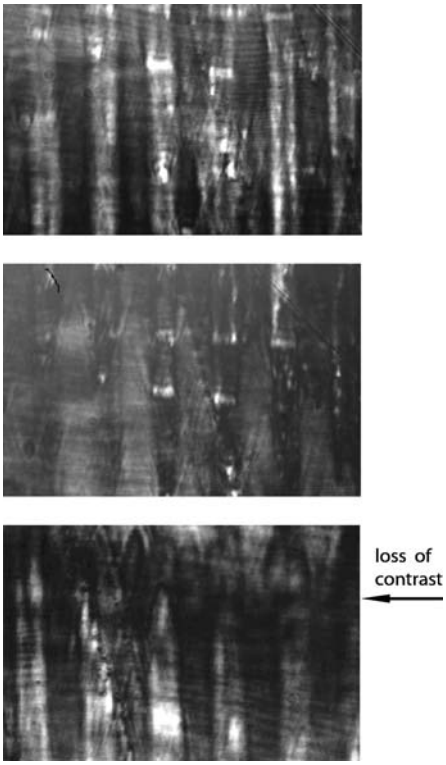


Fig. 14.27. Widefield surface plasmon images of protein monolayer grating sample. **a** Incident angle less than plasmon resonance angle, so protein layer appear bright. Air backing. **b** Incident angle greater than plasmon resonance angle, so protein layer appear dark. Air backing. **c** Variable angle of incidence across sample, with contrast inversion. Water backing. Image width 60 microns

width of the protein stripes is 5 μm. Images were also taken with complementary techniques: conventional bright field microscopy, differential interference contrast (DIC) microscopy and atomic force microscopy (AFM). The images obtained with the complementary optical techniques, as well as having poorer contrast compared with SP microscopy, involve illumination from above the sample whereas the SP images are taken through the coverslip, clearly a major advantage for biological imaging. The AFM images confirm the thickness of the protein layer to be 5 nm. Both the bright and dark field SP images show excellent contrast compared to the other optical techniques. We note the texture of the images, the dark field image is highly sensitive to edges as we might expect as the zero order is blocked. There are also some features that are visible in the dark field image that cannot be seen in the bright field. The arrows in Fig. 14.29a and b indicate a region where the protein coverage is rather poor; it appears below the limit of detectability in the bright field image but can still be observed in the dark field.

14.6.3 Scanning Fluorescence Surface Wave Microscopy

One of the great advantages of surface wave and SP microscopies lies in the ability to perform label-free detection. Despite this the large field enhancement makes this

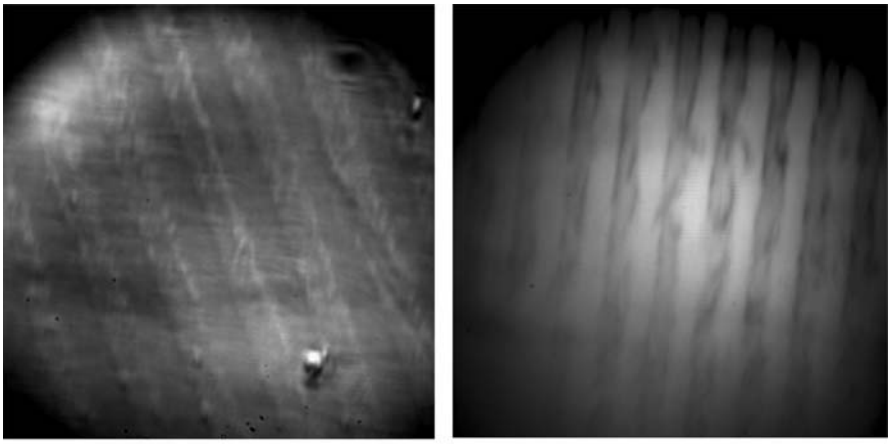


Fig. 14.28. Images of protein grating in air and water. **a** Air backing: picture width approximately 60 microns. **b** Water backing: picture width approximately 90 microns

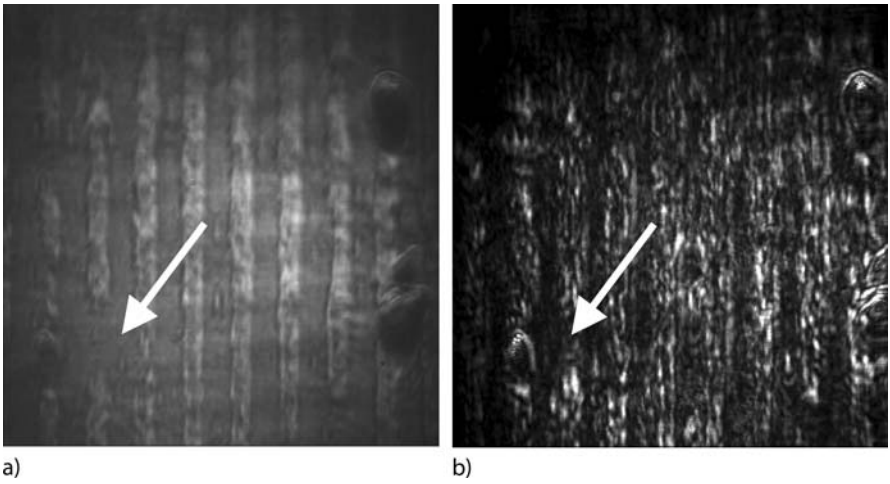


Fig. 14.29. Bright and dark field surface plasmon images of protein grating. Picture width 90 microns. **a** Bright field image. **b** Dark field image. Arrowed region on images shows region of poor coverage showing greater sensitivity of dark field system

technique promising for fluorescence applications especially those involving two photon absorption where the local field enhancement will be raised to the fourth power. Consider the system shown in Fig. 14.30, this is essentially a scanning two photon microscope with beam paths that allow one to insert optical components to modify the beam profile. Quite clearly the key element in such a system is a structure that supports surface waves. The coverslip may have a gold layer deposited as discussed earlier in the chapter, in our experiments, however, we used the struc-

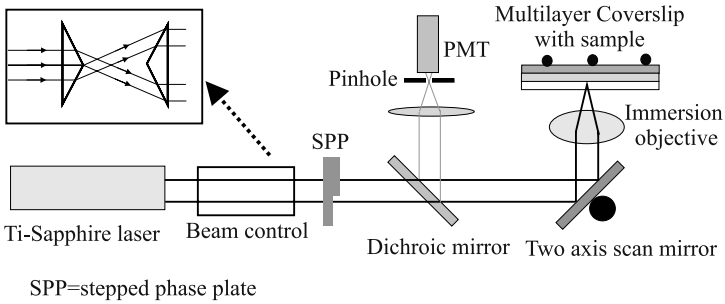


Fig. 14.30. Schematic diagram of two photon fluorescence surface wave microscope

ture shown, whose reflectivity is shown in Figures 14.4 and 14.5. This surface wave structure has two main advantages compared to a gold layer supporting SPs. Firstly, it is transparent to the emitted radiation, thus allowing efficient single ended operation. Secondly, gold without an additional layer of dielectric can quench fluorescence. Even though it is possible to get measurable fluorescence through a gold layer (Stefani *et al.* 2005) these reasons mean a surface wave structure is preferred.

In Fig. 14.4 we showed the reflection characteristics of the surface wave structure used in our experiments. Figure 14.31 shows the amplitude and phase of the transmission coefficient for incident *p*-polarized light. We observe a large a peak in the amplitude of the transmission coefficient whose value is approximately 8.5, which will give a two photon excitation approximately 5000 times larger than for a plane wave incident at the same angle. There is also a phase shift in the transmission coefficient indicating lateral transport of energy, note that observed phase shift only changes through 180 degrees rather than the 360 degrees observed in reflection; this is illustrated clearly in Fig. 14.12. This arises simply because there is no 'specular' component in the transmission coefficient only a contribution due to the surface wave whose phase only changes by 180 degrees.

Modelling the response of the two photon microscope surface wave has been described in (Somekh 2002). Briefly this involves a vector diffraction analysis of the field interacting with the fluorophore. The field distribution at the focus of a high nu-

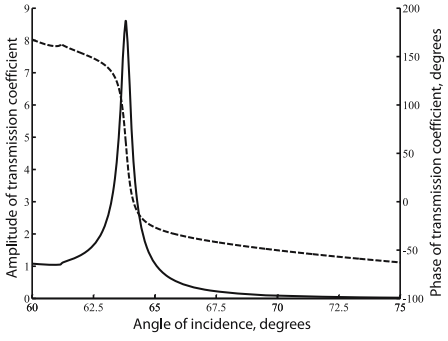


Fig. 14.31. Calculate reflection coefficient for *p*-polarized light for guided waves structure described in Fig. 14.4. *Solid curve* (amplitude), phase (*dashed*)

merical aperture objective, taking account of the vectorial nature of the waves, can be calculated following the approach presented by many authors, for instance, Hopkins (1943) and Richards and Wolf (1959). More recently, Török *et al.* (1995) have extended this approach to account for focusing through media of different refractive indices.

In essence, each point of illumination in the back focal plane is regarded as a point source which gives rise to a plane wave of appropriate polarisation state and angle of incidence depending on the azimuthal angle and radial position respectively.

This leads to the following expression for the field distribution at the $z = 0$ plane corresponding to the interface between n_1 and n_2.

$$E_n = \int_0^{2\pi} \int_0^{\theta_{\max}} A(\theta, \phi) f_n(\theta, \phi) \exp in_1 k(x \sin\theta \cos\phi + y \sin\theta \sin\phi + z \cos\theta) \mathrm{d}\theta \mathrm{d}\phi \tag{14.6}$$

where E_m denotes the x, y, z components of the electric field vector.

$$\begin{aligned} &\text{For } E_x \quad f_x(\theta, \phi) = \cos^2\phi \cos\theta \tau_p(\theta) + \sin^2\phi \tau_s(\theta) \\ &\text{For } E_y \quad f_y(\theta, \phi) = \cos\phi \sin\phi [\cos\theta \tau_p(\theta) - \tau_s(\theta)] \\ &\text{For } E_z \quad f_z(\theta, \phi) = -\cos\phi \sin\theta \tau_p(\theta) \end{aligned}$$

Where k is the wavevector in free space, $A(\theta, \phi)$ represents the source of excitation in the back focal plane, for an aplanatic system an additional factor of $\cos^{1/2}\theta$ should be incorporated to account for the change in area subtended by an element of the incident beam as it changes direction passing through the objective. Other polarisation states are readily incorporated using the Jones matrix on the distribution in the back focal plane. $\tau_p(\theta)$ and $\tau_s(\theta)$ represent the transmission coefficients for p- and s-polarisations.

Equation (14.6) is similar to that developed by Török *et al.* (1995), there are certain issues that need clarifying to appreciate how the expression is used for the problem of SP and surface wave imaging imaging. Imaging through a stratified medium, as addressed by Török *et al.* (1995) is of considerable importance in conventional two photon microscopy, since the aberration of the focused beam as the light passes through regions of different refractive index has a powerful influence on the field distribution and the resulting fluorescent yield. Our problem looks formally similar; the key physical difference here is that the layer through which the light travels prior to reaching the sample is extremely thin compared to the optical wavelength, so aberrations from this source are less important (although accounted for). The form of the transmission coefficients $\tau_s(\theta)$ and, particularly $\tau_p(\theta)$ are crucial in our case and they have a different meaning from the interface transmission coefficients as described in the matrix formulation of Török *et al.* (1995). In Eq. (14.6) the coefficients are the total electric field transmission coefficients for the incident beam to the interface between the last layer and the semi-infinite backing (n_4 in Fig. 14.1b),

thus accounting for multiple reflections at the interfaces as well as the propagation within the layers.

Once the field distribution is calculated the fluorescent yield will be proportional to the sum of the squares of the field raised to the second power for two photon excitation. If the fluorophores have a preferred dipole axis is, of course, necessary to resolve the field along the dipole axis can be calculated by raising the component of the field the fourth power. The effects on the point spread function for two photon imaging with different polarisation models are discussed by Higdon *et al.* (1999). Calculated point spread functions have been presented in Somekh (2002) and rather than repeat these we will show experimental results which bear out the predictions.

The effect of the pinhole is naturally incorporated by multiplying the response of non-confocal response by the point spread function associated with the pinhole, this confers a considerable improvement for two photon surface wave microscopy because the point spread function operates at the emission wave length. This effect is borne out in the experimental results shown in the following section.

Point spread functions of the surface microscope were obtained using fluorescent beads (Molecular Probes, PS-Speck 175 ± 5 nm diameter, 505 nm excitation/515 nm emission) attached to thin layer of polylysine attached to the modified layered surface wave coverslips. An annular distribution of linearly polarised light is input onto the back focal plane. This was achieved using either with the double axicon arrangement shown in the inset or a simple annular mask. In principle the double axicon concentrated all the optical power into the annular ring, although in practice imperfections in the components meant that it did not give the improvement expected. The point spread function without the stepped phase plate is shown in Fig. 14.32, where one can see a double-peaked distribution, this is discussed in Kano *et al.* 1998 and is modelled in Somekh (2002). This arises from the fact that the axial component of the light is dominant for the surface waves produced and the axial components cancel on the optical axis, as shown in Fig. 14.33a. A phase plate in the back focal plane in which the two semi-circles of the distribution have a phase shift relative phase shift of 180 degrees imposed on them means that the axial com-

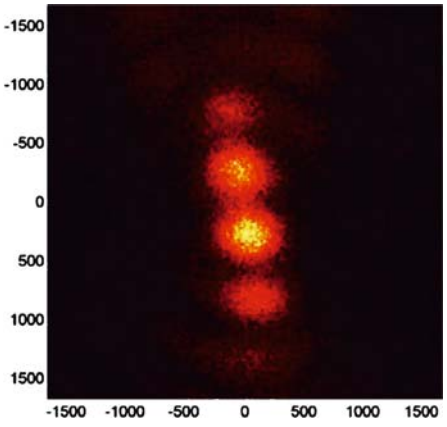


Fig. 14.32. Measured two-photon PSF without stepped phase plate and pinhole (distances in nanometres)

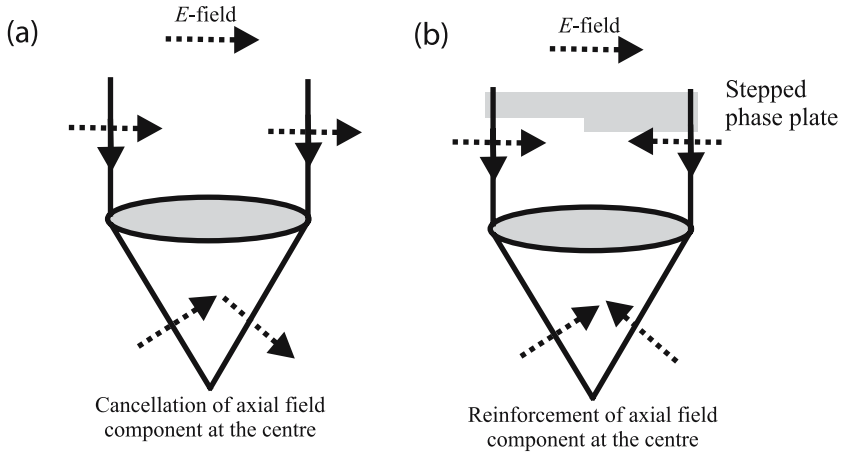


Fig. 14.33. Schematic daigram showing axial components of field distribtion. **a** Cancellation of axial components without phase plate, **b** Reinforcement of axial components with phase plate

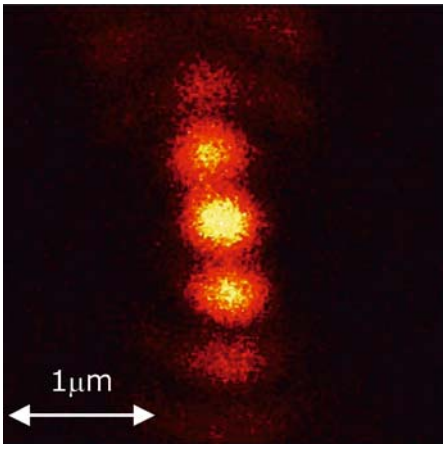


Fig. 14.34. Measured two-photon PSF with stepped phase plate and without pinhole (distances in nanometres), showing single peaked distribution

ponents reinforce (Fig. 14.33b), giving a single mainlobe and side lobes as shown in Fig. 14.34. The stepped phase plate used in these experiments was a step in a glass slide which was designed to give 180 degrees phase shift at 800 nm wavelength, since the wavelength used in these experiments was 925 nm the actual phase shift was reduced accordingly. This did not appear to have a seriously deleterious effect on the observed point spread function. We did, however, carry out other experiments to verify the operation of the phase plate in which the deviation from the perfect 180 degree phase shift appeared more critical. In these experiments the phase plate was rotated by 90 degrees from the 'correct' position. This imposed a symmetry where the phase flips as in successive quadrants which results in nulls along both x- and y-directions. The experimental distribution is shown in Fig. 14.35a. Simulation, us-

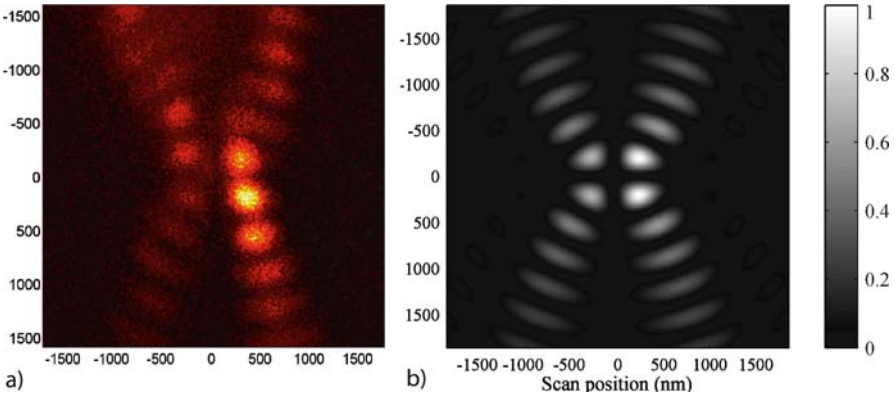


Fig. 14.35. Measured two-photon PSF with crossed phase plate. **a** Experimental distribution. **b** Calculated distribution allowing for 156 degree phase shift by phase plate

ing the formulation of Eq. (14.6) of this situation with a perfect 180 degree phase shift shows a perfectly symmetrical 'X' structure. We note that in the experimental curves that there is a considerable difference in brightness between the quadrants, Fig. 14.35b is a simulation where the phase shift imposed by the plate is 156 degrees corresponding to the longer wavelength used in our experiments. We see that the symmetry is broken. It is also important to note that the simulation represents the square of the axial field whereas two photon excitation depends on the fourth power, so that any asymmetry will be enhanced in the experimental measurements. Clearly misplacing the stepped phase plate is not useful for imaging, it does, however, indicate that the phase plate performs the task for which it was designed.

It is clear that inclusion of a pinhole will greatly improved the imaging performance, firstly reducing the sidelobes and narrowing the width of the mainlobe. Figure 14.36 shows the experimental point spread function obtained with a pinhole diameter of 20 microns. This corresponds to a diameter equal to approximately 1 Airy spot diameter of the emitted fluorescence signal centered around 515 nm.

Figure 14.34 shows a point-spread function with a FWHM of 264 nm along its narrowest dimension, implying a resolution of 132 nm. The introduction of a pinhole further sharpens the central lobe of the distribution to give a FWHM of 221 nm. There is a dramatic improvement in image quality because the sidebands are substantially reduced.

Figure 14.37 is a non-confocal image of showing regions of enhanced fluorescence on artificial membranes. These regions are cholestrol rich regions observable in both artificial membranes and cell membranes and are thought to play an important role in cell signalling.

One of the most interesting features of the response of the surface wave microscope is its response to defocus, this is a feature that we will bring out in more detail when considering the interferometric surface wave microscope in Sect. 14.7. Figure 14.38 shows the variation of fluorescence as a function of defocus (without

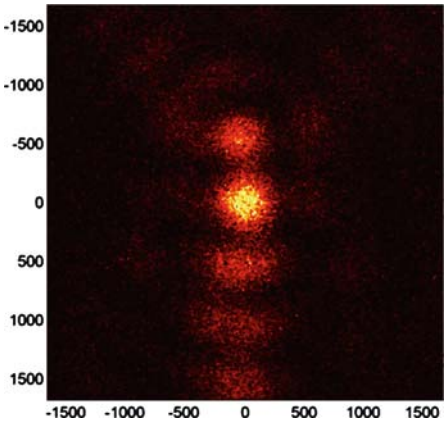


Fig. 14.36. Point spread function obtained with a 1 Airy diameter (fluorescence emission) pinhole

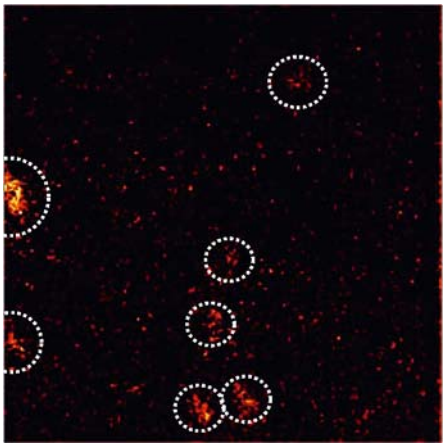


Fig. 14.37. Surface wave image of artificial lipid bilayer (50% PC, 50% cholesterol) exhibiting lipid rafts (*circled*). Image area is $20 \times 20\,\mu m$

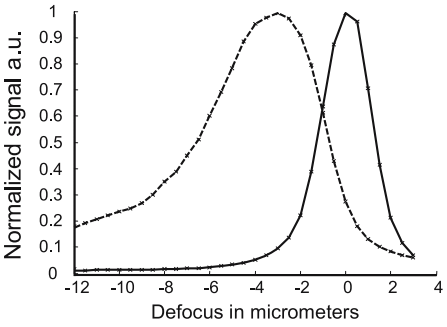


Fig. 14.38. Experimental conventional two-photon (*dashed*) and surface-wave response (*solid*) to defocus

the pinhole in place). The in-focus position was determined using the microscope in a conventional two-photon imaging mode, with a conventional beam expander replacing the axicon arrangement, an image stack was then acquired at differing levels of defocus. The parts of the illumination not contributing to the surface waves were then eliminated and the defocus stack repeated. We observe that the maximum signal

is observed away from the focal position, as predicted in Somekh, 2002. This arises from the more efficient excitation of surface waves on the sample, as the sample is defocused a phase gradient is imposed on the incident beam, at negative defocus this more closely matches the *k*-vector of the surface wave so that the individual point sources add in phase and propagate towards the optical axis. This is the same effect that was discussed in Sect. 14.3, where the surface wave contribution reaches a maximum when the incident wave number matches that of the surface wave. As expected on the other side of defocus the signal falls off very rapidly, and in this case propagation away from the optical axis is favoured. We note also that the width of the point spread function does not degrade noticeably as a function of defocus because the fluorescent emission arises from the region where the surface waves come to a focus. The results presented in Fig. 14.38 have been predicted using Eq. (14.6), and very good agreement is obtained when a small amount of loss is introduced in the layers, this presumably accounts for both ohmic loss and scattering. The behavior with respect to defocus is quite similar to the interference microscope discussed in Sect. 14.7.

The fact that the surface waves can propagate through a considerable distance prior to reaching the optical axis does not concern us particularly with two photon excitation since the emitted light is proportional to the square of the local intensity; the effects of aberration prior to propagation do not appear to significantly degrade the images. Another point that is well worth noting is that effective depth of field of this microscope is very large indeed. This arises from the fact that once the surface waves are excited they will come to a focus regardless of where they are excited (provided of course they are not scattered or attenuated). In essence, the surface wave microscope is rather like the focusing of a Bessel beam (Durnin *et al.*, 1987), where the range of axial spatial frequencies is very small. Effectively, a Bessel beam is formed when there is only one axial spatial frequency present, variation around the azimuth will lead to different beam profiles some will have nulls at the center and others will have maxima. The key feature of only one axial spatial frequency means that as the sample is defocused each azimuthal component will acquire the same additional phase shift, so the *relative* phase shift between the different components will be unaltered, thus accounting for the huge depth of field. For surface wave microscopy the surface waves are excited at a single axial spatial frequency so defocusing the sample will not affect their relative phase, this means that the width of the surface wave focus will be insensitive to defocus. The situation however differs from a Bessel beam focusing in free space in one important respect; the strength of the waves propagating towards the axis will increase as the sample is moved from the focus of the object. From a practical point of view this enormous depth of field is very useful. At the simplest level it means that the quality of the optics is not very critical, symmetric aberrations do not present a problem. More importantly it greatly relaxes the problems of aligning two tightly focused beams since it reduces the problem to a plane rather than a three dimensional focus. This would be a very significant advantage in microscope schemes involving two beams, such as, for instance, STED (Dyba *et al.*, 2003).

14.7 Objective Lens Interferometric Techniques

14.7.1 Scanning Interferometry

This subsection describes the use of interferometry for SP microscopy. Clearly an interferometer gives one the opportunity to measure the phase of the signal, however, by the far most significant advantage of interferometry lies in the confocal imaging properties, which allows us to obtain far better spatial resolution compared to non-interferometric methods.

Consider a scanning heterodyne interferometer such as the one shown in Fig. 14.39. The interference signal is produced by interfering the reference beam with the signal emerging from the objective lens. This has the same transfer function as the confocal microscope, this is explained in detail in Somekh, 1993 and references therein. Briefly, we can see that the interference signal reaches a maximum value when the beam returning from the reference beam and the objective are parallel to each other. When the sample is defocused the beam from the objective will become curved and there will be a reduction in the interference signal. Mathematically, this is identical to a confocal pinhole[7] and indeed the reference beam is often referred to a 'virtual' pinhole.

Figure 14.39 shows a schematic diagram of the actual scanning heterodyne interferometer used in the experiments. The interferometer was illuminated with a 633 nm HeNe laser, one arm of the interferometer passed through an oil immersion objective (Zeiss CP-Achromat 100x/1.25oil), the other arm was formed by reflection from a reference mirror. Two Bragg cells, driven at frequencies differing by 10kHz,

[7] The interference signal is proportional to the field, so the squared modulus is the same as the output from a confocal microscope sensitive only to intensity.

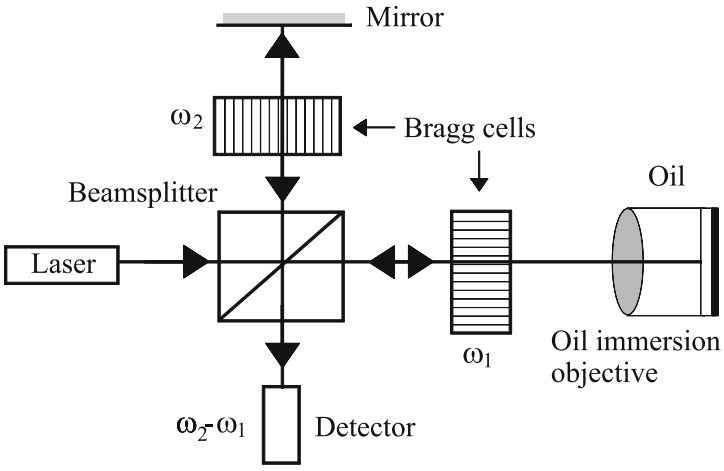


Fig. 14.39. Heterodyne interferometer configuration used for $V(z)$ measurements and high resolution surface plasmon imaging

were used to effect the frequency shifting, so that the interference signal was detected at the beat frequency between the sample and reference arms. The use of two Bragg cells not only allows the interference signal to be detected at a very convenient frequency but also reduces the effects of spurious reflections in the Bragg cells (Migdall *et al.*, 1990). The samples were mounted in an x, y, z piezo-scanning stage with resolution of 10 nm. The scan stages were controlled by a PC and the output from the heterodyne interferometer was converted to an amplitude and phase signal in a lock-in amplifier, which was, in turn, sent to the PC for display. One advantage of the heterodyne system is that the heterodyne signal contains the interference term alone and no other processing is necessary as is the case with, say, phase stepping techniques. This is one of the advantages compared to widefield interometric techniques discussed later in this section.

We now consider how we might expect an interferometer to behave in the presence of a sample that supports SPs. The contrast of the SP microscope can be explained by the so-called $V(z)$ effect, where the output response, V, varies as a function of defocus, z. This effect has been used to evaluate the pupil function of objective lenses and to measure aberrations, Zhou *et al.* (1997). The effect is also very important in scanning acoustic microscopy where surface acoustic (Rayleigh) waves are excited by a large numerical aperture acoustic lens (Atalar, 1978, Parmon and Bertoni, 1979). The resulting phase difference between ray paths corresponding to excitation and reradiation of surface waves and rays reflected normally from the sample provides a means of accurately measuring wave velocities. This effect also provides a powerful means to explain image contrast in the acoustic microscope, Ilett *et al.* (1984), Somekh *et al.* (1985). We will show that the situation when imaging structures that support SPs is analogous to the acoustic case, although there are important differences, in particular, with respect to the polarisation of the incident beam.

In order to calculate the $V(z)$ response of a heterodyne interferometer we first need to calculate the field reflected to the back focal plane of the microscope objective, see Fig. 14.18. Consider linearly polarised light incident on the back focal plane of the microscope objective. The field amplitude passing through the objective is assumed to vary radially as P(r), where the pupil function may include amplitude and phase variations (introduced by, for example, aberrations). For simplicity we normalise the aperture of the objective to unity. As discussed in Sect. 14.5 a point excitation in the back focal plane can be considered to be transformed to a plane wave in a specific direction and polarisation state after passing through the objective.

Consider a plane wave emerging from a point r, ϕ, in the back focal plane. The field reflected back along the original polarisation direction due to this source will be given by

$$E_{\text{reflected}} = P^2(\sin\theta)\left[R_p(\sin\theta)\cos^2\phi + R_s(\sin\theta)\sin^2\phi\right]\exp(j2n_1k\cos\theta z) \quad (14.7)$$

The pupil function appears squared since the light passes through the objective twice. $R_p(\sin\theta)$ and $R_s(\sin\theta)$ refer to the Fresnel reflection coefficients for p- and s-polarisations. The terms $\cos^2(\phi)$ and $\sin^2(\phi)$ refer to the fact that the proportion of

each incident polarisation changes with azimuthal angle. The *p*-incident light is the component resolved radially and the *s*-incident light is resolved along the tangent. The interference signal arises from the component of the field resolved along the direction of the original polarisation direction; so it is necessary to resolve each reflected component a second time. The final term in the exponential refers to the phase shift due to defocus, z, k is the wavenumber in free space ($2\pi/\lambda_{\text{free}}$), where λ_{free} is the free space wavelength. Negative z refers to movement of the sample towards the objective.

The field reflected in the back focal plane is interfered with a frequency shifted reference beam E_{o} so that the frequency shifted output from the heterodyne interferometer is given by:

$$2Re\left(E_{\text{reflected}}E_o^*\right)$$

The instantaneous interference signal from the heterodyne interferometer is a time varying signal proportional to the instantaneous change in intensity, and is thus real. The amplitude and phase of this signal can, of course, be represented as a complex phasor. We can thus represent $V(z)$, omitting constant terms and assuming a uniform reference beam as:

$$V(z) = \iint_{\text{lens aperture}} P^2(\sin\theta)\left[R_p(\sin\theta)\cos^2\phi + R_s(\sin\theta)\sin^2\phi\right] \times \exp(j2n_1 k\cos\theta z)\mathrm{d}(\sin)\theta)\mathrm{d}\phi \tag{14.8}$$

The interesting thing about this expression is that, as we will see, the modulus of the $V(z)$, $|V(z)|$, gives a considerable amount of information about the behavior of the SPs. Despite this it is still necessary to detect the modulus of the *interferometer* output in order to get this information. Moreover, it is the phase behavior of the reflectance functions that give $V(z)$ its interesting properties. In all the experimental and theoretical $V(z)$ curves presented in this article we only present $|V(z)|$; since the phase is not used modulus signs are omitted for brevity.

Figure 14.40 shows a theoretical $V(z)$ curves calculated using Eq. (14.8). This curve is formed from contributions from and R_p and R_s respectively, we note that excitation of SPs represented by the phase change in R_p induces a periodic ripple at negative defocus, whose periodicity depends on θ_p, whereas no such structure is visible in the $V(z)$ formed from R_s. For a full aperture objective there will be equal contributions from *s*- and *p*-incident polarisations, which still gives the SP induced ripple shown albeit with reduced relative amplitude compared with pure *p*-incidence.

When the sample is defocused curvature is introduced to the reflected wavefront; the ripple arises from the deformation of the curved wavefront by the phase change in the reflection coefficient. This means that there are two principal components corresponding to the output signal: the reflection close to normal incidence and the reflection around θ_p. This effect has been explained for the case of the scanning acoustic microscope by Atalar (1979). These two contributions may be represented by ray paths A and B (Fig. 14.41); as the defocus changes the relative path length

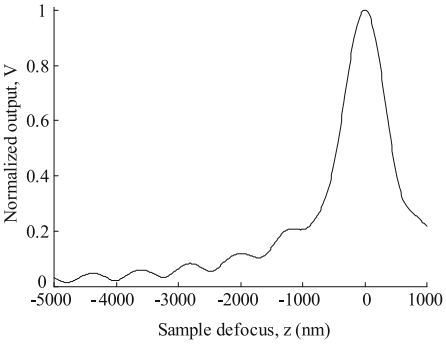


Fig. 14.40. Theoretical V(z) curve corresponding to reflectance functions shown in Figs. 14.2 and 14.3. Note the periodic ripple at negative defocus

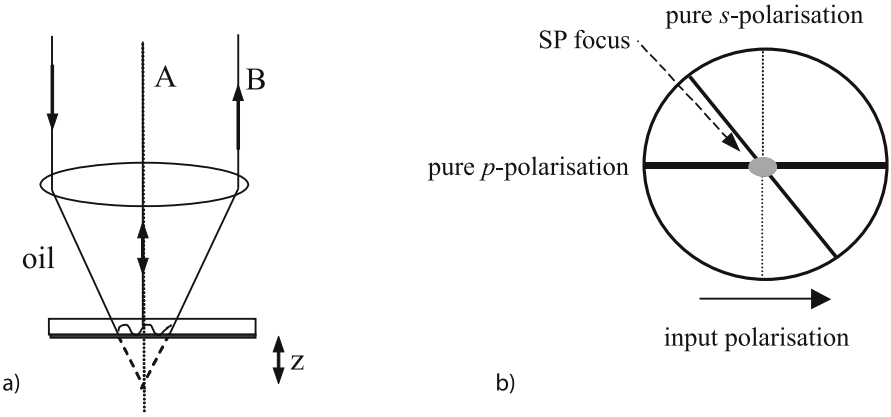


Fig. 14.41. Ray paths showing principal contributions contributing the ripple in the $V(z)$ curve. **a** Rays in plane of incidence showing normally reflected contribution (**A**) and surface plasmon contribution (**B**). **b** Schematic of sample showing surface plasmons excited over a ring, from which they can focus to the optical axis

between the rays changes. Ray paths A and B do not interfere with each other directly because they return in different positions, however, they each interfere with a common reference beam (Offside *et al.* 1989). This means that the strength of the interference signal will oscillate periodically, going through a complete cycle when the phase difference between A and B changes by 2π. The periodicity, Δz of the oscillations in the $|V(z)|$ signal can be obtained by noting that a defocus, z, introduces a phase shift in the wave travelling along the optical axis of $2kn_1z$; the phase difference in the wave exciting the SP is $2kn_1 \cos\theta$; Δz is obtained by setting the difference in these two phase shifts equal to 2π. The periodicity, Δz, is thus given by:

$$\Delta z = \frac{\lambda_{\text{free}}}{2n_1(1 - \cos\theta_p)} \tag{14.9}$$

This expression is similar to the analogous expression for the $V(z)$ periodicity in the acoustic microscope.

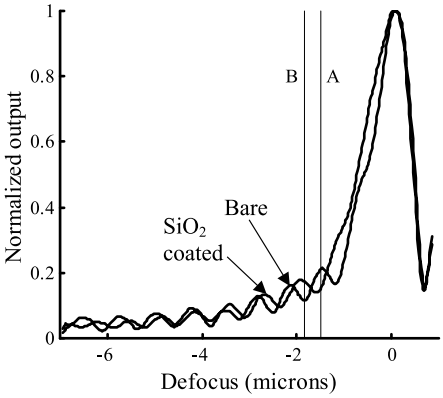


Fig. 14.42. Experimental $V(z)$ curves obtained on samples coated with gold (*bare*) and gold with an additional 20 nm layer of SiO_2 (*coated*)

Figure 14.42 shows $V(z)$ curves obtained on a coverslip coated with 44 nm of gold and a similar sample on which a further layer of 20 nm of silicon dioxide was deposited on the gold. A very thin layer of approximately 1 nm of chromium was deposited between the gold and the coverslip to improve adhesion. The presence of the dielectric perturbs θ_p, which, in turn, perturbs the periodicity of the ripple. The measured periodicity of the ripple for the bare metal film was 761 nm, which is extremely close to the value of 758 nm predicted from Eq. (14.9). The periodicity obtained on the dielectric coated sample was 676 nm compared to 672 nm predicted from a metal layer coated with 20 nm of silicon dioxide. The experiments indicate that the presence of the dielectric increase θ_p from 43.6 degrees to 46.4 degrees with a corresponding increase in the k vector of the excited SP.

We have also inserted apertures in the back focal plane of the illumination optics to vary the proportion of *s*- and *p*-polarisations. These show that, as expected, when a narrow aperture is placed normal to the direction of the incident polarisation, most of the *p*-polarised light is suppressed and as expected the ripples in the $V(z)$ curve are substantially suppressed.

We now turn our attention to how the $V(z)$ curve can be related to imaging with the heterodyne interferometer. The $V(z)$ curve indicates that at focus there is very little contrast between different samples, but as the sample is defocused the different periodicities of the ripple patterns means that there is contrast between a bare metal film and a dielectric coated metal film. For instance, defocus A on Fig. 14.42, shows a region where the bare region will appear bright compared to the coated region. For larger defocuses the contrast will reverse, at, for instance, defocus B. Although the $V(z)$ curves indicate that the microscope will detect contrast between a bare metal surface and a dielectric coating, they do not give any indication concerning the lateral resolution. When the sample is in focus the wavefront of the returning beam will make a substantial contribution to the output interference signal *except* where the phase changes rapidly over the region where the SPs are excited. For this reason the interferometer is not expected to give much SP contrast in focus. It is thus necessary to defocus the sample to obtain good contrast.

The advantage in terms of contrast of operating with a negative defocus may, at first sight, appear to be a problem with the technique since defocus, of course, results in spreading of the focal spot. The SPs are, however, excited on a ring so that they propagate towards the optical axis where they come to a focus. This indicates that there will be a large increase in intensity over a highly localized diffraction limited region. The diameter of the spot will be small (although not equal) in both horizontal and vertical directions, since the polarisation state, and hence the strength of SP excitation, varies with azimuthal angle. It is this self-focusing that leads to the possibility of high resolution imaging in the defocused state. The fact that the SPs are excited over a ring means that the SP microscope can be thought of as a 2π (radians) microscope by analogy with the 4π (stereoradians) microscope used in 3-D confocal microscopy (Hell *et al.*, 1994).

A series of images were obtained on the structure shown schematically in Fig. 14.25d. A piezo-drive stage was used to scan in increments of 0.02 μm. The images shown in Fig. 14.43a,b and c were obtained by scanning the objective relative to the coverslip. Figure 14.43a was obtained with the sample close to the focus and shows barely discernible contrast as expected from the $V(z)$ curves. Figure 14.43b was obtained at a defocus of −1.5 μm (A on Fig. 14.42) and shows the stripes (bare metal) bright, whereas Fig. 14.18c taken at −1.85 μm (B on Fig. 14.42) shows the stripes dark. Below each of the images horizontal line traces starting 1.6 μm from the top left hand corner of the images are shown. The image corresponding to Fig. 14.43a shows little contrast as expected. The traces corresponding to Fig. 14.43b and c show opposite contrast to each other and, more importantly, the transition across the stripes in both cases is less than 0.5 μm. Since the traces are horizontal and the stripes are not exactly vertical the actual transitions are at least 10% sharper than this value. These results clearly indicate that the resolution

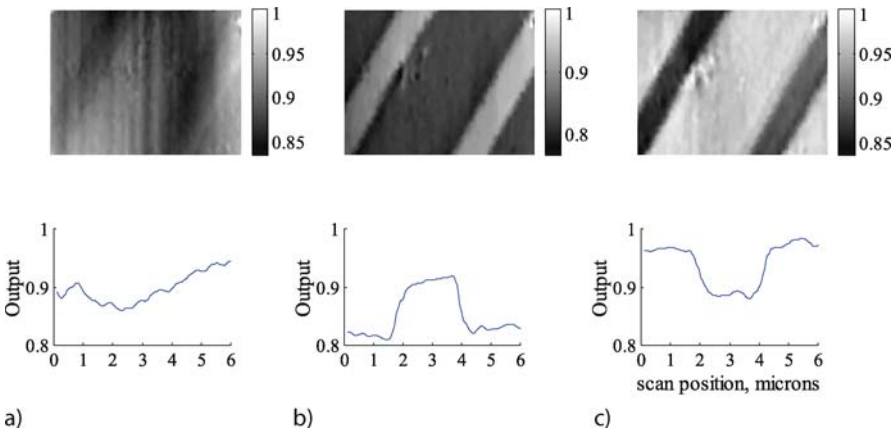


Fig. 14.43. Image of sample shown schematically in Fig. 14.25d. Picture width 10 microns. **a** In focus-little contrast. **b** Defocus −1.5 μm-good contrast bare region appears bright. **c** Defocus −1.85 μm-good contrast coated region appears bright

is limited by diffraction rather than the propagation length of the SPs. These effects are discussed in more detail in Somekh *et al.* (2000a, 2000b).

It is interesting to estimate the sensitivity of the interferometric $V(z)$ measurement to small variations in layer thickness. We consider a heterodyne interferometer limited operating in the shot noise limit. Provided the reference beam is sufficiently powerful the signal to noise ratio can be readily shown to be given by:

$$\mathrm{SNR} = \frac{\eta P}{h\nu B} \tag{14.10}$$

Where η is the quantum efficiency, P is the optical power detected from the sample, h is Planck's constant, ν is the optical frequency and B is the measurement bandwidth. Using values of $\eta = 0.8$, $P = 0.1$ mW and $B = 8$ kHz and a wavelength of 633 nm we obtain a signal to noise ratio of 3.2×10^9. This means that the number of resolvable voltage levels is approximately 5.6×10^4. We now look at the signal change on the $V(z)$ curve, such as that shown in Fig. 14.14, at a fixed defocus corresponding to approximately 1.8 microns as a layer of SiO_2 is deposited onto bare metal. This gives a signal change of approximately 1 part in 500 per nm. This indicates that the $V(z)$ should be capable of resolving changes in layer thickness of less than 10^{-2} nm. Clearly, in practice a system may give a noise performance a few dBs worse than the shot noise limit. On the other hand, using 1 point only of the $V(z)$ to measure a change in layer thickness is rather crude. We are confident therefore that the $V(z)$ technique can be used for measuring small fractions of a monolayer on localised regions.

14.7.2 Widefield Interferometric Techniques

From Subsect. 14.7.1 it is clear that the confocal transfer function confers highly desirable properties when performing SP microscopy. Indeed, the resolution is hardly influenced by the SP propagation length but is now diffraction limited.

It is highly desirable to be able to achieve these properties with widefield operation. Confocal transfer functions can be achieved in a widefield configuration using a speckle interferometer as shown schematically in Fig. 14.24. To construct the interferometer we now incorporate the reference arm shown in the dashed box of Fig. 14.24. The microscope is Köhler illuminated from the ground glass surface of the diffuser. A speckle pattern is then imaged onto the back focal planes of the microscope objective and also the reference objective[8]. The spatial variation in the back focal plane ensures that a wide area of the sample is illuminated. The back focal plane distribution ensures that another speckle pattern illuminates the sample and the reference mirror. These distributions are imaged through the system back to the image plane, where diffraction limited speckle patterns from sample and reference are projected. Figure 14.44 shows the interfering speckle patterns from the

[8] It is not necessary to match the objectives indeed, a low NA objective can be used in the reference arm. All that is necessary is that the returning speckle patterns match. The long temporal coherence greatly relaxes the alignment conditions.

sample and reference beams interfering to form fringes formed by tilting the reference beam, when the diffuser is stationary. Rotating the diffuser produces different speckle patterns which are averaged away in the CCD camera leaving the clean fringes shown in Fig. 14.44b.

In (Somekh *et al.* 2000) we showed that the interference term from the speckle interferometer is a widefield analogue of the confocal output from a heterodyne interferometer. This interference term may be readily extracted using a standard phase stepping algorithm, for this reason the piezoelectric actuator was inserted behind the reference mirror.

In physical terms the confocal response may be understood by regarding the interference terms as the correlation between the reference arm and the sample arm. When the sample is defocused the speckle pattern from the sample decorrelates so that the interference signal decreases. The rate at which the interference signals decorrelate is directly related to the speckle dimension in the axial direction, which is in turn determined by the square of the numerical aperture of the objective (Somekh *et al.*, 2000). In principle, then the speckle interferometer provides a means to get similar performance in a widefield configuration to a scanning one. There is, however, one practical difference related to the dynamic range of the detectors. From the $V(z)$ curves in Fig. 14.40 and 42 we can see that the interference signal level in the region of interest is small compared to the in focus value. If we want to see small changes in the signal then this places even more significant demands on the dynamic range of an optical detector. This presents no serious problem with a photodetector as used in the scanning heterodyne interferometer, where the desired signal and the background are separated. The CCD camera used in the widefield configuration suffers from the large background reducing the smallest resolvable change. The simplest solution is to use a mask in the back focal plane. Following the view that output from interferometer is formed as the phasor sum of two beams A and B of Fig. 14.41 we simply use a mask with small aperture in the middle and annulus to allow the low angles of incidence and the angles around the SP resonance to pass. The intermediate regions do not make a significant contribution to the $V(z)$ response when the sample is above the focus, so the output above

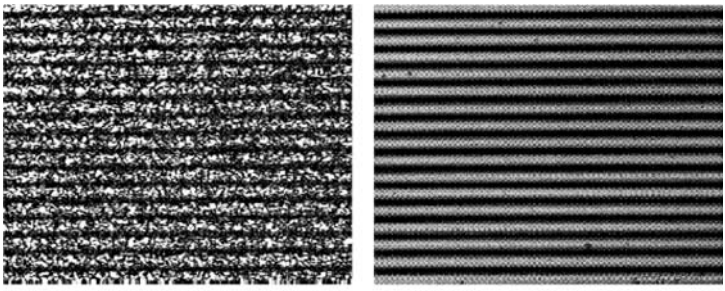


Fig. 14.44. Speckle fringe pattern with obtained with widefield interferometer. **a** Stationary diffuser. **b** Moving diffuser

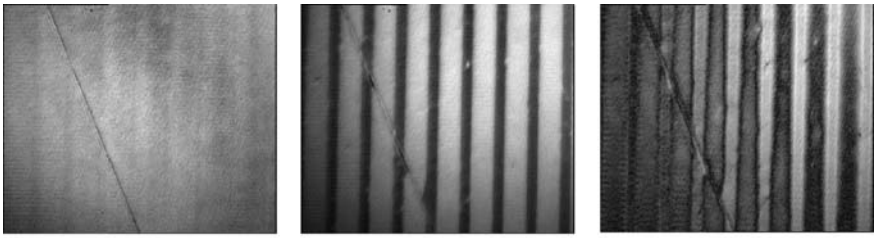


Fig. 14.45. Images with widefield SP inteferometric microscope of sample in Fig. 14.25d. Input polarisation parallel to grating vector. Picture width 65 microns. **a** Reference focus. **b** −2.4 microns from reference. **c** −3.5 microns from reference

focus is relatively unaffected. The signal at focus is, of course, reduced but as shown with the scanning configuration this gives relatively poor image contrast.

The images of Fig. 14.45a,b,c show widefield interferometric images of the grating structure of Fig. 14.25d. These images were obtained with the input polarisation parallel to the grating vector. Figure 14.46a,b and c show three images of the grating taken with the incident polarisation perpendicular to the grating vector. These images were selected from a stack of images taken at different defocuses over approximately 6microns of defocus, in both cases the images show frequent contrast reversals, the contrast, however, differs for each orientation, so for this reason different defocus values were chosen for each orientation. For both orientations, however, the lateral resolution was better than the non-interferometric microscope on account of the confocal operation. The lateral resolution is similar to the scanning configuration as expected. Figure 14.47 shows two $V(z)$ curves extracted from a sequence of images taken with the incident polarisation perpendicular to the grating vector (Fig. 14.46). The solid curve corresponds to the region where the sample is bare and the dotted curve corresponds to the coated region. The $V(z)$ curve is distinctly less beautiful than those obtained with the scanning configuration largely because the annulus affects the region around focus. The observed periodicity corresponds to excitation angles of 47.8 degrees and 46.3 degrees for the coated and uncoated re-

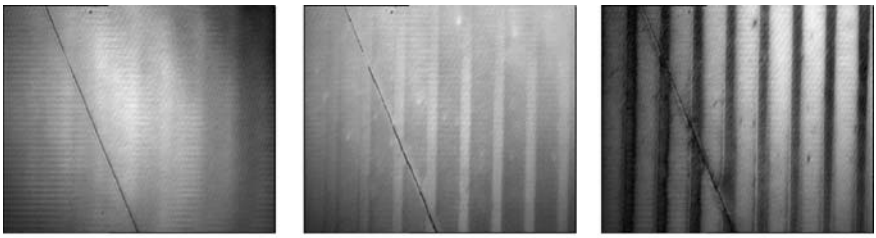


Fig. 14.46. Images with widefield SP inteferometric microscope of sample in Fig. 14.25d. Input polarisation perpendicular to grating vector. Picture width 65 microns. **a** Reference focus. **b**−1.2 microns from reference. **c** −2.4 microns from reference

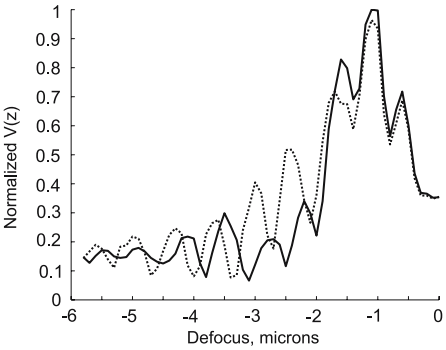


Fig. 14.47. $V(z)$ curves for input polarisation perpendicular to grating vector extracted from image stack. Solid curve on bare uncoated gold, dotted curve on coated region

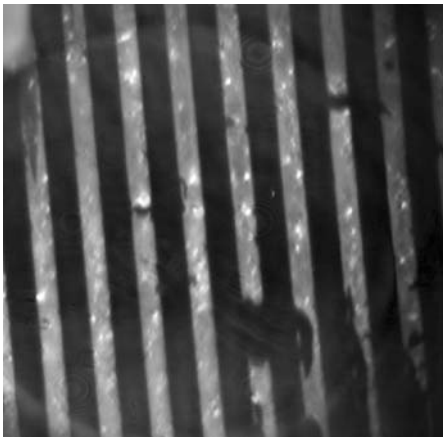


Fig. 14.48. Interferometric microscope image of protein grating, image width 65microns

gions respectively. This is a rather different from the values expected from the bare material but shows the expected difference between the bare gold and the coated region. The curves do indicate the rich contrast available from this form of SP microscopy.

Figure 14.48 shows a a widefield SP image of a 5micron protein grating monolayer taken at a defocus of approximately -1 μm, there is strong image contrast and, more importantly, superior lateral resolution compared to the non-interferometric microscope.

14.8 Discussion and Conclusions

Many aspects of SP and surface wave microscopy have been discussed in this chapter. We have concentrated on imaging modes that use non-localised surface waves essentially on planar substrates. In order to give a more rounded view of the field some brief mention of areas not covered is needed. For this reason we will very briefly mention the following areas: (i) total internal reflection (fluorescence) microscopy, TIR(F)M. (ii) Localized SPs and (iii) 'Exotic' techniques.

14.8.1 Relationship of SP Methods with TIR(F)M Methods

Total internal reflection microscopy is far more mature than surface wave microscopy (Axelrod, 1984, Toomre and Mannstein, 2001, Axelrod, 2001, see also Chap. 8 in this book). It is therefore certainly not appropriate to review this area but this section will indicate the relation between the two techniques. The transmission coefficient for *p*-polarized light is shown in Fig. 14.49, this is obtained under similar conditions to Fig. 14.31 except that the guiding layers are no longer present. The transmitted field shows quite a large peak above the critical angle, we also note that although there is a phase variation it is, however, considerable weaker compared to the case where a surface wave is excited. The mean gradient is smaller indicating that considerable energy is only transported for short distances, moreover, the gradient does not remain constant so the displaced field will be distorted. Modelling, the situation with annular excitation we see that when the sample is moved above the focus there is only a tiny increase in signal, of approximately 3%; this indicates that the self-focusing and improved generation efficiency are extremely weak compared to the case when SPs are excited. On the other hand, the fact that TIR occurs for all angles above the critical angle indicates that the control of the incident angle is much less stringent (also allowing for tuning of the penetration depth).

Total internal reflection microscopy is usually operated with fluorescent labels and the technique is powerful and less experimentally demanding compared to the SP and SW fluorescent methods discussed earlier. This technique should, however, have advantages for pump/probe experiments and harmonic imaging where the field enhancement is crucial. For practical biological imaging with fluorescent labels single photon and two photon TIRFM represent a more convenient means for imaging cell surfaces unless the techniques are to be combined with other more exotic techniques.

Where SP techniques have a major advantage compared to TIR methods is for label-free detection. A great merit of SP methods is that the two contrast mechanisms mentioned in Sect. 14.3 operate. For instance, the interferometers discussed in Sect. 14.7 give strong quantitative contrast arising from the change in *k*-vector of the travelling wave. This is demonstrated in the $V(z)$ curves whose periodicity is a function of wave number of the propagating surface wave modes. TIR methods are

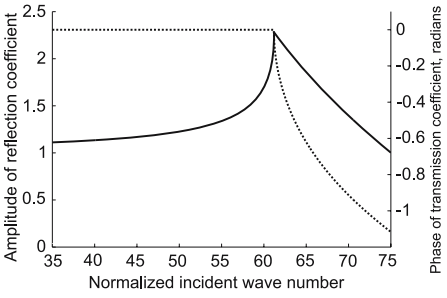


Fig. 14.49. Transmission coefficient for uncoated glass layer ($n = 1.518$) corresponding to total internal reflection. Solid curve amplitude, dotted curve phase

insensitive to this contrast mechanism, so that label-free binding studies, material adhesion etc is best performed with SPs. This is not to say that TIR methods cannot be used for label-free imaging, but only when the scattering mechanism (mechanism 2 of Sect. 14.3) predominates. Indeed TIR can be very effective for monitoring local scatters such as the passage of particles through, for instance, a cell membrane (Byrne *et al.* 2007).

In summary despite the instrumental similarity SP, SW and TIR techniques have distinct areas of application, which usually depends on the contrast mechanisms one wishes to exploit.

14.8.2 Localized SPs

We have concentrated on SPs and SWs propagating along a plane. Another area of growing interest is the use of localized SPs for microscopy. Typically, these involve noble metal particles with diameter between $10 - 100\,\text{nm}$. Localized SPs propagate on the surface of these nanoparticles, and like non-local SPs their propagation properties is affected by the local environment within a distance comparable to the particle diameter (Hutter and Fendler, 2004). These changes are usually addressed by a change in the scattering cross-section as a function of interrogating wavelength. Clearly the very localization of SPs on a nanoparticle means that many of the trade-offs between resolution and sensitivity discussed in this chapter do not arise. The challenges associated with these particles are more related to optimization of signal levels and positioning the particles at the desired locations.

Both localized and non-localized SPs rely on field enhancement and environmental sensitivity for their operation, they both represent powerful tools for label-free imaging, and there are many areas in, say, binding studies where they are competing techniques. It should be noted that although the physics of the environmental interactions are similar the instrumentation necessary to exploit their properties are rather different.

14.8.3 'Exotic' Techniques

The methods discussed here have concentrated on obtaining diffraction limited resolution with surface waves, we will briefly discuss two methods which potentially break the normal diffraction limit by (i) utilizing evanescent waves in image formation and (ii) using surface waves with extraordinarily small wavelength so that diffraction limit becomes extremely small. Both methods are in their infancy and are a long way off from being practical imaging techniques, on the other hand, their potential is such that it would be a pity not to mention them.

The first method involves the use of the 'Pendry' lens (Pendry, 2000), referred to briefly in footnote 1. The essential idea is that if both the permittivity and permeability can be made negative then a material has a negative refractive index. The result is that evanescent waves can be amplified with the result that 'perfect' image contruction is possible. Consider a sample with very fine features, the

light scattered from this will consist of propagating waves and evanescent waves which will decay as they propagate through 'free' space. A slab of negative index will regenerate the evanescent waves so that, in principle at least, the original spectrum is recovered in the image plane. Such negative index materials are not presently obtainable at optical wavelengths (although in the microwave region these can be constructed, Wiltshire *et al.* 2001). It has been shown, however, by Pendry 2000, that the electric and magnetic fields are decoupled in the near field so that only the permittivity need to be considered hence plasmonic materials such as silver can be used as (imperfect) 'Pendry' lens. Fang *et al.* (2005) have performed an elegant if somewhat contrived experiment where a fine featured structure was imaged through a 35 nm of PMMA and a 35 nm slab of silver. The image was projected onto photoresist, which was read with an atomic force microscope. The results showed that the image resolution was considerably enhanced compared to free space propagation, indicating that the plasmon structure had enhanced the resolution.

The second technique that should be mentioned involves genuine generation of far field imaging with SPs (Smolyaninov *et al.*, 2005a and 2005b). The method uses a small drop of glycerine placed on a gold coated prism. The sample to be examined is positioned on the gold under the droplet. At first sight this might appear to be rather like a Kretschmann confiuguration, it is, however, very different. A key concept behind the experiment is that the k-vector of the SP becomes vanishingly large when the dielectric constant of the metal is equal and opposite to that of the dielectric (assuming no losses). The glycerine is used because it best meets this condition. The illuminating light is converted to a very short wavelength SP mode by the sample. The droplet subsequently performs its second function; that is acting as a mirror for the surface waves thus forming a greatly magnified (if distorted) image of the sample. Since the surface roughness in the sample scatters light back into free space the image can then be viewed with a conventional objective, since the *magnified* image can be faithfully represented with propagating light. Lateral resolution consistent with the predicted SP wavelength of 70 nm has been observed. Clearly, the experiment as it stands is a long way from routine use, however, the observation that ultrashort wavelength SPs excited at optical wavelengths can be exploited is extremely signficant and is sure to find considerable future application.

In summary this chapter has reviewed some of the recent developments in SP and surface wave microscopy. In particular, the need to monitor samples for long periods with large bandwidth means that there is a growing need to extract maximum information from each pixel. This sets a limit on the minimum number of photons that need to be observed, on the other hand, fluorophores are limited in the maximum number of photons they can emit before photobleaching. Label-free detection will thus become increasingly important and it is in this area that SP techniques are expected to play a major role in biological imaging.

Acknowledgements

The author wishes to thank the Paul Instrument fund, the Engineering and Physical Sciences Research Council and the Wellcome Trust for financial support. A particular debt is due to Dr Shugang Liu who obtained our first real results. I am very pleased to acknowledge the contribution of many colleagues who have made a large contribution to this work: Gerard Byrne, Jason Goh, Mahadi Abdul Jamil, Mark Pitter, Chung Wah See, Graham Stabler, Tzvetan Velinov, Kelly Vere and Jing Zhang. Finally, I would finally like to thank Prof. Paul O'Shea of the School of Biology, University of Nottingham and Dr. Morgan Denyer of Bradford University, School of Pharmacy, for their touching belief that these techniques would one day be useful.

References

1. Atalar A (1978) An Angular Spectrum Approach to contrast in the reflection acoustic microscopy, *J. Appl. Phys.* **49**, 5130–5139
2. Atalar A (1979) A physical model for acoustic signatures, *J. Appl. Phys.* **50**, 8237–8239
3. Axelrod D, Burghardt TP and Thompson NL (1984) Total internal reflectance fluorescence. *Ann. Rev. Biophys. Bioeng.* **13**, 247–268
4. Axelrod D, 'Total Internal Reflection Fluorescence Microscopy in Cell Biology', *Traffic*, **2**, pp. 764–774, 2001
5. Azzam RMA and Bashara NM (1986) *Ellipsometry and polarized light,* Elsevier
6. Berger CEH, Kooyman RPH and Greve J, (1994) Resolution in Surface plasmon microscopy, *Rev. Sci. Instr.* **65**, 2829–2835
7. Bertoni HL amd Tamir T (1973), Unified theory of Rayleigh-angle phenomena for acoustic beams at liquid-solid interfaces, *Appl. Phys. A* **2**(4), 157–172
8. Bertoni HL and Somekh MG (1986) Ray-optical analysis of spherical focusing transducers for acoustic microscopy, *IEEE T Ultrason. Ferr.* **33**, 91
9. Born M and Wolf E, *Principles of Optics,* 7th expanded edition Pergamon (1999)
10. Bozhevolnyi SI (1996), Localization phenomena in elastic surface-polariton scattering caused by surface roughness, *Phys Rev B,* 54 (11): 8177–8185
11. Brekhovskikh LM (1960) *Waves in layered media,* Academic Press, New York, London
12. G.D. Byrne, S. Stolnik, M.C. Pitter, J. Zhang, M.G. Somekh (2007), Label-free total internal reflection microscopy for long term imaging of nanoparticle drug uptake by live cells, to be submitted to J. Microscopy
13. Campagnola PJ, Mei-de Wei, A Lewis and Loew LM(1999), High-Resolution Nonlinear Optical imaging of Live Cells by Second harmonic generation, *Biophysical Journal* **77**, 3341–3349
14. Durnin J, Micelli JJ, Eberly JH, Diffraction-free beams (1987), *Phys. Rev. Lett.* **58** (15), 1499–1501
15. Dyba M, Jakobs S, Hell SW (2003), Immunofluorescence stimulated emission depletion microscopy, *Nat Biotechnol* 2003, **21**, 1303–1304
16. Fanton JT, Opsal J, Willenborg DL, Kelso SM, Rosencwaig A (1993), Multiparameter measurements of Thin-Films using beam-profile reflectometry, *J. Appl. Phys.* **73**, 7035–7040
17. Flanagan MT and Pantell RH (1984), Surface Plasmon Resonance and immunosensors, *Elect. Lett*. **20**, 969–970

18. Giebel KF, Bechinger C, Herminghaus S, Riedel M, Leiderer P, Weiland U and Bastmeyer M (1999), Imaging of cell/Substrate Contacts of Living cells, *Biophysical Journal* **76**, 509–516
19. Green RJ , Tasker S, Davies J, Davies MC, Roberts CJ and Tendler SJB (1997), Adsorption of PEO-PPO-PEO Triblock Copolymers at the Solid/Liquid, *Langmuir* **13**, 6510–6515
20. Grigorenko AN, Beloglazov AA, Nikitin PI, Kuhne C, Steiner G and Salzer R (2000), Dark field surface plasmon resonance microscopy *Opt. Comms.* **174**, 151–155
21. Grigorenko AN, Nikitin PI, Kabashin AV (1999), Phase jumps and interferometric surface plasmon resonance imaging, *Appl. Phys. Lett.* **75** (25), 3917–3919
22. Hecht B, Bielefeldt H, Novotny L, Inouye Y and Pohl DW (1996), Local Excitation, Scattering, and Interference of Surface Plasmons, *Phys. Rev. Lett* **77**, 1889–1892
23. Hell S W, Lindek S, Stelzer E H K, (1994) Enhancing the axial resolution in far-field light microscopy-2 photon 4 PI confocal fluorescence microscopy, *J. Mod. Opt.* **41**, 675–681
24. Higdon PD, Török P and Wilson T, (1999) Imaging properties of high aperture multiphoton fluorescence optical microscopes, *J. Microsc.* **193** (Part.2), 127–141
25. Hopkins H H, (1943) The Airy disc formula for systems of high relative aperture, *Proc Phys Soc.* **55**, 116–128
26. Hutter E and Fendler JH (2004), Exploitation of Localized Surface Plasmon Resonance, Advanced Materials 16 (19), 1685–1706.
27. Ilett C, Somekh MG and Briggs GAD (1984), Acoustic microscopy of elastic discontinuities *Proc. Roy. Soc.-Series A*, **393**, 171–183
28. Kano H and Knoll W (1998), Locally excited surface-plasmon-polaritons for the thickness measurement of LBK films, **153**, 235–239
29. Kano H and Knoll W (2000), A scanning microscope employing localized surface-plasmon-polaritons as a sensing probe *Opt. Commun.* **182**, 11–15
30. Kano H, Mizuguchi S and Kawata S (1998), Excitation of Surface-plasmon polaritions by a focused Laser Beam, *J. Opt. Soc. Am*. B **15**, 1381–1386
31. Knoll W, Interfaces and Thin films as seen by bound Electromagnetic waves (1998), *Ann. Rev. Phys. Chem.* **49**, 569–638
32. Konopsky VN (2000), Operation of surface plasmon near-field microscope with gold and silver tips in tapping mode: demonstration of subtip resolution, *Optics Communications* **185**, 83–93
33. Konopsky VN, Kouyanov KE, Novikova NN (2001), Investigations of the interference of surface plasmons on rough silver surface by scanning plasmon near-field microscope, *Ultramicrscopy* **88** (2) 127–138
34. Johnson PB and Christy RW (1972) Optical constants of noble metals, *Phys. Rev.* **6** (12) 4370–4379
35. Kovacs G (1982), 'Optical excitation of surface plasmon-polaritons in layered media' in Electromagnetic surface modes, edited by A D Boardman, John Wiley and Sons Ltd
36. Mansfield SM and Kino GS (1990), Solid Immersion microscope, *Appl. Phys. Lett.* **57** (24), 2615–2616
37. Migdall AL, Roop B, ZhengYC, Hardis JE and Xia GJ (1990), Use of heterodyne detection to measure optical transmittance over a wide range, *Appl. Opt*. **29**, 5136–5144
38. Notcovich AG, Zhuk V and Lipson SG, Surface plasmon resonance phase imaging, *Appl. Phys. Lett.* **76** (13), 1665–1667 (2000)
39. Tanaka T and Yamamoto S (2003), Laser-scanning surface plasmon polariton resonance microscopy with multiple detectors, *Appl. Opt.*, **42** (19), 4002–4007

40. Terris BD, Mamin HJ, Rugar D, Studenmund WR and Kino GS (1994), Near-field optical-data storage using a solid immersion lens, *Appl. Phys. Lett*. **65** (4) 388–390.
41. Offside MJ, Somekh MG, See CW (1989), Common Path Scanning Heterodyne Optical Profilometer for absolute phase measurement, *Appl. Phys. Lett.* **55** (20), 2051–2053
42. Parmon W and Bertoni HL (1979), Ray interpretation of the material signature in the acoustic microscope, *Elect. Lett.* 15, 684–686
43. Pendry JB (2000), 'Negative Refraction Makes a Perfect Lens', *Phys. Rev. Lett.* **85** (18), 3966–3969
44. Raether H (1988), Surface plasmons on smooth and rough surfaces and on gratings, Springer-Verlag, Berlin
45. Richards B and Wolf E, (1959) Electromagnetic diffraction in optical systems, II Structure of the image in an aplanatic system, *Proc. Roy. Soc. Lond.* **253**, 358–379
46. Rothenhausler B and Knoll W (1988), Surface-plasmon Microscopy, *Nature* **332**, 615–617
47. Sawyer NBE, Morgan SP, Somekh MG, See CW, Cao XF, Shekunov BY and Astrakharchik E (2001), Wide field amplitude and phase confocal microscope with parallel phase stepping, *Rev. Sci. Inst.* **72**, 3793–3801
48. See CW, Somekh MG and Holmes R, (1996) Scanning optical microellipsometer for pure surface profiling, *Appl. Opt.* **35**, 6663–6668
49. Shatalin SV, Juskaitis R, Tan JB, Wilson T (1995), Reflection Conoscopy and microellipsometry of isotropic thin-film structures, *J. Microsc.* **179**, 241–252.
50. Smolyaninov II, Elliott J, Zayats AV and Davis CC (2005a), Far-Field Optical Microscopy with a Nanometer-Scale Resolution Based on the In-Plane Image Magnification by Surface Plasmon Polaritons, *PRL* **94**, 057401, 1–4
51. Smolyaninov II, Davis CC, Elliott J and Zayats AV (2005b), Resolution enhancement of a surface immersion microscope near the plasmon resonance, *Opt. Lett.* **30** (4), 382–384
52. Somekh MG (2002), Surface plasmon fluorescence microscopy: an analysis, *J. Microsc*, **206**, Pt 2, 120–131
53. Somekh MG, Bertoni HL, Briggs GAD and Burton NJ (1985), A two-dimensional imaging theory of surface discontinuities with the scanning acoustic microscope, *Proc. Roy. Soc. Lond. Series* A **401** 29–51
54. Somekh MG , Liu S and Velinov TS (2000a), Optical $V(z)$ for high resolution "2π" surface plasmon microscopy, *Opt. Lett*., **25**, 823–825
55. Somekh MG, Liu SG, Velinov TS and CW See (2000b), High resolution scanning surface plasmon microscopy, *Appl. Opt.* **39** 6279–6287
56. Somekh M G, See C W and Goh J (2000), Wide field amplitude and phase confocal microscope with speckle illumination, *Opt. Comm.* **174**, 75–80
57. Stabler G, Somekh MG, See CW (2004), High-resolution wide-field surface plasmon microscopy, *J. Microsc.* **214** Part 3, 328–333
58. Stefani FD, Vasilev K, Bocchio N, Stoyanova N, Kreiter M (2005), Surface-plasmon-mediated single-molecule fluorescence through a thin metallic film, *Phys. Rev. Lett.* **94** (2): Art. No. 023005
59. Toomre D and Manstein DJ (2001) Lighting up the cell surface with evanescent wave microscopy, *Trends in Cell Biology* **11** (7), 298–303
60. Török P, Varga P, Laczik Z and Booker G R, (1995) Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive indices: and integral representation, *J.O.S*. *A***12**, 325–332
61. Velinov T, Somekh MG and Liu S (1999), Direct far-field observation of surface-plasmon propagation by photoinduced scattering, *Appl. Phys. Lett.* **75**, 3908–3910

62. Velinov T, Somekh MG, Liu S (1999), Direct far-field observation of surface-plasmon propagation by photoinduced scattering, *Appl. Phys. Lett.* **75**, 3908–3910
63. Yeatman EM (1996), Resolution and sensitivity in surface plasmon microscopy and sensing, *Biosensors Bioelectron* **11**, 635–649
64. Yeatman EM and Ash EA (1987), Surface Plasmon Microscopy, *Elect. Lett*. **23**, 1091–1092
65. Zhang J (2006), High resolution Solid Immersion Lens Microscopy and its application to Surface Plasmon Resonance Imaging, Ph. D thesis University of Nottingham
66. Zhang J, See CW, Somekh MG, Pitter MC, and Liu SG (2004), Wide-field surface plasmon microscopy with solid immersion excitation, *Appl. Phys. Lett*. **85** (19), 5451–5453
67. Zhang J, Pitter MC, Liu S, See CW and Somekh MG (2006), Surface plasmon microscopy with a two-piece solid immersion lens: bright and dark field, *Appl. Opt.* to be published Nov 2006
68. Zhou H and Sheppard CJR (1997), Aberration measurement in confocal microscopy: Phase retrieval from a single intensity measurement, *J. Mod Opt*, **44**, 1553–1561

15 Optical Coherence Tomography

Adam M. Zysk and Stephen A. Boppart

15.1 Introduction

Optical coherence tomography (OCT) is an established imaging technique for a wide range of biological, medical, and material investigations [1–3]. Advances in the use of OCT for microscopy applications have provided researchers with a means by which biological specimens and non-biological samples may be visualized in cross-section. OCT was initially developed for imaging biological specimens because it permits the imaging of tissue microstructure *in situ*, yielding micron-scale image resolution without the need for excision of a specimen or for tissue processing. OCT is analogous to ultrasound B-mode imaging except that it uses low-coherence light rather than high-frequency sound and performs imaging by measuring the backscattered intensity of light from structures in tissue. OCT can image tissue or specimens in cross-section, as is commonly done in ultrasound, or in *en face* sections, as in confocal and multi-photon microscopy. The OCT image is a gray-scale or false-color multi-dimensional representation of backscattered light intensity that represents the differential backscattering contrast between different tissue types on a micron scale. Because OCT performs imaging using light, it has a one to two order-of-magnitude higher spatial resolution than ultrasound and does not require contact with the specimen or sample. The use of light also enables characterization of the spectroscopic and birefringence properties of tissue and cellular structures.

OCT was originally developed and demonstrated in ophthalmology for high-resolution tomographic imaging of the retina and anterior eye [4–7]. Because the eye is transparent and is easily accessible using optical instruments and techniques, it is well suited for diagnostic OCT imaging. OCT is routinely used for the diagnosis of retinal disease because it can provide images of retinal pathology with micron-scale resolution. Clinical studies have been performed to assess the application of OCT for a number of macular diseases [5–7]. OCT is especially promising for the diagnosis and monitoring of glaucoma and macular edema associated with diabetic retinopathy because it permits the quantitative measurement of changes in the retina or the retinal nerve fiber layer thickness. Because morphological changes often occur before the onset of physical symptoms, OCT provides a powerful approach for the early detection of these diseases.

Since the application of OCT in ophthalmology, OCT has been applied for imaging in a wide range of nontransparent tissues [8–12]. In tissues other than the eye,

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

the imaging depth is limited by optical attenuation due to scattering and absorption. Ophthalmic imaging is typically performed at a center wavelength of 800 nm. Because **optical scattering** decreases with increasing wavelength, OCT imaging in nontransparent tissues is possible using 1.3 μm or longer wavelengths. In most tissues, imaging depths of 2 – 3 mm can be achieved using a system detection sensitivity of 100 to 110 dB. Imaging studies have also been performed in virtually every organ system to investigate applications in cardiology [13–16], gastroenterology [17,18], urology [19,20], neurosurgery [21], and dentistry [22], to name a few. High-resolution OCT using short coherence length, short pulsed light sources has been demonstrated and axial resolutions of less than 1 μm have been achieved [23–28]. High-speed OCT at image acquisition rates of up to 200 frames per second have been achieved, depending on the image size [29]. OCT has been extended to perform Doppler imaging of blood flow [30–32], birefringence imaging to investigate laser intervention [33–35], and spectroscopic imaging for the evaluation of sub-cellular structures [36]. Forward imaging delivery systems, including transverse imaging catheter/endoscopes, needles, and handheld devices, have been developed to enable *in vivo* internal body OCT imaging in animal models and human patients [37–45].

15.2 Principles of Operation

OCT is based on optical ranging, the high-resolution, high dynamic range detection of backscattered light as a function of optical delay. In contrast to ultrasound, the velocity of light is extremely high. Therefore, the time delay of reflected light cannot be measured directly and **interferometric detection** techniques must be used. One method for measuring time delay is to use **low coherence interferometry** or **optical coherence domain reflectometry**. Low-coherence interferometry was first developed in the telecommunications field for measuring optical reflections from faults or splices in optical fibers [46]. Subsequently, the first applications in biological samples included one-dimensional optical ranging in the eye to determine the location of different ocular structures [47,48].

The time delay of reflected light is typically measured using a **Michelson interferometer** (Fig. 15.1). Other interferometer designs, such as a Mach-Zehnder interferometer, have been implemented to optimize the delivery and collection of the OCT beam [49,50]. The light reflected from the specimen or sample is interfered with light that is reflected from a reference path of known path length. **Interference** of the light reflected from the sample arm and reference arm of the interferometer can occur only when the optical path lengths of the two arms match to within the coherence length of the optical source. Depth scanning can be performed in the time- or spectral-domain. **Time-domain OCT** systems vary the reference arm pathlength, inducing changes in the depth from which the backscattered light from the sample is detected. Measured at the interferometer output port, the signal is electronically band-pass filtered, demodulated, digitized, and stored on a computer [2]. **Spectral-domain OCT** systems keep the reference path stationary and calculate the

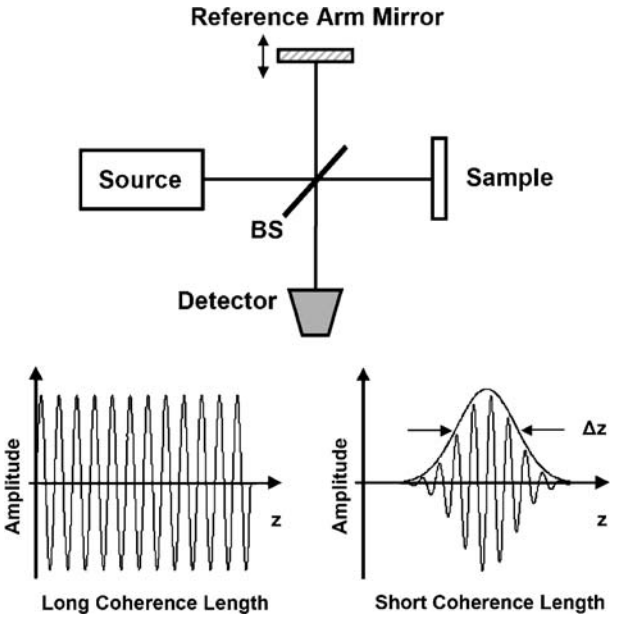


Fig. 15.1. Low coherence interferometry. Using a short coherence length light source and a Michelson-type interferometer, interference fringes are observed only when the path lengths of the two interferometer arms are matched to within the coherence length of the light source

axial signal intensity from changes in the interference spectrum by way of the time-frequency relationship [51,52]. Spectral-domain detection may be accomplished by using a spectrometer for measurement of the interfered light from a broadband source [53], or by rapidly and repeatedly sweeping a narrow-linewidth laser source through a gain-medium spectrum in a mode called **swept-source OCT** [54,55].

The position of the incident beam on the specimen is scanned in the transverse plane and multiple axial measurements are performed. This generates a two-dimensional (2-D) data array that represents the optical backscattering through a cross-sectional plane in the specimen (Fig. 15.2). The logarithm of the backscatter intensity is then mapped to a false-color or gray-scale and displayed as an OCT image. The interferometer in an OCT instrument can be implemented using a fiber optic coupler and beam-scanning can be performed with small mechanical galvanometers in order to yield a compact and robust system (Fig. 15.3).

In contrast to conventional microscopy, the axial resolution in OCT images is determined by the coherence length of the light source. The axial point spread function of the OCT measurement, as defined by the signal detected at the output of the interferometer, is the electric-field autocorrelation of the source. The **coherence length** of the light is the spatial width of the field autocorrelation and the envelope of the field autocorrelation is equivalent to the Fourier transform of its power spectrum. Thus, the width of the **autocorrelation function**, or the axial resolution, is inversely

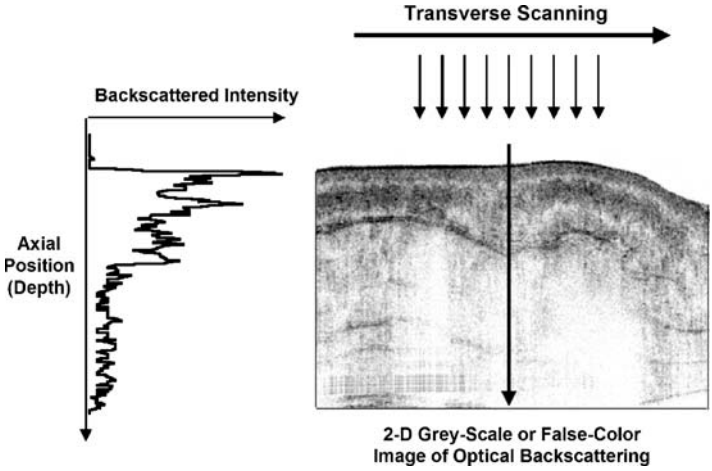


Fig. 15.2. OCT image formation. An OCT image is based on the spatial localization of variations in optical backscatter from within a specimen. Images are acquired by performing axial measurements of optical backscatter at different transverse positions on the specimen and displaying the resulting two-dimensional data set as a gray-scale or false-color image

proportional to the width of the power spectrum. For a source with a Gaussian spectral distribution, the axial resolution Δz is given by

$$\Delta z = \frac{2\ln 2}{\pi} \cdot \frac{\lambda^2}{\Delta\lambda} , \tag{15.1}$$

where Δz and $\Delta\lambda$ are the full-widths at half-maximum of the autocorrelation function and power spectrum respectively and λ is the center wavelength of the optical source. Figure 15.4 illustrates the dependence of the axial resolution on the bandwidth of the optical source. High axial resolution requires broad bandwidth optical sources. Curves are plotted in Fig. 15.4 for three commonly used wavelengths. Higher resolutions are achieved with shorter wavelengths, however, shorter wavelengths are more highly scattered in biological specimens, resulting in less imaging penetration.

The transverse resolution in an OCT imaging system is determined by the focused spot size in analogy with conventional microscopy and is given by

$$\Delta x = \frac{4\lambda}{\pi} \cdot \frac{f}{d} , \tag{15.2}$$

where d is the beam diameter incident on the objective lens and f is the focal length of the objective. High transverse resolution can be obtained by using a large numerical aperture and focusing the beam to a small spot size. The transverse resolution is also related to the depth of focus or the confocal parameter $2z_R$ (two times the **Raleigh range**),

$$2z_R = \frac{\pi\Delta x^2}{2\lambda} . \tag{15.3}$$

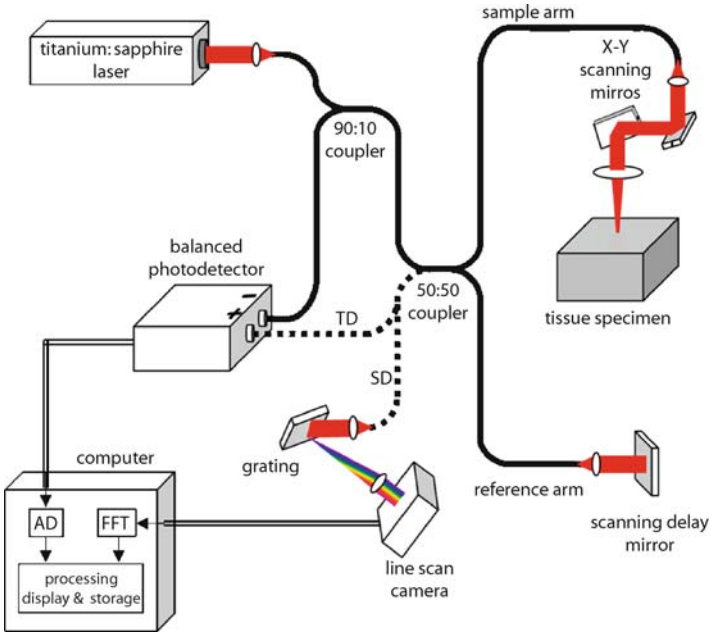


Fig. 15.3. Schematic of an OCT system implemented using fiber optics. The Michelson interferometer is implemented using a fiber-optic coupler. Light from the low-coherence source is split and sent to a sample arm with a beam delivery instrument and a reference arm with an optical pathlength scanner. Reflections from the arms are combined and the output of the interferometer is detected with either a photodiode for time-domain (TD) detection or a spectrometer for spectral-domain (SD) detection. The signals are processed by a computer and displayed as an image

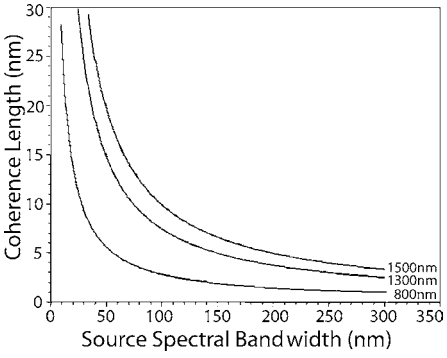


Fig. 15.4. Dependence of coherence length (axial resolution) on optical source bandwidth. Curves are plotted for 800 nm, 1300 nm, and 1500 nm, three common wavelengths used in OCT. High axial imaging resolution is achieved with broad spectral bandwidths and shorter wavelengths. Shorter wavelengths, however, are more highly absorbed in biological tissue, decreasing imaging penetration depth

Thus, increasing the transverse resolution results in a reduced depth of field. Typically, the **confocal parameter** or depth of focus is chosen to match the desired depth of imaging. High transverse resolutions may be utilized in OCT, however, the short

depth of field requires spatially tracking the focus in depth along with the axial OCT scanning.

Finally, the detection signal-to-noise (SNR) is given by the optical power backscattered from the sample (P_{SAM}) divided by the noise equivalent bandwidth (NEB)

$$SNR = 10\log\left(\frac{\eta}{\hbar\omega}\frac{P_{\mathrm{SAM}}}{NEB}\right), \tag{15.4}$$

where η is the quantum efficiency of the detector, ω is radian frequency of the source, and $\hbar$ is Plank's constant divided by 2π.

Depending upon the desired signal to noise performance, incident powers of $1-10\,\mathrm{mW}$ are required for OCT imaging. It is possible to acquire 2048×2048 pixel images at a rate of over 14 frames per second and a signal-to-noise ratio of 100 dB with $1-10\,\mathrm{mW}$ of incident optical power [56]. If lower data acquisition speeds or signal-to-noise can be tolerated, power requirements can be reduced accordingly.

15.3 Technological Developments

Since the inception of OCT in the early 1990's, there have been rapid technological developments aimed at improving the imaging resolution, acquisition rate, and methods of beam delivery to the tissue or sample. Investigators have also explored other imaging methods using the principles of OCT to extract information from the tissue or sample. Some of these methods have included acquiring optical **Doppler** signals from moving scatterers or structures, obtaining images based on the **polarization state** of the incident and returned light, mapping **refractive index** distributions, and extracting **spectroscopic** information based on the local absorption or scattering properties of the tissue [30,33,57–59].

15.3.1 Optical Sources for High-Resolution Imaging

The majority of OCT imaging systems to date have used **superluminescent diodes** (SLDs) as low coherence light sources [60,61]. Superluminescent diodes are commercially available at several wavelengths including 800 nm, 1.3 μm, and 1.5 μm, and are attractive because they are compact, are highly efficient, and have low noise. However, output powers are typically less than a few milliwatts, limiting fast real-time image acquisition rates. Additionally, the available bandwidths are relatively narrow, permitting imaging with $10-15$ micron resolution. Recent advances in short-pulse solid-state laser technology make these sources attractive for OCT imaging in research applications. **Femtosecond solid-state lasers** can generate tunable, low-coherence light at powers sufficient to permit high-speed OCT imaging. Short pulse generation has been achieved in **titanium:sapphire** ($\mathrm{Ti{:}Al_2O_3}$) across the full wavelength range from 0.7 μm to 1.1 μm and over more limited tuning ranges near 1.3 μm and 1.5 μm in **chromium:forsterite** ($\mathrm{Cr^{4+}{:}Mg_2SiO_4}$) and **chromium:**

ytterbium-aluminum-garnet (Cr^{4+}:YAG) lasers, respectively. OCT imaging with axial resolutions of 0.5-2 μm and 5 μm has been demonstrated at 800 nm and 1.3 μm, respectively, using Ti:Al_2O_3 and Cr^{4+}:Mg_2SiO_4 sources [25–28]. A comparison between the power spectra and autocorrelation functions of a 800 nm center wavelength SLD and a short pulse (≈5.5 fs) titanium:sapphire laser is shown in Fig. 15.5. An order-of-magnitude improvement in axial resolution is noted for the short pulse titanium:sapphire laser source [26]. More compact and convenient sources, such as rare-earth-element-doped fiber sources, have also been investigated, yielding very broad bandwidths in the 1.1 μm to 1.6 μm spectral range and resolutions

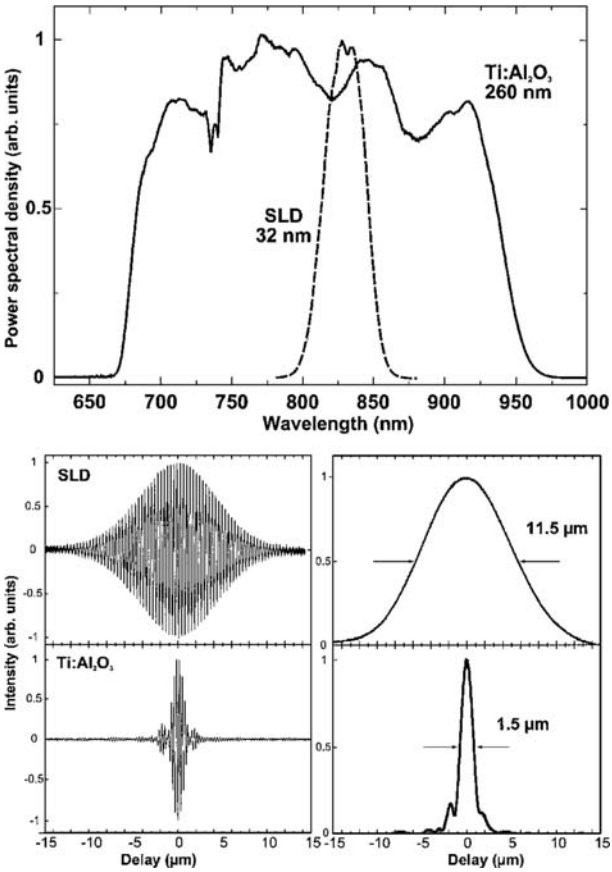


Fig. 15.5. High-resolution OCT imaging. A comparison of optical output spectrum (*top*), interference signals (*bottom left column*), and envelopes (*bottom right column*) for a Kerr-lens modelocked titanium:sapphire (Ti:Al_2O_3) laser versus a superluminescent diode (SLD) is shown. The broad optical bandwidth of the titanium:sapphire laser (260 nm) permits a free-space axial resolution of 1.5 μm. In comparison, the superluminescent diode with a 32 nm spectral bandwidth permits an axial resolution of 11.5 μm. Figure reprinted with permission [26]

of up to 2 μm [62–64]. Fiber-optic bandwidth broadening has also been investigated using titanium:sapphire and neodymium:glass laser sources [65–67]. **Nonlinear** and **ultrahigh-numerical-aperture fibers** have been used to increase the source bandwidth by several orders of magnitude, up to hundreds of nanometers of spectral bandwidth, representing a supercontinuum optical source [68]. The titanium:sapphire ($Ti:Al_2O_3$) laser technology is also routinely used in **multi-photon microscopy** (MPM) applications for its high peak intensities to enable multi-photon absorption and subsequent emission of fluorescence from exogenous fluorescent contrast agents [69,70]. Combined OCT and multi-photon microscopy has been used to provide complementary image data using a single optical source [53,56,71].

New sources have also been developed for high speed spectral-domain data acquisition [54,55]. Rapid scanning of a narrow output spectrum over a broad bandwidth, which enables **swept-source OCT**, has typically been implemented by scanning the output of a semiconductor optical amplifier (SOA) across a grating with a galvanometer or polygonal-scanning mirror. Systems have achieved rapid sweep rates of up to 290 kHz, instantaneous spectral linewidths of approximately 0.07 nm, and scanning bandwidths of over 100 nm [72–74]. For example, the Fourier domain mode locked swept source, constructed using a long fiber ring cavity, semiconductor amplifier, and Fabry-Perot filter, has achieved OCT imaging at acquisition rates up to 232,000 axial scans per second [74]. As the development of new swept-source laser technologies continues, the rapid acquisition rates that they enable are expected to lead to their broad adoption in real-time OCT imaging systems.

15.3.2 Spectroscopic OCT

OCT techniques can be used to extract spatially-distributed spectroscopic information from within the tissue specimen or material sample [57]. In standard OCT imaging, the amplitude of the envelope of the field autocorrelation is acquired and used to construct an image based on the magnitude of the optical backscatter at each position. **Spectroscopic OCT** information can be obtained by digitizing the full interference signal and performing digital signal processing on various time-frequency distributions. The Morlet wavelet transform was first used, rather than the short-time Fourier transform, because it reduced windowing artifacts [75,76]. Spectroscopic data can be extracted from each point within the specimen. The spectral center of mass can be calculated and compared to the original spectrum from the laser source. Shifts from the center of mass are displayed on a 2-D image using the multi-dimensional hue-saturation-luminence (HSL) color space. At localized regions within the tissue, a color is assigned with a hue that varies according to the direction of the spectral shift (longer or shorter wavelength) and a saturation that corresponds to the magnitude of that shift. The luminance is held constant.

In scattering media, longer wavelengths of light are scattered less. Therefore, for a homogeneous scattering sample, one would expect shorter wavelengths to be scattered near the surface and a smooth color-shift to occur with increasing depths as longer wavelengths are scattered. In more heterogeneous samples, such as tissue, scattering objects such as cells and sub-cellular organelles produce variations in the

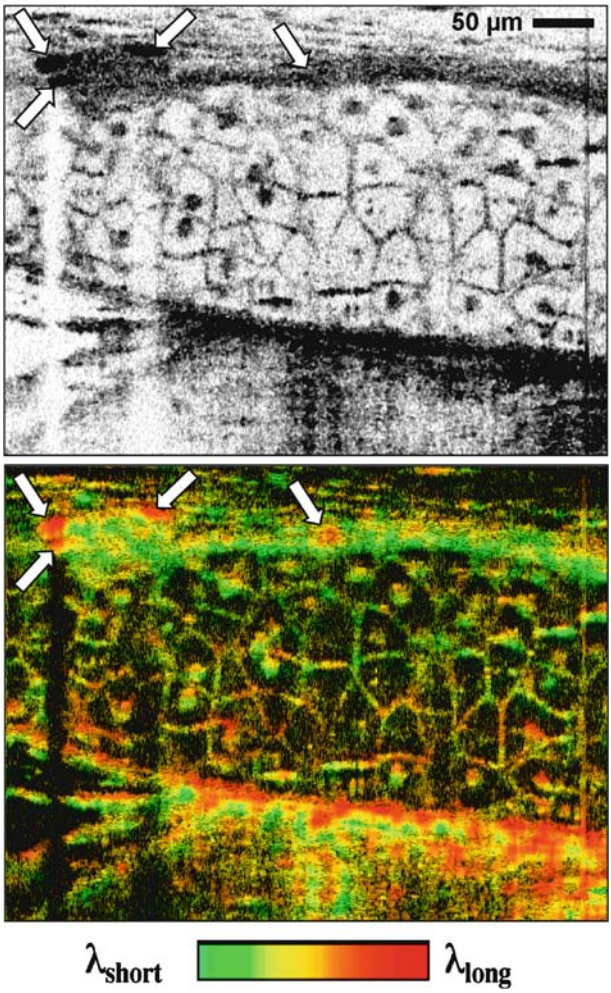


Fig. 15.6. Spectroscopic OCT imaging. Conventional OCT imaging (*top*) and spectroscopic OCT imaging (*bottom*) of *in vivo* mesenchymal cells in a *Xenopus laevis* (African frog) tadpole. Conventional OCT images represent the optical backscatter intensity while spectroscopic OCT images represent local changes in the absorption or scattering properties of the incident optical spectrum. Melanocytes (*arrows*) appear red because the melanin within these cells absorb shorter wavelengths of light. The color scale represents the shift of the center of gravity of the optical spectrum for each pixel in the image. Figure reprinted with permission [57]

spectroscopic OCT data (Fig. 15.6) [57,77]. Although images can indicate changes in the spectroscopic properties of the tissue, further investigation is needed to determine how the spectroscopic variations relate to the biological structures and how this information can be used for diagnostic purposes.

15.3.3 Real-Time Volumetric OCT Imaging

Short-pulse solid-state laser technology not only provides broad spectral bandwidths for high-resolution OCT imaging, but also higher output powers to enable fast **real-time OCT** imaging. Higher incident powers are required to maintain equivalent signal-to-noise ratios when scanning at a faster rate. Linearly translating a reference arm mirror provides axial scan frequencies of approximately 100 Hz, depending on the mirror size and the translating **galvanometer**, but is problematic at higher rates. In the past, several investigators utilized rotating glass cubes, piezoelectric modulators, and multi-pass optical cavities to increase axial scan rates while maintaining scan ranges of 1 – 2 mm [78–80]. Advances in spectral-domain detection technologies have made these techniques obsolete, however, bringing significant increases in scan rate and SNR [51,52]. Figure 15.7 shows the potential SNR advantages of **spectral-domain** detection.

With the introduction of these high-speed scanning capabilities, technological limitations have shifted from high-speed scanning methods to computational data processing and management. A set of two hundred 500 × 500 pixel 16-bit images, which can now be acquired in approximately one second, requires 100 MB of digital storage without compression. Additionally, the Fourier-domain scan line data acquired in spectral-domain systems must be transformed into the spatial domain prior to display. Therefore, displaying the same data set in real-time requires that 100,000 Fast-Fourier Transform (FFT) operations be performed. The computer resources required to process and store data acquired at this rate are substantial. The most advanced spectral-domain systems are capable of sustaining this acquisition rate for only about 3 seconds [29]. Beyond the computational challenges presented by this new technology are the clinical issues that arise when attempting to evaluate large data sets. For example, SD-OCT of the entire human esophagus will yield a very high-resolution 3-D data volume spanning a length of about 20 cm. Assessment and storage of this entire data set in the manner currently employed for radiological imaging would be exceptionally costly. Clinical standards must be established in order to alleviate potential expense and liability concerns.

Depth-priority scanning, commonly implemented in OCT, is performed by acquiring a single axial scan before translating the beam laterally and repeating the depth-acquisition. An alternative method for generating OCT images is **transverse-**

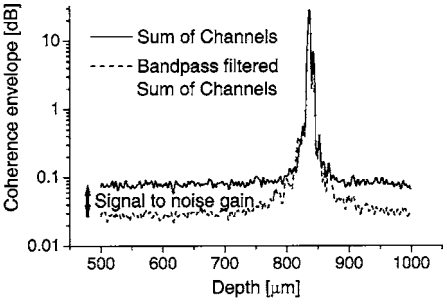


Fig. 15.7. Signal-to-noise advantage gained through spectral-domain (SD) detection. A plot of the coherence envelope from four spectrally-dispersed detection channels is shown. The SNR advantages of SD detection, gained by bandpass filtering each channel individually, is evident. Modified figure reprinted with permission [51]

priority scanning [81,82]. For this method, the optical pathlength in the reference arm is held momentarily constant, apart from a low-amplitude, high-frequency modulation to enable time-domain heterodyne detection. The OCT beam in the sample arm is then scanned laterally across the specimen in both the X and Y planes while the amplitude of the interference signal is recorded. The result, in contrast to a cross-sectional plane in depth, is a horizontal (*en face*) plane at a constant depth within the specimen. This is equivalent to optical sectioning in traditional, confocal, and multi-photon microscopy techniques. **Three-dimensional (3-D) OCT** imaging using transverse-priority scanning can be obtained by stepping the position of the reference arm mirror after each *en face* image is acquired [83].

The OCT images produced from transverse-priority scanning are readily correlated with confocal or multi-photon microscopy images, which are typically acquired with a similar scanning method. *En face* OCT imaging with high-numerical aperture objectives has been termed optical coherence microscopy (OCM) [84]. This method can also utilize higher numerical aperture objective lenses to provide high transverse resolutions since a large depth of focus is not needed as in depth-priority OCT scanning. Recently introduced computational techniques are poised to change these tradeoffs by providing depth-invariant transverse resolution. Interferometric synthetic aperture microscopy (ISAM) accounts for scattering, diffraction, and beam parameters in the raw OCT data to computationally improve image quality [131]. The technique yields high-resolution data over large volumes without requiring mechanical focus scanning or modification of existing hardware.

15.3.4 Optical Coherence Microscopy

Optical coherence microscopy combines the advantages of confocal microscopy with the principles of low-coherence interferometry [84,85]. High contrast and detection sensitivity are achieved via rejection of out-of-focus light, resulting in improved optical sectioning capabilities deep within highly-scattering specimens. Recent advancements include the integration of **spectroscopic** processing and **phase-contrast** into OCM imaging as shown in Figs. 15.8 and 15.9 [36,86].

The axial **point-spread function (PSF)** in OCM is narrower, given by the product of the field autocorrelation PSF and the microscope objective PSF, and enhances the optical sectioning capabilities of the microscopy technique. Figure 15.10 shows experimental measurements of the axial point-spread function with a 20× microscope objective (NA = 0.4) [84]. This data was obtained by scanning a mirror through the focal plane, with and without coherence gating. Considering a single-backscatter model for photons, two limits define when OCM can outperform confocal microscopy. A parametric plot of these theoretical limits is shown in Fig. 15.11 [84]. The first limit, for shorter mean-free-paths, defines when the coherence gating of OCM enhances the confocal microscopy image. This occurs only when the collected scattered light from outside of the focal plane dominates the light collected from within the focal plane. The second limit, for longer mean-free-paths, is determined by the maximum depth at which single-backscattered photons can be

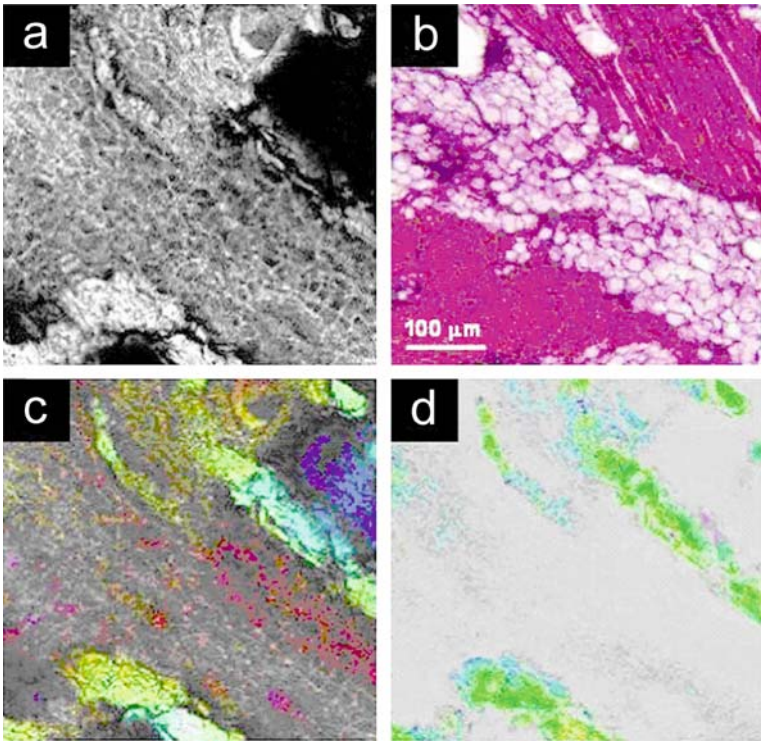


Fig. 15.8. Spectroscopic optical coherence microscopy (OCM). An OCM image (**a**) of tissue with regions of adipose (center) and muscle (*upper right, lower left*) acquired with spectral domain detection can be compared with a light microscope image (**b**) of the corresponding H&E stained histology. Spectroscopic processing by two techniques, metameric (**c**) and light scattering spectroscopy (**d**), reveals differences in scatterer size and organization. Figure reprinted with permission [36]

detected given quantum detection and the tissue laser damage thresholds. For approximately 5 – 15 scattering mean-free-paths, the use of OCM enables imaging of structures that would otherwise be obscured by scattering in conventional confocal microscopy.

While the use of OCM can improve the optical sectioning capability of confocal microscopy, further enhancement is likely with MPM. For several years investigators have recognized that OCT and MPM can utilize a single laser source for multimodality imaging. An integrated OCM and two-photon-excitation fluorescence microscope has been used to image cells in *Drosophila* embryos and **green fluorescent protein (GFP)**-vinculin transfected fibroblasts (Fig. 15.12) [36,71]. OCT and MPM provide complementary image data. OCT can image deep through transparent and highly-scattering structures to reveal the 3-D structural information. OCT, however, cannot detect the presence of fluorescent particles. In a complementary manner, MPM can localize **fluorescent probes** in three-dimensional space. MPM can de-

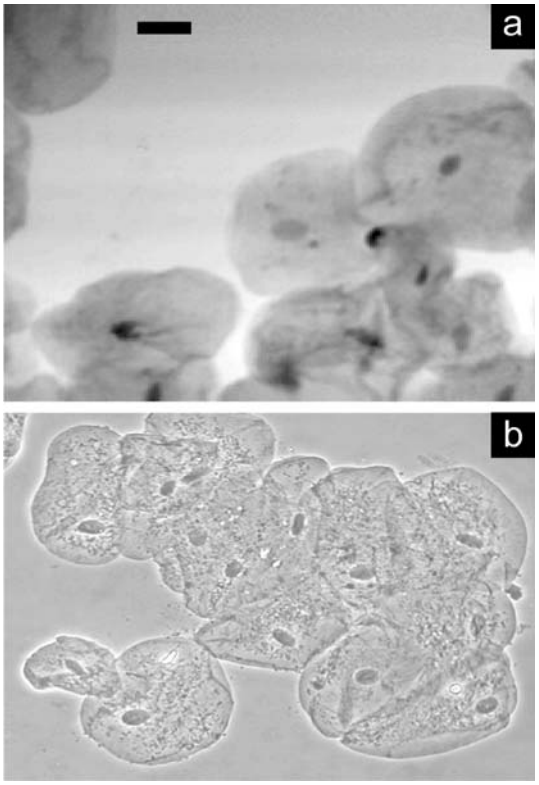


Fig. 15.9. Differential phase contrast optical coherence microscopy (OCM). The human epithelial cheek cells shown here were imaged with phase contrast OCM (**a**) and phase contrast light microscopy (**b**). The *scale bar* is 20 μm in length. Modified figure reprinted with permission [86]

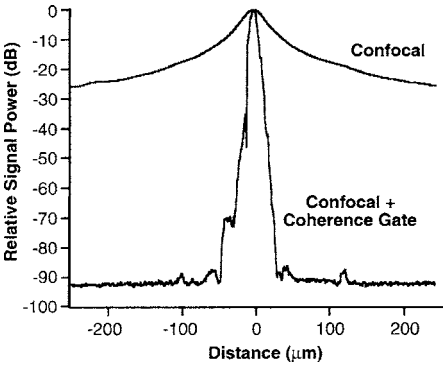


Fig. 15.10. Comparison of confocal and optical coherence microscopy point-spread functions. Experimental confocal microscopy data was obtained using a 20× microscope objective (N.A. = 0.4). A mirror was scanned through the focal plane and reflected signal power was measured with and without coherence gating. Figure reprinted with permission [84]

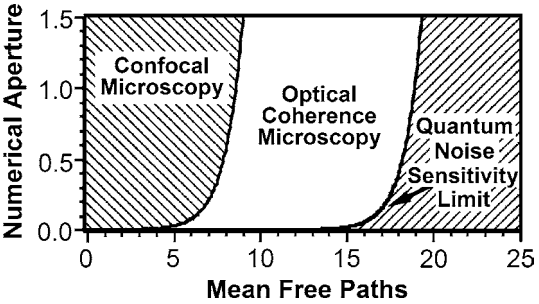


Fig. 15.11. Theoretical parametric plot of limits in optical coherence microscopy. Assuming a single-backscatter model for photons, this plot illustrates that optical coherence microscopy can enhance detection in highly-scattering media, beyond the limits of confocal microscopy alone. The right-most limit is bounded by the quantum noise sensitivity. Figure reprinted with permission [84]

tect the fluorescence, but not the microstructure nor the location of the fluorescence relative to the microstructure. Hence, the development of an integrated microscope capable of OCT and MPM uniquely enables the simultaneous acquisition of microstructural data and the localization of fluorescent probes.

15.3.5 Beam Delivery Systems

The OCT imaging technology is modular in design. This is most evident in the various optical instruments through which the OCT beam can be delivered to the tissue or sample. Because OCT systems can be implemented using fiber optics, single optical fibers can be used to deliver the OCT beam and collect the reflected light. The OCT technology can readily be integrated into existing optical instruments such as **research and surgical microscopes** [71,87], ophthalmic **slit-lamp biomicroscopes** [4], **catheters** [40], **endoscopes** [41–44,88,89], **laparoscopes** [38], needles [39,90,130], and **hand-held imaging probes** [38].

Imaging penetration is determined by the optical absorption and scattering properties of the tissue or specimen. The imaging penetration in OCT ranges from tens of millimeters for transparent tissues such as the eye to less than 3 mm in highly-scattering tissues such as skin. To image highly-scattering tissues deep within the body, novel beam-delivery instruments have been developed to relay the OCT beam to the site of the tissue to be imaged. An OCT catheter has been developed for insertion into biological lumens such as the gastrointestinal tract [40]. Used in conjunction with endoscopy, the 1 mm diameter catheter can be inserted through the working channel of the endoscope for simultaneous OCT and video imaging [41–43]. **Minimally-invasive** surgical procedures utilize laparoscopes which are long, thin, rigid optical instruments to permit video-based imaging within the abdominal cavity. Laparoscopic OCT imaging has been demonstrated by passing the OCT beam through the optical elements of a laparoscope [38]. Deep solid-tissue imaging

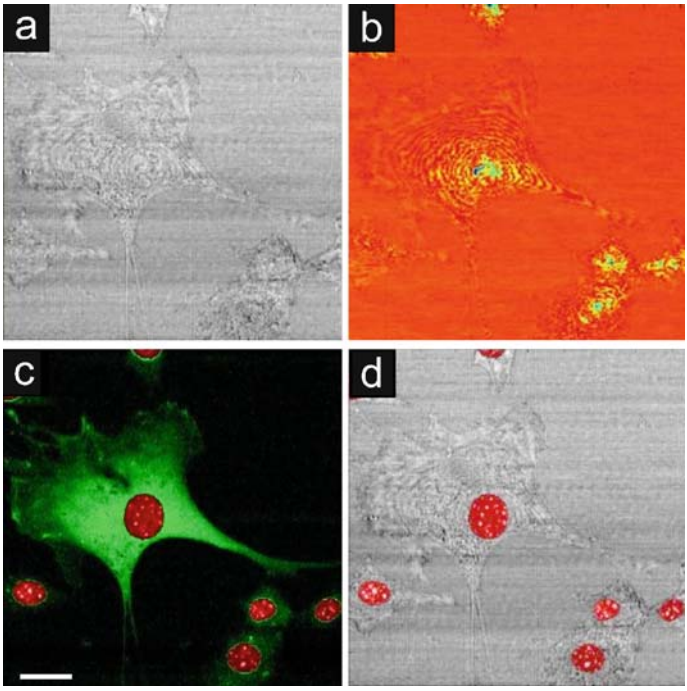


Fig. 15.12. Single cell optical coherence microscopy (OCM) and multi-photon microscopy (MPM) imaging of a green fluorescent protein (GFP)-vinculin transfected fibroblast. An OCM image acquired with spectral-domain detection (**a**) and spectroscopic analysis (**b**) show localized regions of strong spectral scattering. MPM detection yields signals from GFP and DNA nuclear dye (**c**) which can be combined with OCM imaging (**d**). Figure reprinted with permission [36]

is possible with the use of **fiber-needle probes** [39,90,130]. Small (400 μm diameter) needles housing a single optical fiber and micro-optical elements, including angle-polished rotating GRIN lens pairs, have been inserted into solid tissues to acquire OCT axial-scans or images. Recently, microfabricated **micro-electro-optical-mechanical systems (MEOMS)** technology has been used to miniaturize the OCT beam scan mechanism [88].

15.3.6 Contrast Agents and Molecular Imaging

While the use of bioluminescent and **fluorescent probes** has become commonplace in optical molecular imaging, OCT and other *in vivo* optical imaging modalities have had limited use of molecular probes. Novel nonfluorescent and nonbioluminescent molecular imaging probes for OCT have recently been developed [91]. These probes may alter the local optical scattering or absorption properties of the tissue, modulate these properties in a predictable manner, or be detected using spectroscopic OCT

principles. Nonlinear interferometric imaging techniques have also recently been developed to detect endogenous molecules.

Light scattering probes are either sensed directly by detecting scattered light or indirectly by detecting their attenuation properties. These probes work by introducing local refractive index changes using, for example, air, metals, or oils in **microbubbles** or engineered **microspheres** containing scattering nanoparticles of gold, melanin, carbon, or iron-oxide in the shell or core [92,93]. Novel ferromagnetic agents have also been developed for use in **magnetomotive OCT**, a modification of traditional OCT imaging wherein a modulated magnetic field is applied to a sample in order to perturb susceptible agents and thereby gain improved detection sensitivity. Magnetomotive detection has been shown to be highly sensitive to low concentrations (220 μg/g) of magnetite particles in tissue phantoms and in single cells [94]. *In vivo* magnetomotive imaging has been demonstrated using ***Xenopus laevis*** tadpoles and iron oxide particles, as shown in Fig. 15.13 [95]. Absorbing probes, which modify the endogenous regional absorption properties of tissue, may be detected by optical pumping of modifiable probes or spectroscopic evaluation of absorbers with specific spectral signatures (See Fig. 15.14, for example) [96]. **Methylene blue**, which can be excited by 650 nm light to shift its absorption band from 650 nm to 830 nm, is one example of a modifiable absorption probe [97]. The simplest spectroscopic probes are **NIR dyes**, but tunable plasmon-resonant nanoparticles also may hold promise.

While typical OCT measurements are limited to the detection of linear elastic scattering properties of tissue, techniques have been developed to exploit the nonlinear optical effects exhibited by biomolecules. Only processes emitting light that

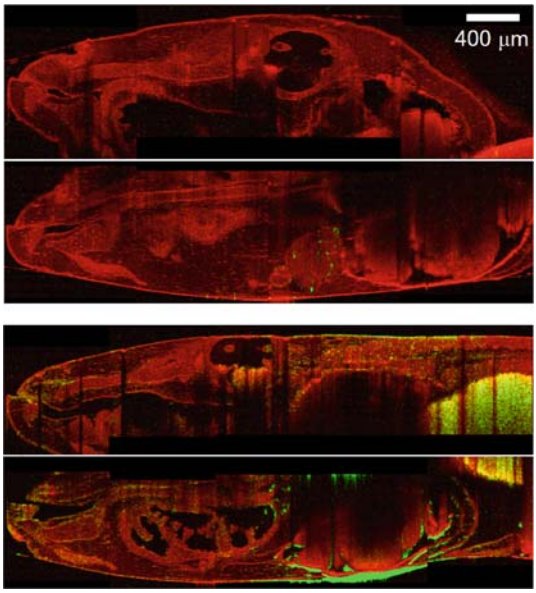


Fig. 15.13. *In vivo* magnetomotive OCT. These composite images show a 3-day-old *Xenopus laevis* tadpole that was exposed to magnetic nanoparticles in its tank water. The *red channel* displays structural OCT data and the *green channel* shows the magnetomotive signal. Significant uptake of magnetic agents is evident in the gastrointestinal tract. Note that the magnetomotive signal in the heart is due to motion artifacts only, as the mean magnetomotive signal in the heart is zero

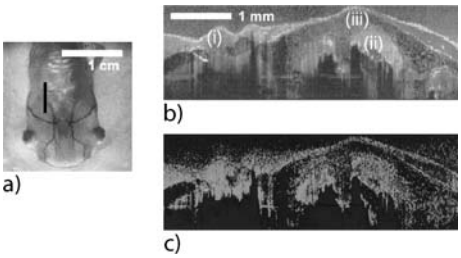


Fig. 15.14. Spectral triangulation in OCT. A *Xenopus laevis* tadpole imaged in the region marked with the *dark line* in (**a**) using OCT (**b**) and molecular contrast agent mapping of indocyanine green (**c**). Evident morphological features include the parabranchial cavity (*i*), gill arches (*ii*), and opercular field (*iii*). Modified figure reprinted with permission [96]

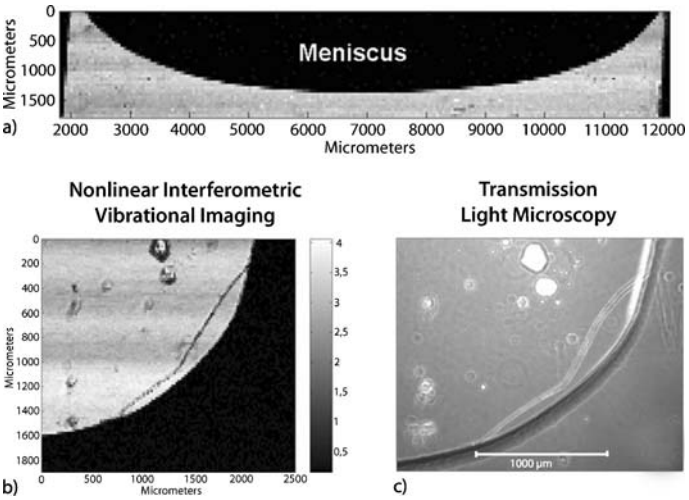


Fig. 15.15. Nonlinear interferometric vibrational imaging (NIVI) images of a quartz cuvette filled with acetone. The imaging system was tuned to a vibrational resonance of 2925 cm^{-1} corresponding to the C–H stretch of acetone. NIVI images of the meniscus formed at the air-acetone interface inside the cuvette (**a**) and the curved bottom of the cuvette (**b**) are shown along with a light microscope image (**c**) showing the visible optical transparency of the sample. Signal is only produced in the acetone, and not in the quartz, air bubbles, or other contaminants. Figure reprinted with permission [91]

remains coherent with the incoming radiation may be used with OCT. **Second harmonic generation (SHG),** wherein a nonlinear material converts two photons of one frequency to one photon of twice that frequency, and **coherent anti-stokes Raman scattering (CARS)**, in which signal is derived from the nonlinear resonances among molecular bonds, are two processes that have been investigated [98–100]. When coupled with the interferometric gating of OCT, these methods may provide molecularly sensitive imaging of three-dimensional tissue structure and function. As shown in Fig. 15.15, an OCT system utilizing a CARS contrast mechanism, which is referred to as **nonlinear interferometric vibrational imaging (NIVI)**, has high sensitivity to specific molecular species [101].

15.4 Applications

In parallel with the development of OCT technology has been the investigation of a wide range of applications in the fields of biology, medicine, and material science. A brief overview of these fields with representative images follows.

15.4.1 Developmental Biology

OCT has been demonstrated in the field of developmental biology as a method to perform high-resolution, high-speed imaging of developing morphology and function. Cellular-level imaging is possible, providing a non-invasive technique for visualizing **cellular processes** such as **mitosis** and **migration**. Imaging studies have been performed on several standard biological animal models commonly employed in developmental biology investigations including embryos and eggs from ***Rana pipiens*** (Leopard frog), ***Xenopus laevis*** (African frog) [102], and ***Brachydanio rerio*** (zebra fish), as well as the murine (mouse) model [103–108].

A series of cross-sectional images acquired *in vitro* from the dorsal and ventral sides of a Stage 49 (12 day) ***Rana pipiens*** **(Leopard frog)** tadpole are shown in Fig. 15.16 [103]. These images were acquired using a SLD source operating at 1.3 μm center wavelength. Axial and transverse resolutions were 16 μm and 30 μm, respectively. Features of internal architectural morphology are clearly visible in the images. The image of the eye differentiates structures corresponding to the cornea, lens, and iris. The corneal thickness is on the order of 10 μm and can be resolved due

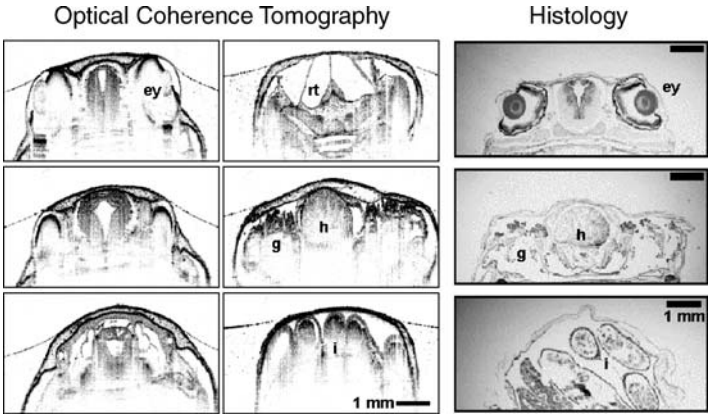


Fig. 15.16. OCT of developing biology. OCT images and corresponding histology acquired from a Stage 49 (12 day) *Rana pipiens* (Leopard frog) tadpole. OCT images in the *left* and *middle columns* were acquired with the OCT beam incident from the *dorsal* and *ventral sides* of the tadpole, respectively. Structures including the eye (ey), respiratory tract (rt), gills (g), heart (h) and gastrointestinal tract (i) are observed. The corresponding histology in the right column correlates strongly with the structures observed using OCT. Figure reprinted with permission [103]

to the differences in index of refraction between the air, cornea, and aqueous humor. By imaging through the transparent lens, the incident OCT beam images several of the posterior ocular layers including the ganglion cell layer, retinal neuroblasts, and choroid. Other identifiable structures include several different regions of the brain as well as the ear vesicle. The horizontal semicircular canal and developing labyrinths can be observed. Internal morphology not accessible in one orientation due to the specimen size or shadowing effects can be imaged by re-orienting the specimen and scanning in the same cross-sectional image plane. The images in the middle column in Fig. 15.16 were acquired with the OCT beam incident from the ventral side to image the respiratory tract, ventricle of the heart, internal gills, and gastrointestinal tract.

Work has also been conducted to assess the development of mouse embryos. Figure 15.17 shows OCT images of a fresh embryo specimen and the correlation to corresponding **histology**. Histological images are acquired by euthanizing the specimen, immersing the specimen in a chemical fixative, and physically sectioning thin (2 – 5 micron-thick) slices using a microtome. The slices are placed on a microscope slide, selectively stained to highlight particular features, and viewed with light microscopy. The correlations between OCT and histology images are strong, suggesting that OCT images can accurately represent the *in vivo* specimen morphology. The potential exists to repeatedly image specimens to quantify organo- and morphogenesis throughout development. Technologies such as OCT are likely to become increasingly important in **functional genomics**, relating genotype to morphology (phenotype) and function in living specimens.

OCT images represent the optical backscatter intensity from regions within the tissue or sample. Because OCT relies on the inherent optical scattering changes to produce image contrast, no exogenous contrast agents or fluorophores are necessary. This permits long-term sequential imaging of development *in vivo* without loss of specimen viability. Repeated imaging of the *Xenopus laevis* hind limb patterning and formation has been demonstrated (Fig. 15.18) without loss of specimen viability [109]. In this example, formation of the cartilaginous skeletal system and the developing vascular system was observed. By imaging subtle differences in backscattering intensity, morphological features millimeters deep within specimens can be clearly delineated.

OCT images used to characterize morphological features within biological specimens show static structures even though they may have been acquired from *in vivo* specimens. *In vivo* imaging in living specimens, particularly in larger organisms and for medical diagnostic applications, must be performed at high speeds to eliminate motion artifacts within the images. **Functional OCT imaging** is the quantification of *in vivo* images which yield information characterizing the functional properties of the organ system or organism. Studies investigating normal and abnormal **cardiac development** have been frequently limited by an inability to access cardiovascular function within the intact organism. OCT has been demonstrated for the high-resolution assessment of structure and function in the developing *Xenopus laevis* (African frog) cardiovascular system and, more recently, in the chick embryo [108].

The morphology of the *in vivo* cardiac chambers can be clearly delineated and image acquisition rates are fast enough to capture the cardiac chambers in mid-cycle. With this capability, images can be acquired and displayed in real-time to produce a movie illustrating the dynamic, functional behavior of the developing heart. Additionally, **four-dimensional imaging**, which captures both three-dimensional structure and movement of the developing heart has been demonstrated in chick embryos (Fig. 15.19). OCT, like technologies such as computed tomography and magnetic resonance imaging, provides high-speed *in vivo* imaging, allowing quantitative assessment of dynamic activity, such as ventricular ejection fraction.

15.4.2 Cellular Imaging

Although previous studies have demonstrated *in vivo* OCT imaging of tissue morphology, most have imaged tissue at $\sim 10-15\,\mu$m resolutions, which does not allow differentiation of cellular structure. The ability of OCT to identify the mitotic activity, the nuclear-to-cytoplasmic ratio, and the migration of cells has the potential to not only impact the fields of cell and developmental biology, but also impact medical and surgical disciplines for the early diagnostics of diseases such as **cancer**.

The ***Xenopus laevis*** (African frog) tadpole has been used to demonstrate the feasibility of OCT for high-resolution *in vivo* cellular and subcellular imaging [26,105]. Many of the cells in this common developmental biology animal model are rapidly dividing and migrating during the early growth stages of the tadpole, providing an

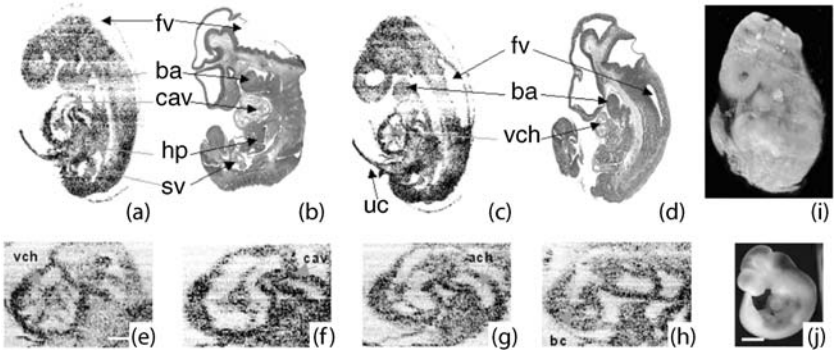


Fig. 15.17. Evaluation of mouse embryo morphology with OCT. Computational sections along sagittal planes extracted from a 3-D OCT dataset of an E10.5 embryo are shown in (**a**) and (**c**) along with corresponding H&E-stained histological sections (**b**, **d**). OCT images of the embryonic heart are shown in (**e**–**h**), with detailed internal structures evident at varying cross-sectional planes. A 3-D volume rendering of the OCT dataset is shown in (**i**) along with a digital photograph of the embryo (**j**). Abbreviations: ba - branchial arch, hp - hepatic primordial, fv - fourth ventricle, sv - subcardinal vein, vch - ventricular chamber, cav - cushion tissue lining the atrio-ventricular canal, ach - atrial chamber, bc - bulbus cordis, and uc - umbilical cord. Scale bar lengths are 200 μm in (**a**–**d**,**i**), 100 μm in (**e**–**h**), and 1 mm in (**j**). Figure reprinted with permission [107]

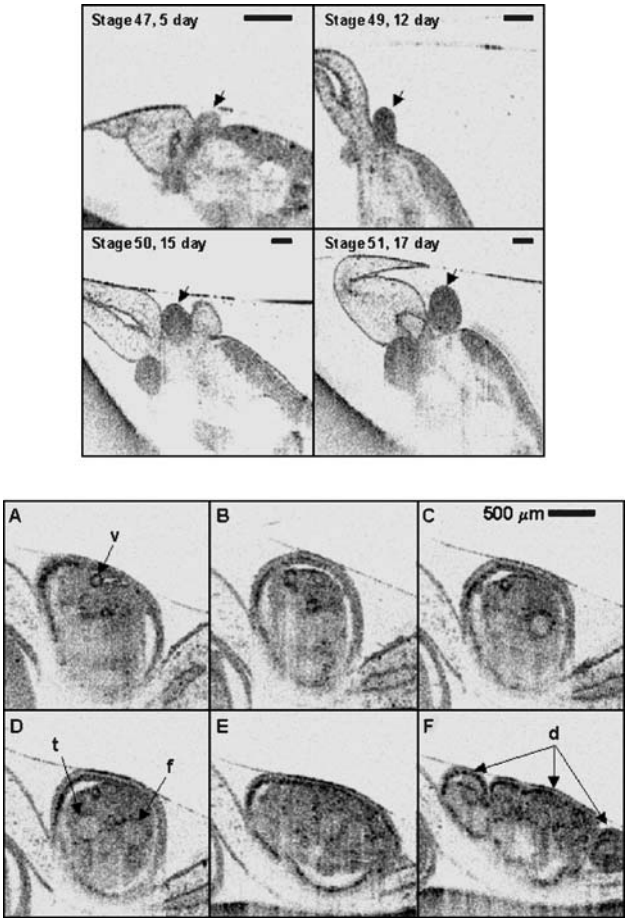


Fig. 15.18. Tracking biological development over time. The top sequence illustrates hind limb development (arrows) in a *Xenopus laevis* tadpole over several weeks. Images were acquired at days 5, 12, 15, and 17. The bottom sequence represents multiple OCT images of a single hind limb acquired at day 28. Images were acquired proximal-to-distal along the length of the hind limb, revealing the vasculature (*v*), the tibia (*t*), the fibula (*f*), and multiple digits (*d*) [109]

opportunity to image dynamic cellular processes. Three-dimensional volumes of high-resolution OCT data have been acquired from these specimens throughout development [105]. From these data sets, cells undergoing mitosis have been identified and tracked in three dimensions (Fig. 15.20). In a similar manner, 3-D data sets have been acquired to track single melanocytes (neural crest cells) as they migrate through the living specimens. The ability of OCT to characterize cellular processes such as mitosis and migration is relevant for cancer diagnostics and the investigation of tumor metastasis in humans.

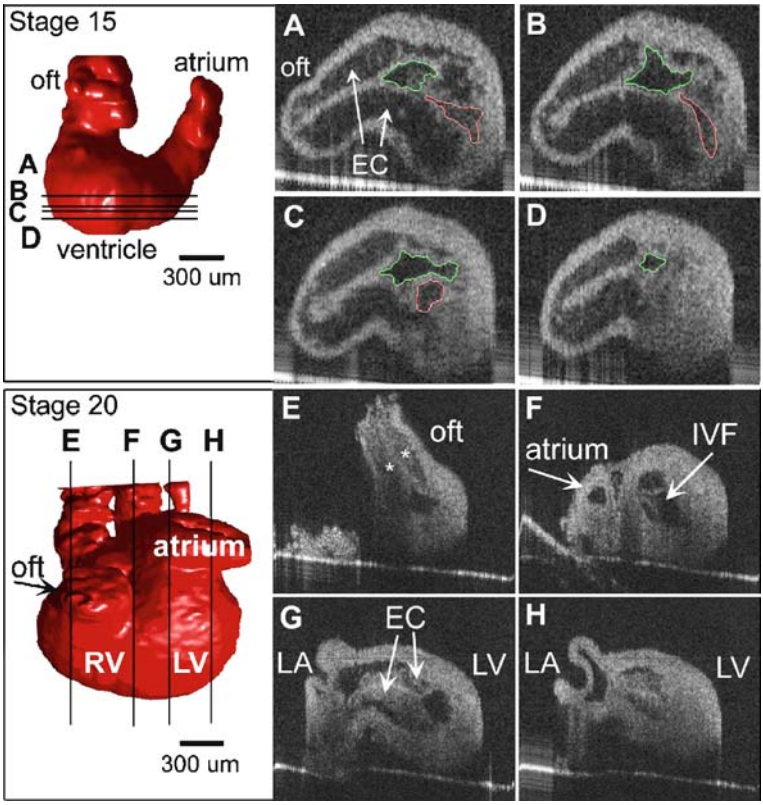


Fig. 15.19. Assessment of cardiac development. OCT images and three-dimensional reconstructions of the embryonic chick heart are shown at two stages of development. The red and green outlines indicate contraction and relaxation and asterisks mark the outflow tract cushions. Abbreviations: EC – endocardial cushions, oft – outflow tract, IVF – interventricular foramen, LA – left atrium, LV – left ventricle, RV – right ventricle. Figure reprinted with permission [108]

An example of cellular-level OCT imaging in these specimens using a state-of-the-art femotosecond laser is shown in Fig. 15.21 [26]. This composite image (0.83×1 mm, 1800×1000 pixels) was acquired using a titanium:sapphire laser with an extremely broad bandwidth (~260 nm). The axial and transverse resolutions for this image are 1 μm and 5 μm, respectively. Because the high transverse resolution reduced the depth-of-focus to 49 μm, separate OCT images were first acquired with the focus at different depths within the specimen. These images were then assembled to produce the composite image shown in Fig. 15.21. This type of image construction is similar to C-mode ultrasound. Cellular features including cell membranes, nuclei, and nuclear morphology are clearly observed up to nearly 1 mm into the specimen. The high-resolution, **optical-sectioning** ability of this system enables 3-D imaging of a single cell, as shown in Fig. 15.22, in which sections were acquired at

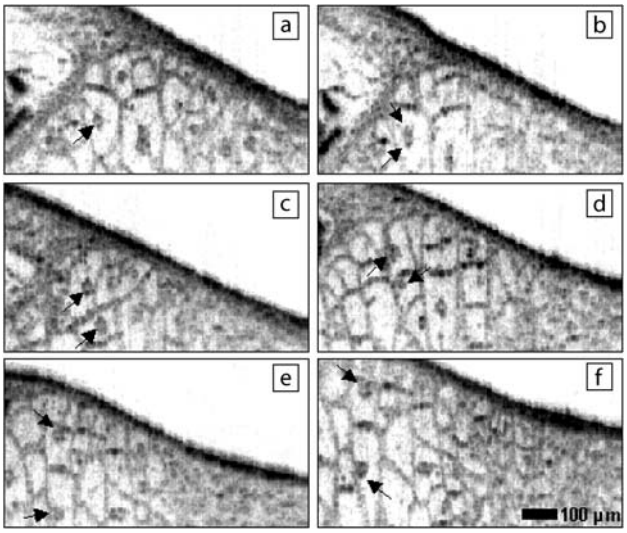


Fig. 15.20. Tracking cell mitosis. OCT imaging of the mesenchymal parent cell in (**a**), which undergoes mitosis, dividing into two daughter cells (**b**–**f**). Following cell division, daughter cells migrate apart and increase in size, preparing for subsequent cell division. Images were extracted from 3-D data sets acquired at 10 min intervals. Figure reprinted with permission [105]

2 micron intervals [110]. Sub-cellular organelles are observed in several of the sections. The images presented here approach the resolutions common in confocal and multi-photon microscopy, but were obtained at greater depths in highly-scattering specimens.

Studies have also been conducted to monitor the migration of cells in engineered tissue scaffolds [111]. Figure 15.23 shows 3-D volume rendered OCT images of a time study of NIH 3T3 fibroblast cells seeded in a porous chitosan scaffold. While OCT is not capable of resolving the features of these cells individually, the dynamics of the cell population are clearly visible from various viewpoints with arbitrary sectioning planes at any angle and direction. These capabilities allow for the investigation of tissue porosity and cell distribution during growth.

15.4.3 Medical and Surgical Microscopy – Identifying Tumors and Tumor Margins

OCT has been used to differentiate between the morphological structure of normal and neoplastic tissue for a wide-range of tumors [21,112–118]. The use of OCT to identify tumors, tumor margins, and metastatic cell groupings *in situ* will represent a significant advancement for medical or **image-guided surgical** applications. OCT has been demonstrated for the detection of **brain** tumors and their margins with normal brain parenchyma, suggesting a role for guiding surgical resection [21].

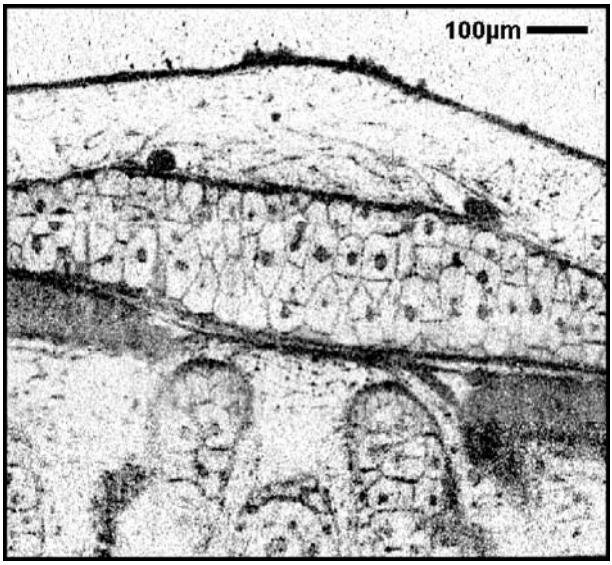


Fig. 15.21. Ultrahigh resolution OCT imaging. Using a titanium:sapphire laser, broad spectral bandwidths enable high axial resolutions. This composite OCT image was acquired with axial and transverse resolutions of 1 and 5 microns, respectively. The high transverse resolution corresponds to a confocal parameter (depth-of-field) of 49 μm. Therefore, multiple images were acquired at increasing depths, then combined to produce this composite image. Figure reprinted with permission [26]

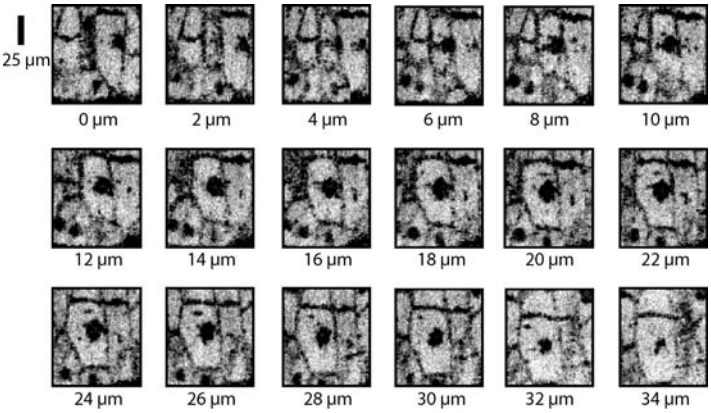


Fig. 15.22. Cell sectioning. At high imaging resolutions, OCT can perform optical sectioning similar to confocal and multi-photon microscopy. Multiple sections at 2 μm intervals are shown for a single *Xenopus laevis* mesenchymal cell. These sections reveal sub-cellular organelles and can be assembled to produce a 3-D data set. Figure reprinted with permission [110]

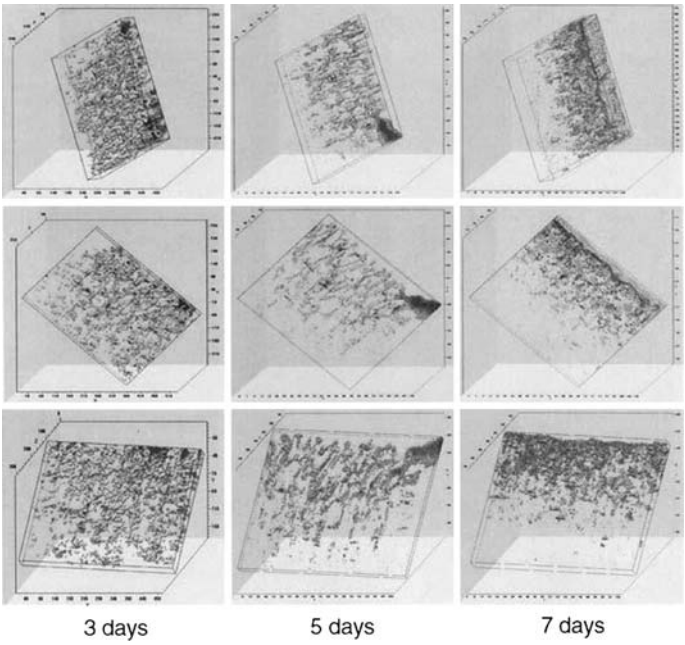


Fig. 15.23. Monitoring engineered tissue. Three-dimensional OCT volumetric images of fibroblast cells seeded in a porous scaffold, as viewed at three rotational angles (*rows*). Tissues culture ages are indicated below each column. Modified figure reprinted with permission [111]

A hand-held surgical imaging probe was constructed for this application. The compact and portable probe permits OCT imaging within the surgical field while the OCT instrument can be remotely located in the surgical suite.

Figure 15.24 shows an *in vitro* specimen of outer human cerebral cortex with metastatic melanoma. The OCT images in Fig. 15.24a and b were acquired through the tumor. These original images were threshold-segmented to identify regions of high backscatter within the tumor. The original images were then overlaid with the segmented data and shown in Fig. 15.24c and d. The OCT images show increased optical backscatter in the region of the tumor (white arrows). Smaller tumor lesions also appear within the image. A shadowing effect is observed below each tumor site due to the increased optical backscatter and the subsequent loss of optical power penetrating beneath the tumor. In Fig. 15.24a and c, the boundary of the tumor can be identified. In Fig. 15.24b and d, the tumor is identified below the surface of normal cortex. The histology in Fig. 15.24e and f confirms the presence and relative sizes of the metastatic tumors.

The system resolution used to acquire the images in Fig. 15.24 was as high as 16 μm, higher than current ultrasound, CT, or MRI intraoperative imaging techniques. This allowed the tumor-cortex interface and the extent of tumor below the surface to be defined with high resolution. At higher imaging resolutions, it may be

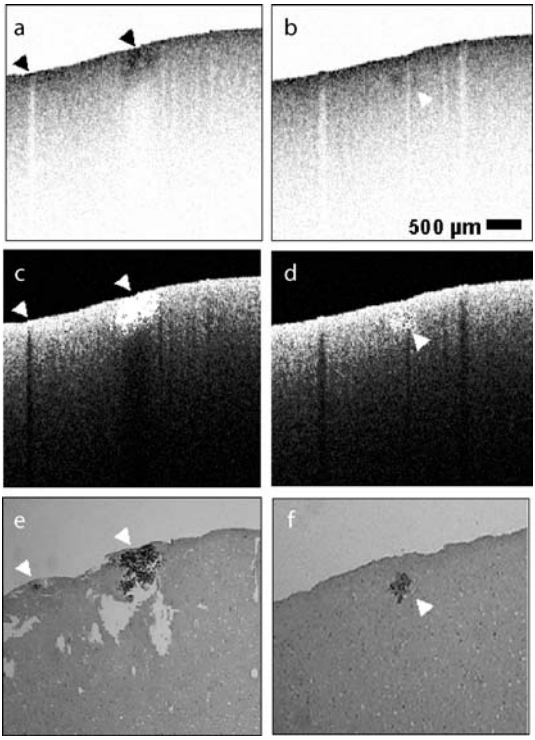


Fig. 15.24. Tumor detection. OCT imaging of metastic melanoma in *in vitro* human cortical brain tissue. Original OCT images (**a**,**b**) were threshold-segmented (**c**,**d**) to highlight backscattering tumor. The correlation with histology (**e**,**f**) is strong. Figure reprinted with permission [21]

possible to image individual tumor cells that have migrated away from the central tumor. OCT, therefore, has the potential to identify tumors and tumor margins on the micron scale and in real-time. OCT also offers imaging performance not achievable with current imaging modalities and may contribute significantly toward the surgical resection of neoplasms.

Investigations have also been made into the use of OCT imaging for **breast** tumor localization. Against the background of healthy fatty tissue and fibrous stroma, breast carcinomas are visible in OCT images due to their increased scattering density. Accurate differentiation, with histological confirmation, has been successfully demonstrated in tumors from animal models and human patients. Potential **breast cancer** applications include biopsy needle guidance, margin detection in open surgery, and the detection of microscopic clusters of metastatic cells in lymph nodes [115–118].

15.4.4 Image-Guided Surgery

The repair of **vessels** and **nerves** is often necessary to restore function following traumatic injury [119]. Although the repair of these sensitive structures is performed with the aid of surgical microscopes and loupes to magnify the surgical field [120], surgeons are limited to the *en face* view that they provide. A technique capable of

subsurface, three-dimensional, micron-scale imaging in real-time would permit the intraoperative monitoring of **microsurgical procedures**. The capabilities of OCT for the intraoperative assessment of microsurgical procedures have been demonstrated [87]. High-speed OCT imaging was integrated with a surgical microscope to perform micron-scale three-dimensional imaging of microsurgical specimens.

The ability of OCT to assess internal structure and vessel patency within an arterial anastomosis (surgical repair) is shown in Fig. 15.25 [87]. Cross-sectional OCT images (2.2×2.2 mm, 250×600 pixel) and 3-D projections of a 1 mm diameter rabbit inguinal artery are illustrated. The two-dimensional cross-sectional images in Fig. 15.25b–e were acquired from the locations labeled in Fig. 15.25a. Figure 15.25b and e were acquired from each end of the anastomosis. The patent lumen is readily apparent. This is in contrast to what is observed in Fig. 15.25d, where the lumen had been occluded by tissue at the site of anastomosis. By assembling a series of cross-sectional 2-D images, a 3-D data set was produced. From this data set, arbitrary planes can be selected and corresponding sections displayed. Figure 15.25f is a longitudinal section from the 3-D data set that confirms the occlusion within the anastomosis site. These results have shown how 2-D OCT images and 3-D OCT projections can provide diagnostic feedback to assess microsurgical anastomoses. This previously unavailable diagnostic ability offers the potential to directly impact and improve patient outcome by incorporating high-speed, high-resolution intraoperative image-guidance during microsurgical procedures.

Surgical intervention requires adequate visualization to identify tissue morphology, precision to avoid sensitive tissue structures, and continuous feedback to monitor the extent of the intervention. The feasibility of OCT to perform image-guided surgical intervention has been investigated [121,122]. High-power **argon laser** (514 nm wavelength) ablation has been used as the representative interventional surgical technique. Figure 15.26 illustrates the use of OCT for **real-time monitoring** of surgical intervention, including **dosimetry** of **laser ablation** [122]. A surgical ablation threshold is documented by a comparison of **thermal injury** in rat kidney and liver for thermal energy doses below and above the ablation threshold. Two sequences with corresponding histology are shown in Fig 15.26. The first exposure of 1 W on *in vitro* rat kidney is halted after 1 s, prior to the ablation threshold. The second exposure of 1 W on *in vitro* rat liver is allowed to continue for 7 s, past the ablation threshold, resulting in ejection of tissue and crater formation. The corresponding histology for the sub-threshold lesion indicates a region of coagulated tissue (arrows). In contrast, the above-threshold lesion histology shows marked tissue ablation and fragmentation within the lesion crater. Below the crater extends a zone of coagulated tissue (arrows) which is not observed in the OCT image due to the poor penetration through the carbonized crater wall.

Although a continuous-wave argon laser was used as the interventional device in this study, the argon laser is only a representative technique for a wide range of instruments and techniques including scalpels, electrosurgery, radiofrequency, microwaves, and ultrasound ablation [123]. OCT imaging was performed at 8 frames per second, fast enough to capture dynamic changes in the optical properties of the

tissue during thermal ablation. These image sequences provide interesting insight into ablation mechanisms for a variety of tissue types. OCT can monitor the extent of thermal injury below the surface of the tissue by imaging the changes in optical backscatter. OCT imaging can, therefore, provide empiric information for dosimetry to minimize the extent of collateral injury. The use of OCT for guiding surgical interventions has the potential to improve intraoperative monitoring and more effectively control interventional procedures.

15.4.5 Materials Investigations

While the majority of OCT applications have been in the fields of biology and medicine, OCT has also been demonstrated in the non-biological areas of materials investigation [124,125], **optical data storage** [126], **microfluidic devices** [127,128], and **tissue engineering scaffolds** [111]. The highly-scattering or reflecting optical properties of many materials prohibit deep imaging penetration using OCT. Many material defects, however, originate at surface or interfacial boundaries, making the use of OCT a possibility for inspection and quality control. For transparent or translucent materials such as plastics or polymer composites, defect imaging at larger depths is feasible.

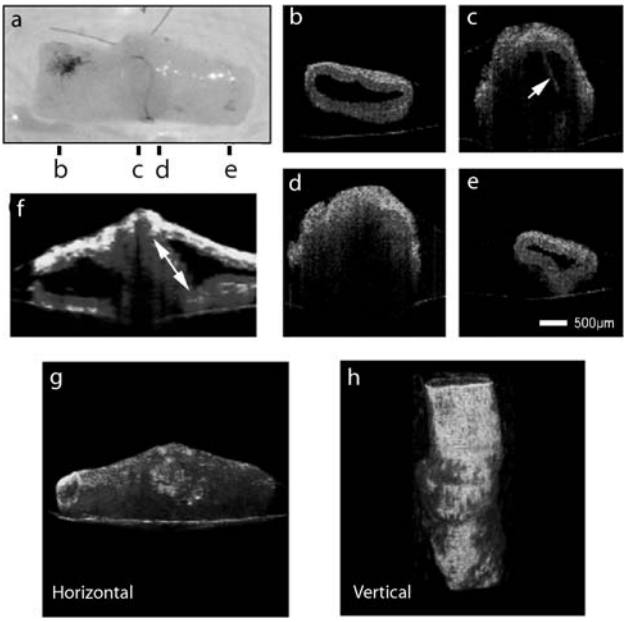


Fig. 15.25. Microsurgical guidance of arterial anastomosis. *Labeled vertical lines* in (**a**) refer to cross-sectional imaging locations (**b**–**e**) from an anastomosis of an *in vitro* rabbit artery. An obstructive flap in (**c**) is marked. By resectioning the 3-D data set, a complete obstruction is observed in (**f**), as indicated by the arrow. Horizontally- and vertically-rotated 3-D projections are shown in (**g**) and (**h**), respectively. Modified figure reprinted with permission [87]

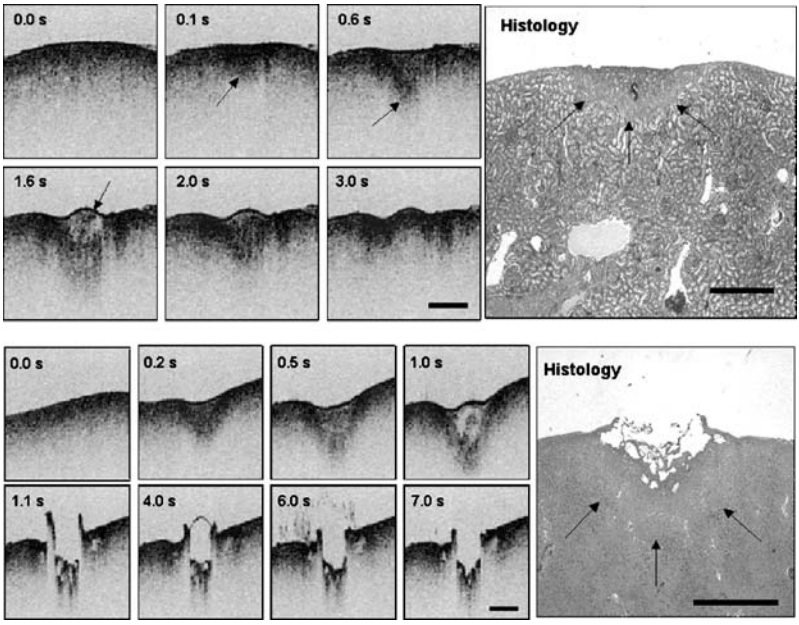


Fig. 15.26. Image-guided surgery. Top OCT image sequence: Argon laser (514 nm, 1 W, 3 s exposure) thermal injury to *in vitro* rat kidney with exposure stopped prior to membrane rupture. Arrows in the corresponding histology indicate zone of thermal damage. Bottom OCT image sequence: Above-threshold argon laser ablation (1 W, 7 s exposure) of rat liver. Arrows in histology image illustrate regions of thermal damage surrounding a region of ablated tissue. *Scale bars* are 1 mm in length. Figures reprinted with permission [122]

Microfluidic mixers, which are used extensively in biochemical processing, rely on rapid and effective mixing mechanisms for proper operation. Most evaluations of the mixing efficiency have been performed by *en face* light microscopy imaging, which is difficult to interpret due to complex mixing geometries. OCT has been used for cross-sectional characterization of the mixing patterns and quantification of critical device parameters [127,128]. Figure 15.27 shows a comparison of light microscopy and OCT images of a vortex mixer. While the mixing efficiency appears high in the light microscope images, the OCT images reveal a far more complex pattern.

The increasing demand for compact, high-density, optical data storage has prompted the investigation of optical sources at shorter visible (blue) wavelengths. The optical ranging capabilities of OCT through scattering materials has the potential for increasing the data storage capacity by assembling multiple layers of optically-accessible data [126,129]. Figure 15.28 illustrates a schematic for the use of OCT in optical data readout as well as images comparing OCT and direct optical detection from a compact disk [126]. The short-coherence length light sources used in OCT can be used for reading high volume density data from multi-layer

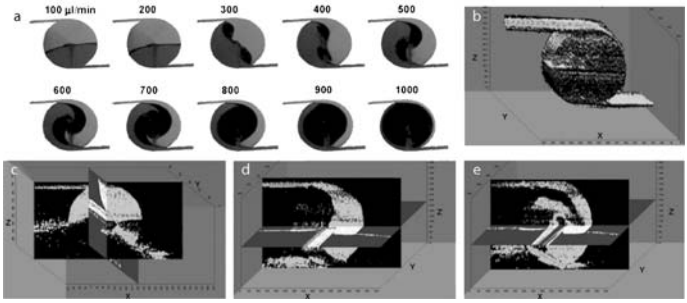


Fig. 15.27. OCT of microfluidic devices. Mixing patterns in a vortex microfluidic mixer are observed with light microscopy (**a**), where inlet flow rates are indicated above, and volume rendered OCT cross-sectional imaging (**b**-**e**). Three-dimensional mixing patterns are shown at flow rates of 100 µl/min (**c**), 500 µl/ min (**d**), and 800 µl/min (**e**). Images are not to scale and have been rotated from the original geometry to show specific features of interest. Figure reprinted with permission [127]

optical disks. Because OCT can share much of the existing hardware and utilize similar scanning techniques present in existing optical data storage system, a high likelihood exists for rapid integration.

15.5 Conclusions

The capabilities of OCT offer a unique and informative means of imaging biological specimens and non-biological samples. The non-contact nature of OCT and the use of low-power near-infrared radiation for imaging make this technique non-invasive and safe, even for extended imaging periods of hours, days, or weeks. OCT relies on the inherent optical contrast generated from variations in optical scattering and index of refraction, but may also be augmented by various exogenous contrast agents, which can target specific tissues of interest or add molecular sensitivity. OCT permits the cross-sectional imaging of tissue and samples and enables *in vivo* structure to be visualized in opaque specimens or in specimens too large for high-resolution confocal or light microscopy.

Imaging at cellular and subcellular resolutions with OCT is an important area of ongoing research. The ***Xenopus laevis*** developmental animal model has been commonly used because its care and handling are relatively simple and it contains cells with a high mitotic index. Cellular imaging in humans, particularly *in situ*, is a challenge because of the smaller cell sizes (10 – 30 µm) compared to larger undifferentiated cells in developing organisms. To the advantage of scientists and physicians, poorly differentiated cells present in many neoplastic tissues tend to be larger, increasing the likelihood for detection using OCT at current imaging resolutions. With further advances in OCT technology, improved discrimination and imaging of more detailed structures should be possible. New laser sources can enhance tissue contrast, improve acquisition rates, and potentially provide functional information since

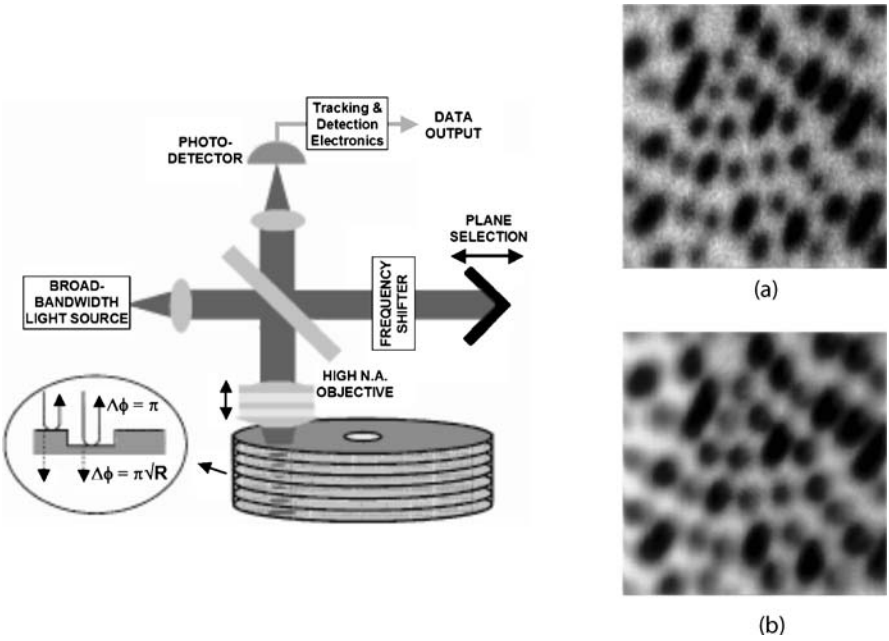


Fig. 15.28. Multi-layer optical data readout. Schematic illustrating the use of OCT for reading optical data from a multi-layer optical disk. Gray-scale images from a CD-ROM layer of optical data acquired with (**a**) OCT and (**b**) direct detection. Modified figure reprinted with permission [126]

tissue scattering and absorbance properties in specimens are wavelength dependent. Short coherence length short pulse laser sources have been used to achieve higher axial resolutions on the order of 1 – 3 microns. Unfortunately, unlike superluminescent diode sources, these high-speed and high-resolution systems have traditionally been relatively large, complex, and costly. Developing compact, portable, turn-key optical sources at near-infrared wavelengths, with broad spectral bandwidths, and with high output powers is an area of ongoing research. Image quality will continue to benefit from techniques such as SD-OCT, swept-source OCT, and OFDI that allow for improved SNR, fast acquisition and volumetric microscopy.

Optical coherence tomography provides high-resolution morphological, functional, and cellular information of biological, medical, and material specimens and samples. OCT represents a multifunctional investigative tool which not only complements many of the existing imaging technologies available today, but over time, is also likely to become established as a major optical imaging modality.

Acknowledgements

We wish to acknowledge our many colleagues that have enthusiastically advanced and utilized OCT technology since its inception. While we have attempted to summarize an ever-expanding body of research and achievement, we regret omissions that were due solely to space limitations. Prof. Stephen Boppart acknowledges the dedicated effort of his students and staff in the Biophotonics Imaging Laboratory at the Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign.

References

1. D. Huang, E.A. Swanson, C.P. Lin, J.S. Schuman, W.G. Stinson, W. Chang, M.R. Hee, T. Flotte, K. Gregory, C.A. Puliafito, J.G. Fujimoto: Science **254**, 1178 (1991)
2. B.E. Bouma, G.J. Tearney, Eds.: *Handbook of Optical Coherence Tomography*. (Marcel Dekker, Inc. 2001)
3. A.F. Fercher, W. Drexler, C.K. Hitzenberger, T. Lasser: Rep. Prog. Phys. **66**, 239 (2003)
4. M.R. Hee, J.A. Izatt, E.A. Swanson, D. Huang, C.P. Lin, J.S. Schuman, C.P. Lin, C.A. Puliafito, J.G. Fujimoto: J. Ophthalmol. **113**, 325 (1995)
5. C.A. Puliafito, M.R. Hee, C.P. Lin, E. Reichel, J.S. Schuman, J.S. Duker, J.A. Izatt, E.A. Swanson, J.G. Fujimoto: Ophthalmology **102**, 217 (1995)
6. C.A. Puliafito, M.R. Hee, J.S. Schuman, J.G. Fujimoto: *Optical Coherence Tomography of Ocular Diseases, 2nd Edition*. (Slack, Inc, Thorofare, NJ 2004)
7. A. Gaudric, B. Haouchine, P. Massin, M. Paques, P. Blain, A. Erginay: Arch. Ophthalmol. **117**, 744 (1999)
8. J.M. Schmitt, A. Knuttel, R.F. Bonner: Appl. Opt. **32**, 6032 (1993)
9. J.M. Schmitt, A. Knuttel, M. Yadlowsky, A.A. Eckhaus: Phys. Med. Biol. **39**, 1705 (1994)
10. J.G. Fujimoto, C. Pitris, S.A. Boppart, M.E. Brezinski: Neoplasia **2**, 9 (2000)
11. J.M. Schmitt, M.J. Yadlowsky, R.F. Bonner: Dermatology **191**, 93 (1995)
12. A.M. Sergeev, V.M. Gelikonov, G.V. Gelikonov, F.I. Feldchtein, R.V. Kuranov, N.D. Gladkova, N.M. Shakhova, L.B. Snopova, A.V. Shakhova, I.A. Kuznetzova, A.N. Denisenko, V.V. Pochinko, Y.P. Chumakov, O.S. Streltzova: Opt. Express 1, **432** (1997)
13. M.E. Brezinski, G.J. Tearney, B.E. Bouma, J.A. Izatt, M.R. Hee, E.A. Swanson, J.F. Southern, J.G. Fujimoto: Circulation **93**, 1206 (1996)
14. G.J. Tearney, M.E. Brezinski, S.A. Boppart, B.E. Bouma, N. Weissman, J.F. Southern, E.A. Swanson, J.G. Fujimoto: Circulation **94**, 3013 (1996)
15. M.E. Brezinski, G.J. Tearney, N.J. Weissman, S.A. Boppart, B.E. Bouma, M.R. Hee, A.E. Weyman, E.A. Swanson, J.F. Southern, J.G. Fujimoto: Heart **77**, 397 (1997)
16. I.K. Jang, B.E. Bouma, D.H. Kang, S.J. Park, S.W. Park, K.B. Seung, K.B. Choi, M. Shishkov, K. Schlendorf, E. Pomerantsev, S.L. Houser, H.T. Aretz, G.J. Tearney: J. Am. Coll. Cardiol. **39**, 604 (2002)
17. G.J. Tearney, M.E. Brezinski, J.F. Southern, B.E. Bouma, S.A. Boppart, J.G. Fujimoto: J. Gastroenterol. **92,** 1800 (1997)
18. J.A. Izatt, M.D. Kulkarni, H-W. Wang, K. Kobayashi, M.W. Sivak: IEEE J. Selected Topics in Quant. Elect. **2**, 1017 (1996)

19. G.J. Tearney, M.E. Brezinski, J.F. Southern, B.E. Bouma, S.A. Boppart, J.G. Fujimoto: J. Urol. **157**, 1915 (1997)
20. E.V. Zagaynova, O.S. Streltsova, N.D. Gladkova, L.B. Snopova, G.V. Gelikonov, F.I. Feldchtein, A.N. Morozov: J. Urol. **167**, 1492 (2002)
21. S.A. Boppart, M.E. Brezinski, C. Pitris, J.G. Fujimoto: Neurosurg. **43**, 834 (1998)
22. B.W. Colston Jr., M.J. Everett, L.B. Da Silva, L.L. Otis, P. Stroeve, H. Nathal: Appl. Opt. **37**, 3582 (1998)
23. X. Clivaz, F. Marquis-Weible, R.P. Salathe: Elec. Lett. **28**, 1553 (1992)
24. B.E. Bouma, G.J. Tearney, S.A. Boppart, M.R. Hee, M.E. Brezinski, J.G. Fujimoto: Opt. Lett. **20**, 1486 (1995)
25. B.E. Bouma, G.J. Tearney, I.P. Bilinski, B. Golubovic, J.G. Fujimoto: Opt. Lett. **21**, 1839 (1996)
26. W. Drexler, U. Morgner, F.X. Kartner, C. Pitris, S.A. Boppart, X.D. Li, E.P. Ippen, J.G. Fujimoto: Opt. Lett. **24**, 1221 (1999)
27. W. Drexler: J. Biomed. Opt. **9**, 47 (2004)
28. B. Povazay, K. Bizheva, A. Unterhuber, B. Hermann, H. Sattmann, A.F. Fercher, W. Drexler, A. Apolonski, W.J. Wadsworth, J.C. Knight, P. St. J. Russell, M. Vetterlein, E. Scherzer: Opt. Lett. **27**, 1800 (2002)
29. W.Y. Oh, S.H. Yun, B.J. Vakoc, G.J. Tearney, B.E. Bouma: Appl. Phys. Lett. **88**, 103902 (2006)
30. Z. Chen, T.E. Milner, S. Srinivas, X. Wang, A. Malekafzali, M.J.C. van Germert, J.S. Nelson: Opt. Lett. **22**, 1119 (1997)
31. J.A. Izatt, M.D. Kulkarni, S. Yazdanfar, J.K. Barton, A.J. Welch: Opt. Lett. **22**, 1439 (1997)
32. J.K. Barton, A.J. Welch, J.A. Izatt: Opt. Express **3**, 251 (1998)
33. J.F. de Boer, T.E. Milner, M.J.C. van Germert, J.S. Nelson: Opt. Lett. **22**, 934 (1997)
34. M.J. Everett, K. Schoenenberger, B.W. Colston Jr., L.B. Da Silva: Opt. Lett. **23**, 228 (1998)
35. K. Schoenenberger, B.W. Colston Jr., D.J. Maitland, L.B. Da Silva, M.J. Everett: Appl. Opt. **37**, 6026 (1998)
36. C. Xu, C. Vinegoni, T.S. Ralston, W. Luo, W. Tan, S.A. Boppart: Opt. Lett. **8**, 1079 (2006)
37. G.J. Tearney, S.A. Boppart, B.E. Bouma, M.E. Brezinski, N.J. Weissman, J.F. Southern, J.G. Fujimoto: Opt. Lett. **21**, 543 (1996)
38. S.A. Boppart, B.E. Bouma, C. Pitris, G.J. Tearney, J.G. Fujimoto, M.E. Brezinski: Opt. Lett. **22**, 1618 (1997)
39. X. Li, C. Chudoba, T. Ko, C. Pitris, J.G. Fujimoto: Opt. Lett. **25**, 1520 (2000)
40. G.J. Tearney, M.E. Brezinski, B.E. Bouma, S.A. Boppart, C. Pitris, J.F. Southern, J.G. Fujimoto: Science **276**, 2037 (1997)
41. B.E. Bouma, G.J. Tearney, C.C. Compton, N.S. Nishioka: Gastrointest. Endosc. **51**, 467 (2000)
42. A. Das, M.V. Sivak Jr., A. Chak, R.C. Wong, V. Westphal, A.M. Rollins, J. Willis, G. Isenberg, J.A. Izatt: Gastrointest. Endosc. **54**, 219 (2001)
43. X.D. Li, S.A. Boppart, J. Van Dam, H. Mashimo, M. Mutinga, W. Drexler, M. Klein, C. Pitris, M.L. Krinsky, M.E. Brezinski, J.G. Fujimoto: Endoscopy **32**, 921 (2001)
44. F.I. Feldchtein, G.V. Gelikonov, V.M. Gelikonov, R.V. Kuranov, A.M. Sergeev, N.D. Gladkova, A.V. Shakhov, N.M. Shakhova, L.B. Snopova, A.B. Terent'eva, E.V. Zagainova, Y.P. Chumakov, I.A. Kuznetzova: Opt. Express **3**, 257 (1998)
45. A.R. Tumlinson, L.P. Hariri, U. Utzinger, J.K. Barton: Appl. Opt. **43**, 113 (2004)

46. K. Takada, I. Yokohama, K. Chida, J. Noda.: Appl. Opt. **26**, 1603 (1987)
47. A.F. Fercher, K. Mengedoht, W. Werner: Opt. Lett. **13**, 186 (1988)
48. C.K. Hitzenberger: Invest. Ophthalmol. Vis. Sci. **32**, 616 (1991)
49. A.M. Rollins, J.A. Izatt: Opt. Lett. **24,** 1484 (1999)
50. B. E. Bouma, G. J. Tearney: Opt. Lett. **24,** 531 (1999)
51. J.F. de Boer, B. Cense, B.H. Park, M.C. Pierce, G.J. Tearney, B.E. Bouma: Opt. Lett. **28**, 2067 (2003)
52. R. Leitgeb, C.K. Hitzenberger, A.F. Fercher: Opt. Express **11**, 889 (2003)
53. M. Wojtkowski, R. Leitgeb, A. Kowalczyk, T. Bajraszewski, A. F. Fercher: J. Biomed. Opt. **7**, 457 (2002)
54. S.R. Chinn, E.A. Swanson, J.G. Fujimoto: Opt. Lett. **22**, 340 (1997)
55. F. Lexer, C.K. Hitzenberger, A.F. Fercher, M. Kulhavy: Appl. Opt. **36**, 6548 (1997)
56. C. Vinegoni, T. Ralston, W. Tan, W. Luo, D.L. Marks, S.A. Boppart: Appl. Phys. Lett. **88**, 053901 (2006)
57. U. Morgner, W. Drexler, F.X. Kartner, X.D. Li, C. Pitris, E.P. Ippen, J.G. Fujimoto: Opt. Lett. **25**, 111 (2000)
58. A.M. Zysk, J.J. Reynolds, D.L. Marks, P.S. Carney, S.A. Boppart: Opt. Lett. **28**, 701 (2003)
59. A.V. Zvyagin, K.K.M.B.D. Silva, S.A. Alexandrov, T.R. Hillman, J.J. Armstrong, T. Tsuzuki, D.D. Sampson: Opt. Express **11**, 3503 (2003)
60. M.L. Osowski, T.M. Cockerill, R.M. Lammert, D.V. Forbes, D.E. Ackley, J.J Coleman: IEEE Photon. Tech. Lett. **6,** 1289 (1994)
61. H. Okamoto, M. Wada, Y. Sakai, T. Hirono, Y. Kawaguchi, Y. Kondo, Y. Kadota, K. Kishi, Y. Itaya: J. Lightwave Tech. **16**, 1881 (1998)
62. B.E. Bouma, L.E. Nelson, G.J. Tearney, D.J. Jones, M.E. Brezinski, J.G. Fujimoto: J. Biomed. Opt. **3**, 76 (1998)
63. M. Bashkansky, M.D. Duncan, L. Goldberg, J.P. Koplow, J. Reintjes: Opt. Express **3**, 305 (1998)
64. K. Bizheva, B. Povazay, B. Hermann, H. Sattmann, W. Drexler, M. Mei, R. Holzwarth, T. Hoelzenbein, V. Wacheck, H. Pehamberger: Opt. Lett. **28**, 707 (2003)
65. D.L. Marks, A.L. Oldenburg, J.J. Reynolds, S.A. Boppart.: Opt. Lett. **27**, 2010 (2002)
66. Y. Wang, I. Tomov, J.S. Nelson, Z. Chen, H. Lim, F. Wise: J. Opt. Soc. Am. A **22**, 1492 (2005)
67. S. Bourquin, A.D. Aguirre, I. Hartl, P. Hsiung, T.H. Ko, J.G. Fujimoto, T.A. Birks, W.J. Wadsworth, U. Bünting, D. Kopf: Opt. Express **11**, 3290 (2003)
68. G. Humbert, W.J. Wadsworth, S.G. Leon-Saval, J.C. Knight, T.A. Birks, P. St. J. Russell, M.J. Lederer, D. Kopf, K. Wiesauer, E.I. Breuer, D. Stifter: Opt. Express **14**, 1596 (2006)
69. W. Denk, J.H. Strickler, W.W. Webb: Science **248**, 73 (1990)
70. D.W. Piston, M.S. Kirby, H. Cheng, W.J. Lederer, W.W. Webb: Appl. Opt. **33**, 662 (1994)
71. E. Beaurepaire, L. Moreaux, F. Amblard, J. Mertz: Opt. Lett. **24**, 969 (1999)
72. S.H. Yun, C. Boudoux, G.J. Tearney, B.E. Bouma: Opt. Lett. **28**, 1981 (2003)
73. W.Y. Oh, S.H. Yun, G.J. Tearney, B.E. Bouma: Opt. Lett. **30**, 3159 (2005)
74. R. Huber, M. Wojtkowski, J.G. Fujimoto: Opt. Express **14**, 3225 (2006)
75. I. Daubechies: IEEE Trans. Inf. Theory **36**, 961 (1990)
76. R. Leitgeb, M. Wojtkowski, A. Kowalczyk, C.K. Hitzenberger, M. Sticker, A.F. Fercher: Opt. Lett. **25**, 820 (2000)
77. C. Xu, P.S. Carney, S.A. Boppart: Opt. Express **13**, 5450 (2005)

78. K.F. Kwong, D. Yankelevich, K.C. Chu, J.P. Heritage, A. Dienes: Opt. Lett. **18**, 558 (1993)
79. L. Giniunas, R. Danielius, R. Karkockas: Appl. Opt. **38**, 7076 (1999)
80. M. Lai: Engineering & Laboratory Notes, Supplement to Optics & Photonics News **13**, 10 (2002)
81. A.G. Podoleanu, J.A. Rogers, D.A. Jackson, S. Dunne: Opt. Express **7**, 292 (2000)
82. S. Bourquin, V. Monterosso, P. Seitz, R.P. Salathe: Opt. Lett. **25**, 102 (2000)
83. J.A. Rogers, A.G. Podoleanu, G.M. Dobre, D.A. Jackson, F.W. Fitzke: Opt. Express **9**, 533 (2001)
84. J.A. Izatt, M.R. Hee, G.M. Owen, E.A. Swanson, J.G. Fujimoto: Opt. Lett. **19**, 590 (1994)
85. B.M. Hoeling, A.D. Fernandez, R.C. Haskell, E. Huang, W.R. Myers, D.C. Petersen, S.E. Ungersma, R. Wang, M.E. Williams: Opt. Express **6**, 136 (2000)
86. C.G. Rylander, D.P. Dav, T. Akkin, T.E. Milner, K.R. Diller, A.J. Welch: Opt. Lett. **29**, 1509 (2004)
87. S.A. Boppart, B.E. Bouma, C. Pitris, G.J. Tearney, J.F. Southern, M.E. Brezinski, J.G. Fujimoto: Radiology **208**, 81 (1998)
88. Y. Pan, H. Xie, G. K. Fedder: Opt. Lett. **26**, 1966 (2001)
89. A.R. Tumlinson, L.P. Hariri, U. Utzinger, J.K. Barton: Appl. Opt. **43**, 113 (2004)
90. J. Wu, M. Conry, C. Gu, F. Wang, Z. Yaqoob, C. Yang: Proc. of the SPIE **6079**, 1J (2006)
91. S.A. Boppart, A.L. Oldenburg, C. Xu, D.L. Marks: J. Biomed. Opt. **10**, 041208 (2005)
92. J.K. Barton, J.B. Hoying, C.J. Sullivan: Acad. Radiol. **9S**, 52 (2002)
93. T.M. Lee, A.L. Oldenburg, S. Sitafalwalla, D.L. Marks, W. Luo, F. Jean-Jacques Toublan, K.S. Suslick, S.A. Boppart: Opt. Lett. **28**, 1546 (2003)
94. A.L. Oldenburg, J.R. Gunther, S.A. Boppart: Opt. Lett. **30**, 747 (2005)
95. A.L. Oldenburg, F. Jean-Jacques Toublan, K.S. Suslick, A. Wei, S.A. Boppart: Opt. Express **13**, 6597 (2005)
96. C. Yang, L.E.L. McGuckin, J.D. Simon, M.A. Choma, B.E. Applegate, J.A. Izatt: Opt. Lett. **29**, 2016 (2004)
97. K.D. Rao, M.A. Choma, S. Yazdanfar, A.M. Rollins, J.A. Izatt: Opt. Lett. **28**, 340 (2003)
98. Y. Jiang, I. Tomov, Y. Wang, Z. Chen: Opt. Photonics News **15**, 22 (2004)
99. C. Vinegoni, J.S. Bredfeldt, D.L. Marks, S.A. Boppart: Opt. Express **12**, 331 (2004)
100. S. Tang, T.B. Krasieva, Z. Chen, B.J. Tromberg: J. Biomed. Opt. Lett. **11**, 020502 (2006)
101. D.L. Marks, S.A. Boppart: Phys. Rev. Lett. **92**, 123905 (2004)
102. P.D. Nieuwkoop, J. Faber: *Normal Table of Xenopus Laevis* (Garland Publishing, Inc., New York 1994)
103. S.A. Boppart, M.E. Brezinski, B.E. Bouma, G.J. Tearney, J.G. Fujimoto: Dev. Biol. **177**, 54 (1996)
104. S.A. Boppart, M.E. Brezinski, G.J. Tearney, B.E. Bouma, J.G. Fujimoto: J. Neurosci. Methods **70**, 65 (1996)
105. S.A. Boppart, B.E. Bouma, C. Pitris, J.F. Southern, M.E. Brezinski, J.G. Fujimoto: Nat. Med. **4**, 861 (1998)
106. Q. Li, A.M. Timmers, K. Hunter, C. Gonzalez-Pola, A.S. Lewin: Invest. Ophthalmol. Vis. Sci. **42**, 2981 (2001)
107. W. Luo, D.L. Marks, T.S. Ralston, S.A. Boppart: J. Biomed. Opt. **11**, 021014 (2006)
108. M.W. Jenkins, F. Rothenberg, D. Roy, V.P. Nikolski, Z. Hu, M. Watanabe, D.L. Wilson, I.R. Efimov, A.M. Rollins: Opt. Express **14**, 736 (2006)

109. S.A. Boppart: Surgical Diagnostic, Guidance, and Intervention Using Optical Coherence Tomography. PhD Thesis, Massachusetts Institute of Technology, Cambridge, MA (1998)
110. S.A. Boppart, W. Drexler, U. Morgner, F.X. Kartner, J.G. Fujimoto: Ultrahigh Resolution and Spectroscopic Optical Coherence Tomography Imaging of Cellular Morphology and Function. In: *Proc. Inter-Institute Workshop on In Vivo Optical Imaging at the National Institutes of Health.* ed. by A.H. Gandjbakhche, 56 (1999)
111. W. Tan, A. Sendemir-Urkmez, L.J. Fahrner, R. Jamison, D.L. Leckband, S.A. Boppart: Tissue Eng. **10**, 1747 (2004)
112. A.M. Sergeev, V.M. Gelikonov, G.V. Gelikonov, F.I. Feldchtein, R.V. Kuranov, N.D. Gladkova, N.M. Shakhova, L.B. Snopova, A.V. Shakov, I.A. Kuznetzova, A.N. Denisenko, V.V. Pochinko, Y.P. Chumakov, O.S. Streltzova: Opt. Express **1**, 432 (1997)
113. S.A. Boppart, A.K. Goodman, C. Pitris, C. Jesser, J.J. Libis, M.E. Brezinski, J.G. Fujimoto: Brit. J. Obstet. Gyn. **106**, 1071 (1999)
114. C. Pitris, A.K. Goodman, S.A. Boppart, J.J. Libus, J.G. Fujimoto, M.E. Brezinski: Obstet. Gynecol. **93**, 135 (1999)
115. N.V. Iftimia, B.E. Bouma, M.B. Pitman, B. Goldberg, J. Bressner, G.J. Tearney: Rev. Sci. Inst. **76**, 064301 (2005)
116. S.A. Boppart, W. Luo, D.L. Marks, K.W. Singletary: Breast Cancer Res. Treat. **84**, 85 (2004)
117. A.M. Zysk, S.A. Boppart: J. Biomed. Opt., **11**, 054015 (2006)
118. W. Luo, F.T. Nguyen, A.M. Zysk, T.S. Ralston, J. Brockenbrough, D.L. Marks, A.L. Oldenburg, S.A. Boppart: Technol. Cancer Res. Treat. **4**, 539 (2005)
119. C. Zhong-wei, Y. Dong-yue, C. Di-Sheng: *Microsurgery*. (Springer Verlag, New York, NY 1982)
120. M.D. Rooks, J. Slappey, K. Zusmanis: Microsurgery **14**, 547 (1993)
121. M.E. Brezinski, G.J. Tearney, S.A. Boppart, E.A. Swanson, J.F. Southern, J.G. Fujimoto: J. Surg. Res. **71**, 32 (1997)
122. S.A. Boppart, J.M. Herrmann, C. Pitris, D.L. Stamper, M.E. Brezinski, J.G. Fujimoto: J. Surg. Res. **82**, 275 (1999)
123. S.A. Boppart, J.M. Herrmann, C. Pitris, D.L. Stamper, M.E. Brezinski, J.G. Fujimoto: Comput. Aided Surg. **6**, 94 (2001)
124. M.D. Duncan, M. Bashkansky, J. Reintjes: Opt. Express **2**, 540 (1998)
125. F. Xu, H.E. Pudavar, P.N. Prasad, D. Dickensheets: Opt. Lett. **24**, 1808 (1999)
126. S.R. Chinn, E.A. Swanson: Opt. Lett. **21**, 899 (1996)
127. C. Xi, D.L. Marks, D.S. Parikh, L. Raskin, S.A. Boppart: Proc. Natl. Acad. Sci. **101**, 7516 (2004)
128. Y.-C. Ahn, W. Jung, J. Zhang, Z. Chen: Opt. Express **13**, 8164 (2005)
129. S.R. Chinn, E.A. Swanson: Opt. Memory Neural Networks **5**, 197 (1996)
130. A.M. Zysk, S.G. Adie, J.J. Armstrong, M.S. Leigh, A. Paduch, D.D. Sampson, S.A. Boppart: Opt. Lett. **32**, 385 (2007)
131. T.S. Ralston, D.L. Marks, P.S. Carney, S.A. Boppart: Nat. Phys. **3**, 129 (2007)

16 Near-Field Optical Microscopy and Application to Nanophotonics

Motoichi Ohtsu

16.1 Introduction

Near-field optics deals with the local electromagnetic interaction between optical field and matters in a nanometric scale. Due to the size-dependent localization and size-dependent resonance, optical near-field is completely free from diffraction of light. The initial idea of near field optics as a super-resolution imaging tool, in a somewhat primitive form, has been proposed in 1928 [1]. After about half a century of silence, basic research has been started early 1980's by several groups in Europe, US and Japan almost independently. A short comment has been written in the author's laboratory notebook on February 26, 1982, pointing out that *the chemical etching of fiber for fabricating a probe is not straightforward, and thus, requires a continuous effort to establish a nano-fabrication technology* [2]. However, after continuous efforts, study of near field optics showed a lot of progress. For example, the single string of DNA molecule has been successfully imaged. The linewidth of the image has been as narrow as 4 nm or even less [3]. As a result of this progress it can be claimed that a new field, the so called "nano-photonics" has been started as a basic technology to support the society of the 21st century. Further, a novel field of "atom-photonics" has also been started, which controls the thermal and quantum motions of atoms precisely by the optical near-field [4,5].

Three examples of technical trends and technical problems for the present and future optical industry are presented here. (1) It has been estimated that the society of the year of 2010 requires the high-density optical storage and readout technology, with the storage density of 1 Tb/in^2 and data transmission rate of 1 Gbps. Since the mark-length for 1 Tb/in^2 is as short as 25 nm, writing such a small pit is far beyond the diffraction limit of light. (2) The progress of DRAM technology requires the drastic improvement of photo-lithography. It is estimated that the size of the fabricated pattern should be as narrow as 50 nm or even narrower. Though novel methods using excimer lasers, EUV, and SR light sources have been developed, this required linewidth is far beyond the diffraction limit of the methods using a conventional visible light source. (3) The progress of optical fiber transmission system requires the improvement of the photonic integration technology. It is estimated that the photonic matrix switching device should be sub-wavelength in its size in order to realize more than 1000×1000 element integration of the devices for the 10 Tbps switching capability in the year of 2015. Since the conventional photonic devices, e.g., diode lasers and optical waveguides have to confine the light in them, their minimum sizes

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

are limited by the diffraction of light. Thus, the size required above is beyond the diffraction limit.

By the three examples shown above, it can be easily understood that a novel optical nanotechnology is required to go beyond the diffraction limit to support the science and technology of the 21st century. Near field optics can be used as such a novel one to meet this requirement. This paper reviews the recent progress of author's works on near-field optical microscopy and its application to chemical vapor deposition, nano-photonic switching, and optical storage/readout by optical near-field for nano-photonics.

16.2 Nano-Scale Fabrication

There has been an interest in application to nano-structure fabrication because of the possibility of realizing nano-photonic integration [6]. For the realization of a device which uses the optical near-field as a carrier for signal transmission, various materials with nanometric size must be integrated laterally on a substrate. For this integration, we need an advanced nano-structure fabrication technique, able to realize spatially high resolution, precise control of size and position, and be applicable for various materials. Photo-enhanced chemical vapor deposition (PE-CVD) method has been attracting attention. Based on photochemical reaction, PE-CVD offers not only the possibility for the lateral integration of different in a truly single growth run [7,8] but also the option of chemical selective growth by varying the wavelength of the light source used. Since the optical near-field energy is concentrated within nanometric dimensions smaller than the wavelength of light [9,10], we are able to deposit various materials of nanometric dimensions by utilizing the photo-decomposition of chemical gases. The combination of optical near-field technology with the PE-CVD, i.e., a near-field optical CVD (NFO-CVD) thus appears to be the most suitable technology for the integration of nanometer scale elements, because it not only allows us to fabricate nanostructures but also gives dual advantage of *in-situ* measurement of the optical properties of the fabricated nanostructures.

16.2.1 Depositing Zinc and Aluminum

Let us discuss the deposition of Zn as an example of NFO-CVD by using a diethlzinc (DEZ) as a parent gas. As NFO-CVD is based on a photodissociation reaction, it is necessary for a reactant molecule to absorb photons with a higher energy than its dissociation energy. This molecule absorbs light with a photon energy higher than 4.59 eV ($\lambda < 270$ nm), and the photodissociation reaction occurs as: $Zn(C_2H_5)_2 + 2.256\,eV \rightarrow ZnC_2H_5 + C_2H_5$, and $ZnC_2H_5 + 0.9545\,eV \rightarrow Zn + C_2H_5$. Thus the second harmonic light (SH light) of an Ar^+ laser ($\lambda = 244$ nm) and an ArF excimer laser ($\lambda = 193$ nm) were used as the light source for photodissociation of DEZ gas. For the purpose of generating the UV optical near-field

with sufficient power density to decompose DEZ gas, an UV fiber with transmission loss as low as 1.1 dB/m at 244 nm was used to fabricate a probe [11]. An UV probe was coated with 200 nm thick Al film after being tapered by chemical etching. The throughput of the probe was 1×10^{-4} and the power density at the probe tip with a sub-100 nm aperture for 1 mW of incident UV light was as high as 1 kW/cm^2. Figure 16.1a, b shows two methods of NFO-CVD, i.e., direct gas phase photodissociation method and prenucleation method, respectively. The advantage of the direct gas phase photodissociation method Fig.16.1a lies in the possibility of selective deposition of various materials by changing parent gases, which is useful for lateral integration [12]. One disadvantage of this method is that the probe tip is also gradually covered with the depositing materials while it is fabricating a nano-structure on the substrate. In our experience, however, this only becomes a problem after a few hours of operation, while only a few seconds are necessary to fabricate a nano-structure. It is therefore not a serious problem. In order to examine the effect of optical near-field energy on the growth, an experiment was performed by varying the illumination time at a constant gas pressure of 1 mTorr and an input SH light power of 15 mW. The value of the optical near-field energy was evaluated by measured SH light power × illumination time. It was confirmed that the dot size depends on the spatial distribution of the optical near-field in a direction lateral to the substrate surface, while depending on the optical near-field energy in the normal direction. One of the most attractive features of this technique is its high spatial resolution. The lateral size of the fabricated pattern depends on the spatial distribution of the optical near-field energy and its reproducibility also depends on the

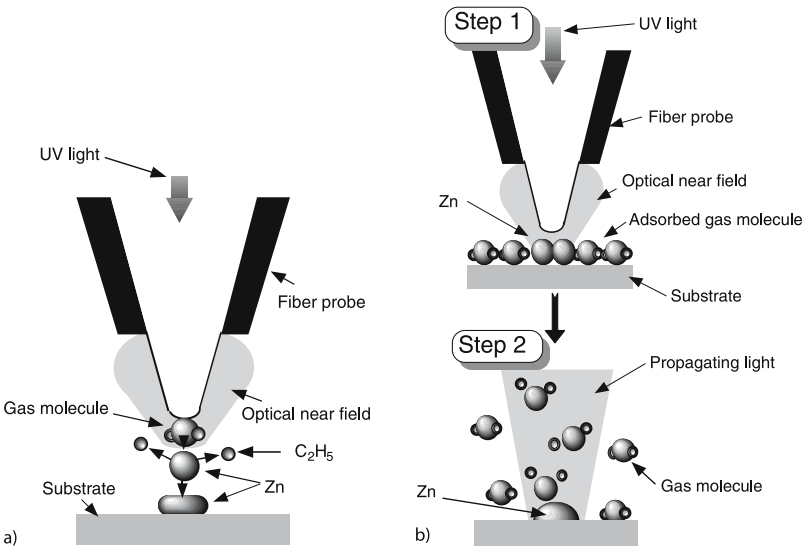


Fig. 16.1. Principles of NFO-CVD. (**a**) Gas phase direct photodissociation method. (**b**) Prenucleation method

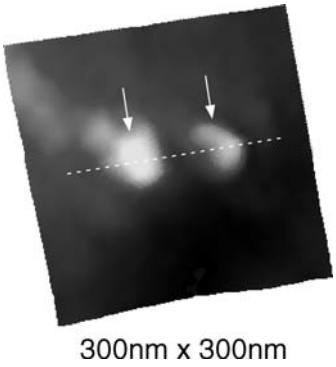


Fig. 16.2. Shear-force image of Zn dots

reproducibility of fabricating probes. Figure 16.2 shows the shear-force image of dots. Two dots with a diameter of 52 nm and 37 nm (FWHM of the cross-sectional profile) were fabricated at a very close distance of 45 nm. Since the measured diameter of the dot image includes the resolution of a vacuum shear-force microscope(VSFM) [13], the intrinsic diameter may be smaller than the value estimated from these figures.

Figure 16.3 shows the shear-force image of the loop-shaped Zn pattern on a glass substrate produced by the prenucleation method Fig. 16.1b [14]. The vacuum chamber was evacuated to below 10^{-5} Torr prior to the prenuclei fabrication stage, then filled with about 10 Torr of DEZ gas, maintaining the pressure for 20 min. Next, the chamber was re-evacuated to the pressure of 10^{-5} Torr, leaving a few adsorbed monolayer on the substrate surface. Prenucleation was performed by delivering the SH light on the substrate covered with adsorbed molecules using a probe. Nuclei of Zn were formed by the decomposition of DEZ gas adsorbed on the substrate

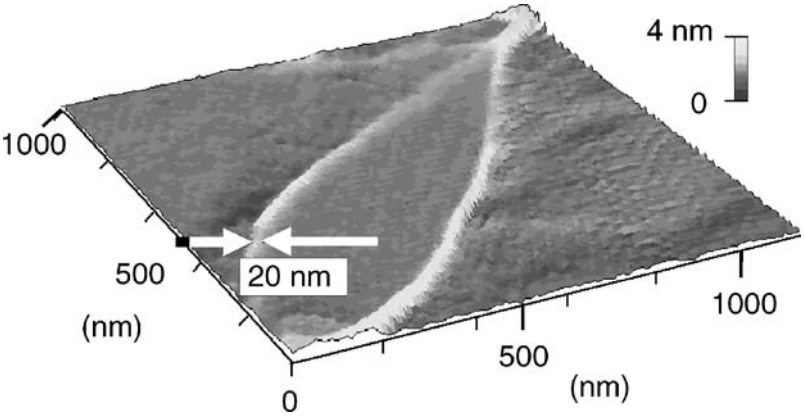


Fig. 16.3. Shear-force image of a loop-shaped Zn deposited on a glass substrate by the prenucleation method

with the optical-near field at the probe tip. In the growth stage after nuclei fabrication, DEZ gas was refilled in the chamber with a few Torr and the unfocused ArF excimer laser with the maximum energy of 10 mJ was directly irradiated on the prenucleated substrate. Then growth proceeded only on the pre-existing nuclei. As this figure shows, the minimum width of the pattern is as small as 20 nm. The width achieved here is two orders smaller than the minimum width reported so far by conventional PE-CVD using a far-field light [15]. Since the measured width includes the resolution of VSFM, the intrinsic width can be smaller than the value estimated from this figure. This method has the advantage of being free from the deposition at the probe tip, but otherwise has the drawback that the lateral integration of various materials is not easy due to using the propagating light in the second step. for the first time by NFO-CVD of nanometer-sized Zn dots with nonresonant light [16]. In order to investigate the deposition effect by nonresonant light, the fiber probe without metal coating on it, i.e., a bare fiber probe, was used for deposition. Therefore, the far-field light leaked to circumference of the fiber probe. The solid curve in Fig. 16.4 shows the cross-sectional profile of a Zn dot on the sapphire step substrate deposited by using the 244 nm-wavelength light, i.e., the light resonant to the absorption spectrum. It has tails of 4 nm height on both side of the dot. The tails correspond to the deposition by the leaked far-field light. The broken curve in this figure shows the profile of the dot deposited by using the light of $\lambda = 488$ nm, i.e., the nonresonant light. This curve has no tails since, in the case of the far-field PE-CVD, Zn deposition using the nonresonant light is not possible [17]. We discuss the possible mechanism of DEZ dissociation and deposition by the nonresonant optical near-field. In the case of using the far-field light, the dissociation of DEZ is induced by photo-absorption and predissociation. In contrast, in the case of using the optical near-field, the dissociation can take place even under the nonresonant condition. Its first reason is two-photon absorption process due to the high energy density of the optical near-field at the apex of the high throughput fiber probe. The

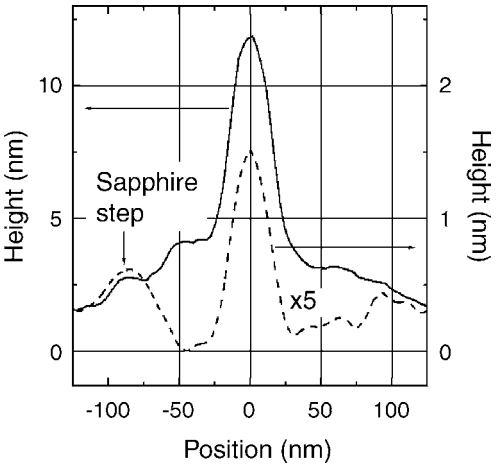


Fig. 16.4. Cross sectional profiles of Zn dots deposited on a sapphire step substrate deposited by the light of $\lambda =$ 244 nm (*solid curve*) and $\lambda = 488$ nm (*broken curve*)

second is the induced transition to the dissociation channel by the apex of the fiber. The third involved the direct coupling between the optical near-field and the dissociation molecular vibration mode. This nonresonant deposition method has high potential for the application to the NFO-CVD of gas sources for which far-field PE-CVD is unrealizable. NFO-CVD also allows us to fabricate nanostructures of several metals (Al, Cr, W, etc.) as well as Zn. For example, Fig. 16.5 shows shear-force image of five Al dots deposited on a sapphire substrate by dissociating the trimethylalminium (TMA) using the light of $\lambda = 244$ nm [18]. The TMA partial pressure and the light power incident into the fiber probe were 10 mTorr and 1 mW, respectively. The FWHM of the dots are typically 25 nm, which is comparable to the apex diameter of the used probe. As demonstrated by the deposition of Zn and Al, various materials can be deposited selectively by only changing the reactant gases. Furthermore, there is no limitation in regard to substrate and deposited materials. Zn and Al were deposited on an insulator substrate like sapphire. These are the advantage of this technique. Another of the most attractive points of this technique

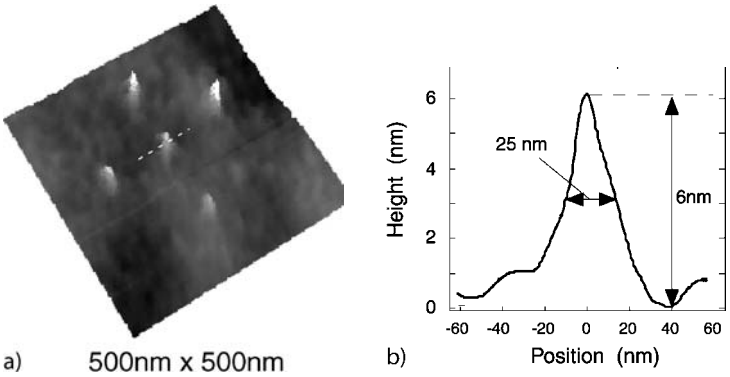


Fig. 16.5. Shear-force image of five Al dots deposited on a sapphire substrate

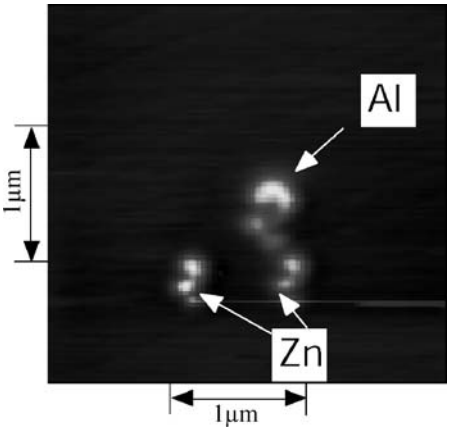


Fig. 16.6. Zn and Al dots deposited closely on a sapphire substrate

is the possibility for in-situ lateral integration of nano-scale structures of different materials with ease, which is difficult in conventional technique. In order to demonstrate the possibility, Zn dots were deposited at the first step, under the DEZ pressure of 10 mTorr. As the second step, Al dots were deposited under TMA pressure of 10 mTorr after evacuating the deposition chamber to 5×10^{-6} Torr. Figure 16.6 shows the result, i.e., shear-force image of deposited Zn and Al dots at closely position on a sapphire substrate. The distance between the Zn and Al dots was as close as 100 nm.

16.2.2 Depositing Zinc Oxide

One of the advantages of this NFO-CVD is the fact that there is no limitation in regard to substrate and deposited materials. That is, nanostructure of oxides, insulators, and semiconductors are deposited as well as metals. As an example, let us demonstrate the nanofabrication by NFO-CVD of ZnO on a sapphire substrate as insulator [19]. As a preliminary experiment, ZnO films were deposited by the PE-CVD method by using a propagating far-field light. Here, we used the reaction between oxygen and DEZ conveyed by a carrier gas (Ar) into the chamber during the irradiation by the propagating SH light (λ = 244 nm) of an Ar^+ laser. A (0001) sapphire was used as a substrate for the epitaxial growth of ZnO. The reaction chamber was filled with the reactant gases at the ratio of DEZ : O of 1 : 10 at a working pressure of 10 mTorr. The chamber pressure was maintained at 10 mTorr during the growth. In order to find the optimal growth conditions, the crystallinity, stoichiometry, optical transmission, and photoluminescence were evaluated. The PE-CVD of ZnO was carried out for 10 min within a range of the substrate temperature from room temperature to 300 °C. The energy density of the laser source and the spot size were 10 mW and 600 μm, respectively. Crystalline films were deposited at substrate temperatures over 100 °C and the films with c-axis oriented crystalline. For films deposited at a substrate temperature above 150 °C, the atomic ratio of Zn : O

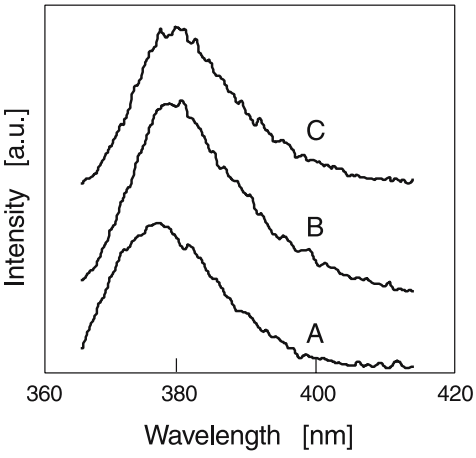


Fig. 16.7. Photoluminescence spectra of ZnO films deposited at the substrate temperature of 150 °C (A), 200 °C (B), and 300 °C (C). They were measured at room temperature

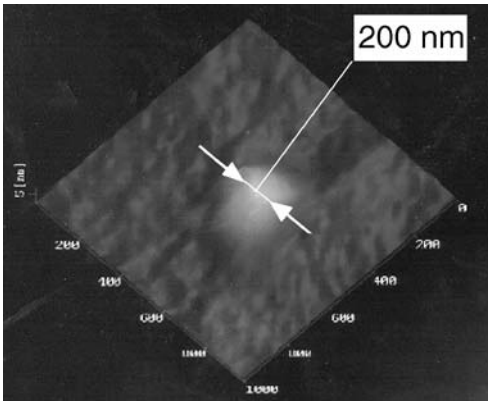


Fig. 16.8. Shear-force image of a ZnO dot deposited on the (0001) sapphire substrate by NFO-CVD

was 1.00 : 1.00 within an accuracy of a few percent. The optical properties were also investigated. Transmission fell off steeply at around 380 nm, a characteristic of high-quality ZnO film. From the plot of the wavelength vs. the absorption coefficient, optical band gap energies ranging from 3.26 to 3.31 eV were estimated, which is identical to the value recorded for high-quality ZnO films [20]. The photoluminescence spectra were measured using the 325 nm line of a CW He-Cd laser. Figure 16.7 shows the emission spectra measured at room temperature from films deposited at the substrate temperature from 150 °C to 300 °C. The emission peak position is coincident with the expected energy of the free exciton, and a strong free exciton emission at 380 nm can be clearly observed even at room temperature [21,22]. This confirms that if we use PE-CVD method at a low temperature, a ZnO film emitting UV light at room temperature can be fabricated. This was the first observation of room temperature UV emission from ZnO films deposited by PE-CVD. Under the growth conditions studied by the preliminary experiments mentioned above, the ZnO nanostructure fabrication was carried out by NFO-CVD. The fabrication was performed by introducing the SH light of an Ar^+ laser onto the sapphire substrate surface through the fiber probe. Figure 16.8 shows the VSFM image of fabricated nanometric scale ZnO. The dot was of a 200 nm diameter and a 5 nm height. The diameter is smaller than the wavelength of the irradiating light source. However, because this value includes broadening due to the resolution of VSFM, the real diameter should be much smaller than that observed.

16.3 Nanophotonic Devices and Integration

Future optical transmission systems require an advanced photonic integrated circuit (IC) for increasing speed and capacity. To meet this requirement, its size should become much smaller than that of a conventional diffraction-limited photonic IC. The concept of such a nano-photonic IC is shown in Fig. 16.9, where metallic wires, light emitters, optical switches, input/output terminals, and photo-detectors are all

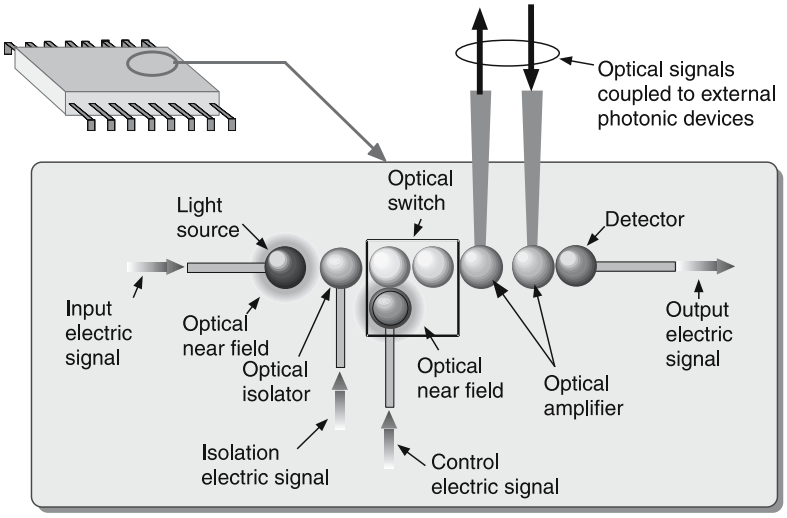


Fig. 16.9. Concept of a planar nano-photonic integrated circuit

controlled by nano-scale single dots and lines [6]. These devices use the optical near-field as a carrier for signal transmission.

As has been demonstrated in the previous sections, the NFO-CVD constitutes a very promising tool for in-situ patterning of nanostructures for this integration because this technique exhibits extraordinarily high controllability and reproducibility in fabricating nanostructures at desired position. This has never been demonstrated by any other conventional self-organized growth technique for semiconductor nanostructures. What is excellent is that as it is based on a photodissociation reaction, selective growth of various materials, i.e., metals, insulators, and semiconductors, can be accomplished by the choice of light source. It allows us to realize a nano-photonic IC composed of nanostructures. This section reviews our recent works on nano-switching devices, which is a key device for nano-photonic integration.

16.3.1 Switching by Nonlinear Absorption in a Single Quantum Dot

In order to demonstrate, e.g., a nonlinear optical switching capability of a single quantum dot (QD), we measured the nonlinear absorption change in a self-assembled single InGaAs QD grown on a (100) GaAs substrate by gas-source molecular beam epitaxy [23]. The average QD diameter was 30 nm and the height was 15 nm. The QD density was about 2×10^{10} dots/cm^2. These QDs were covered with cap layers with a total thickness of 180 nm. As schematically explained in Fig. 16.10a, a probe light (λ = 900–980 nm) was introduced into the back of the sample, and the transmitted light was collected and detected by a high throughput fiber probe with a double tapered structure [24] placed in the vicinity of the sample surface. When the sample was illuminated by a pump light (λ = 635 nm) passing

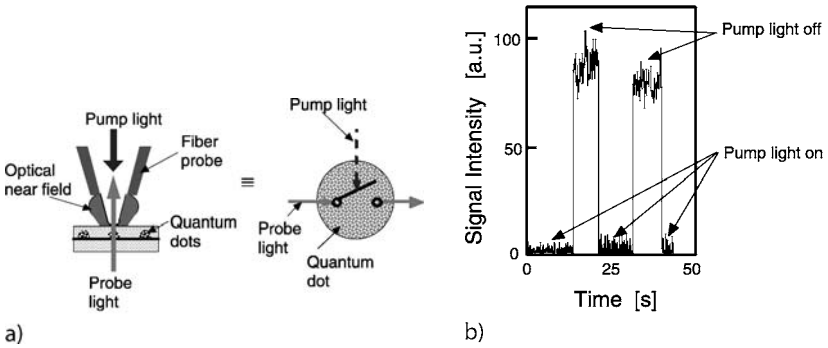


Fig. 16.10. Saturated absorption spectroscopy of a semiconductor quantum dot. (**a**) Schematic explanation of the principle. (**b**) An experimental result of optical switching properties of a single InGaAs quantum dot

through the fiber probe, carriers were generated in the barrier layer, and flew into the QDs. The ground states of the QDs were occupied by the carriers. The resultant reduction of the absorption of the QDs was measured by detecting the transmission change in the probe light. Experiments were carried out in a liquid helium cryostat, where a double modulation/demodulation technique was used for detecting very weak signals. Figure 16.10b shows an experimental result demonstrating a deep modulation of the transmitted probe light power due to irradiation of the pump light. This result confirms that a single QD works like an optical switch, and moreover, the switching operation can be detected by a conventional optical signal detection technique, which is advantageous for application to nano-photonic IC.

16.3.2 Switching by Optical Near-Field Interaction Between Quantum Dots

We have proposed another novel approach towards a nano-photonic switch for both elementary and functional photonic devices [25]. The building blocks of the proposed device consist of three nanometric QDs as illustrated in Fig. 16.11. The discrete energy levels of each dot are described as

$$E_n = E_B + (h^2/8Ma^2)(n_x^2 + n_y^2 + n_z^2), \quad (n_x, n_y, n_z = 1, 2, 3, \cdots), \tag{16.1}$$

where the mass and size are denoted as M and a, respectively. When the sizes of dots 1, 2, 3 are respectively chosen as $a/2$, $a/\sqrt{2}$, and a, two neighboring dots with the distance r_{ij} $(i, j = 1, 2, 3)$ have the same excited energies denoted as E_1, E_2, and E_3 in Fig. 16.11. The energy level E_3 of dot 1, e.g., specified by $(n_x, n_y, n_z) = (1, 1, 1)$, is resonant with states specified by both (2, 1, 1) for dot 2 and (2, 2, 2) for dot 3. Dot 1 is coupled to the input light, which is transmitted by the inter-dot Yukawa interaction [26] to dot 2 that is connected to the output light. Dot 3 is coupled to the control light that governs the switching mechanism. When the control light is on, optical near-field interaction of dot 3 with the other dots are forbidden, since the level E_1

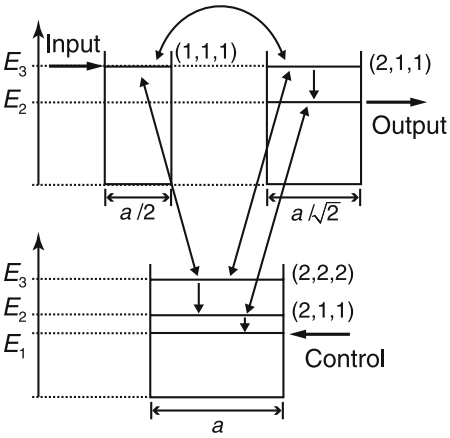


Fig. 16.11. Three-dot configuration as fundamental block of the proposed device. Optical near-field is abbreviated by ONF

of dot 3 is occupied and the intra-dot relaxation time is much faster than the inter-dot transfer time. In contrast, when the control light is off, those interactions are allowed, which results in the population difference in dot 2 producing a difference in the transmission signal. It should be noted that the fast intra-dot relaxation time guarantees a single direction for the signal transmission, and that the frequency conversion from input to output avoids irrelevant cross-talk.

In order to analyze dynamic properties of the device, we solved the following master equations for the population $P_n^j(t)$ of dot j to remain the excited states E_n at time t. We adopted CuCl dots for the following case study though any arbitrary material system is, in principle, suitable. These QDs have discrete energy levels due to quantum size effects, as a result of exciton confinement [27]. The energy levels are described as in (16.1), which means the energy structure reflecting each size of the QDs. Figure 16.12 shows an example of the time evolution of $P_{n=2}^{j=2}(t)$, where the largest dot size is assumed to be 10 nm and the distance between neighboring dots is 10 nm. The intra-dot relaxation time is set as 1 ps. The curve with circles represents the ON condition while the curve with triangles shows the OFF condition. We esti-

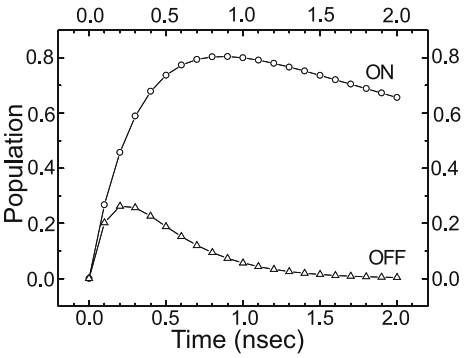


Fig. 16.12. Population of dot 2 as a function of time when control light is ON and OFF

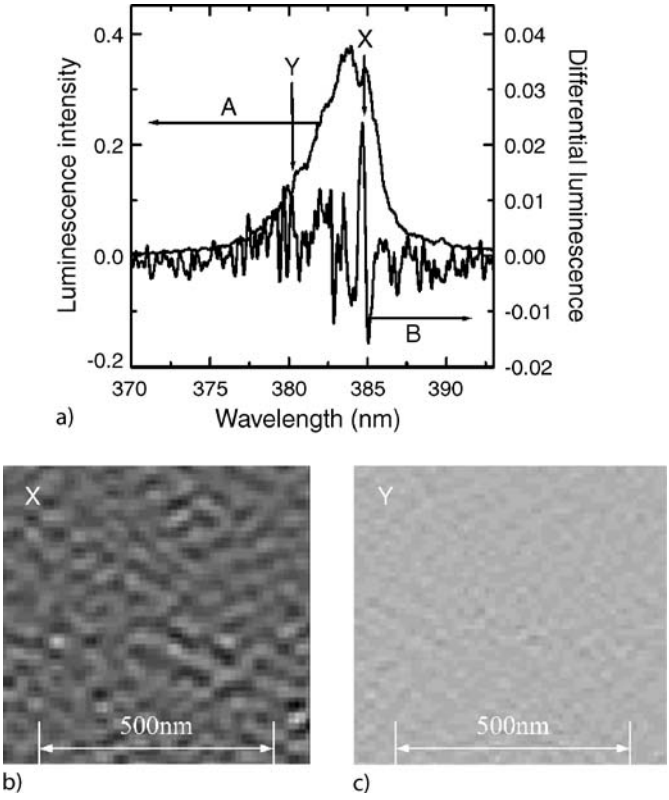


Fig. 16.13. Spectral and spatial information of a sample. (**a**) Curves A and B show the far-field and near-field photoluminescence spectra, respectively, of CuCl quantum dots embedded in NaCl matrix at 15 K. (**b**) and (**c**) Spatial distributions of the photoluminescence peaks X and Y, respectively, marked in a. The spatial distributions for peaks X and Y correspond to the spatial distributions of the quantum dots with sizes of 4.3 nm and 3.4 nm, respectively

mate from this figure that an optical near-field nano-switch can be operated within a few hundred picoseconds, and expect that it would, in principle, be applicable to other material systems like ZnO and GaAs. In order to experimentally verify the switching mechanism and operation proposed, we need to overcome several kind of difficulties. The first step is to identify a specific QD with a desirable size or energy, after observing the spatial and spectral distribution of QDs. Then it is necessary to verify the excitation energy transfer between two QDs, i.e. to detect the desirable frequency conversion from ω_3 to ω_2. Finally we have to show the switching operation of the proposed device. Here we review experimental results on the first step, using CuCl QDs embedded in NaCl matrix. The spatial and spectral characteristics of the sample were investigated by using a near-field optical spectrometer [28]. The curves A and B of Figure 16.13a show the far-field and near-field photolumines-

cence spectra, respectively, of a sample at 15 K that was excited by a He-Cd laser ($\lambda = 325$ nm). The far-field spectrum is inhomogeneously broadened by the size distribution of QDs while the near-field spectrum has the very fine structure that ideally corresponds to a vertical scan at a specific spatial position of the sample.

Spatial distributions of photoluminescence with energies X and Y are shown in Figs. 16.13b and c, respectively. These values of X and Y are respectively equivalent to the dot sizes of 4.8 nm and 3.4 nm, which correspond to a pair of dots 2 and 1, or dots 3 and 2. The spatial resolution of the images is close to the aperture diameter of the fiber probe used. These results indicate that we have established how to identify QDs selected both spatially and spectrally.

16.4 Optical Storage and Readout by Optical Near-Field

The use of an optical near-field for realization of a high-density optical storage system has attracted a great deal of attention. This section propose and demonstrate a new contact slider with a high throughput ratio of near-field intensity for realization high recording density and fast readout to the phase-change medium [29]. Schematics of the slider structure and the data storage system are illustrated in Fig. 16.14. A pyramidal silicon probe array is arranged on the rear pad of the slider. The advantages of such a slider are as follows: (1) The high refractive index of the silicon ($n = 3.67$ at $\lambda = 830$ nm) leads to a short effective wavelength inside the probe, which results in higher throughput and smaller spot size than those of conventional fiber probes made of silica glass [30]. (2) The height of the pyramidal silicon probe array is fabricated to be less than 10 μm so that sufficiently low propagation loss in the silicon is maintained. Furthermore, the probe array has high

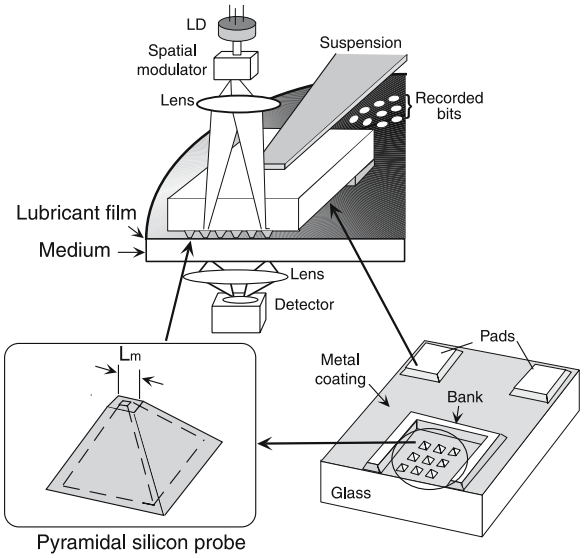


Fig. 16.14. Schematic of the data storage/readout system with a contact slider and a pyramidal silicon probe array (L_m: mesa length)

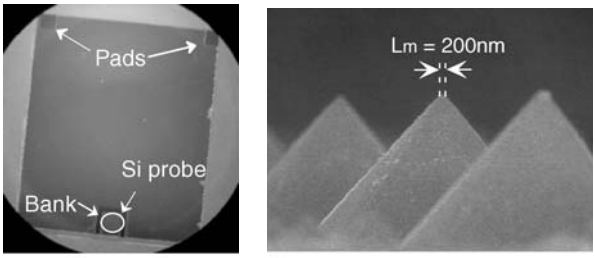


Fig. 16.15. A contact slider. (**a**) An optical image. (**b**) A SEM image of the pyramidal silicon probe array

durability because it is bonded to a thick glass substrate. (3) Compared with those of previously reported pyramidal probes fabricated by use of a focused ion beam [31] or by the transfer mold technique in a pyramidal silicon groove [32], ultrahigh homogeneity in the heights of the probes and pads can be obtained, since the flatness of the probe tip and the upper surface of the pads are determined by the uniformity of the thickness of silicon wafer. (4) Use of a probe array with many elements increases the total data transmission rate by parallel readout [32,33]. In this system the incident light is spatially modulated by an electro-optics method, and the scattered light from a different probe can be read out as a time-sequential signal. Since the key issue in realizing a pyramidal silicon probe array is high homogeneity in the heights of the probes, the probe array is fabricated from a (100)-oriented silicon-on-insulator (SOI) wafer. Figure 16.15a, b show an optical image of the contact slider and a scanning electron microscopic image of the pyramidal silicon probe array fabricated on the slider, respectively. The height dispersions of the probes and pads should be less than 10 nm, because these dispersions are determined by the uniformity of thickness of the SOI wafer. Here the slider is designed by use of the design criteria for a contact-type hard-disk head so that its jumping height over the phase-change medium is maintained at less than 10 nm. Furthermore, since the phase-change medium is fragile, we designed the bank so that the contact stress becomes 100 times weaker than the yield stress of the magnetic disk at a constant linear velocity (CLV) of 0.3 m/s, corresponding to a data transmission rate of 10 MHz for a data density of 1 Tbit/in^2. To increase the readout speed 100 times, i.e., to realize a 1 Gbit/s data transmission rate for data density of 1 Tbit/in^2, we fabricated 100 probe elements on the inner part of the bank for parallel readout.

In recording and readout experiments with the fabricated contact slider, we compared signals transmitted through phase-change marks recorded with a single element of the probe array and focused propagating light. The experimental setup is shown in Fig. 16.16. The contact slider was glued to a suspension. The slider was in contact with a phase-change medium coated with a thin lubricant film (Fomblin Z-DOL). A laser beam (λ = 830 nm) was focused on one element of the probe array on the slider, where the frequency of the rectangularly modulated signal with 50% duty was changed from 0.16 to 2.0 MHz at a CLV of 0.43 m/s. Then the light transmitted through the recording medium was detected with an APD. We used an as-deposited AgInSbTe film as a recording medium. The optical recording powers

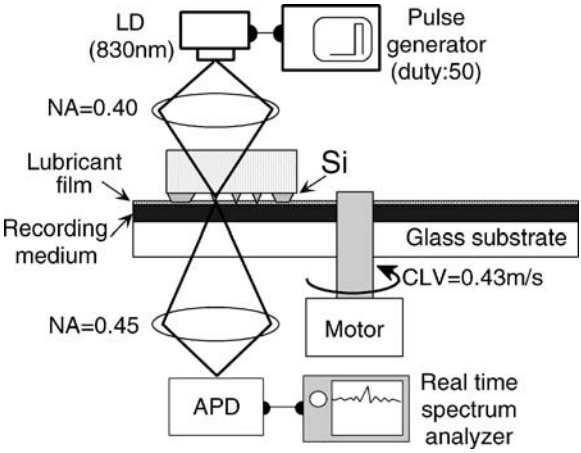


Fig. 16.16. Experimental setup for phase-change storage/readout by the contact slider

for a pyramidal silicon probe with a mesa length L_m of 150 nm (see Figs. 16.1 and 16.17a) and a focused propagating light with an object lens (NA= 0.4) were 200 mW and 15 mW, respectively, which in both cases is the lowest recording power of which we are aware. The optical throughput of the pyramidal silicon probe with a 30 nm thick aluminum coating is 7.5×10^{-2}, which is estimated from the ratio of the optical powers for near- and far-field recordings. Readout was carried out at a CLV of 0.43 m/s, and the constant reading optical powers for the pyramidal silicon probe and the focused propagating light were 20 mW and 3.6 mW, respectively. The resolution bandwidth was fixed at 30 kHz.

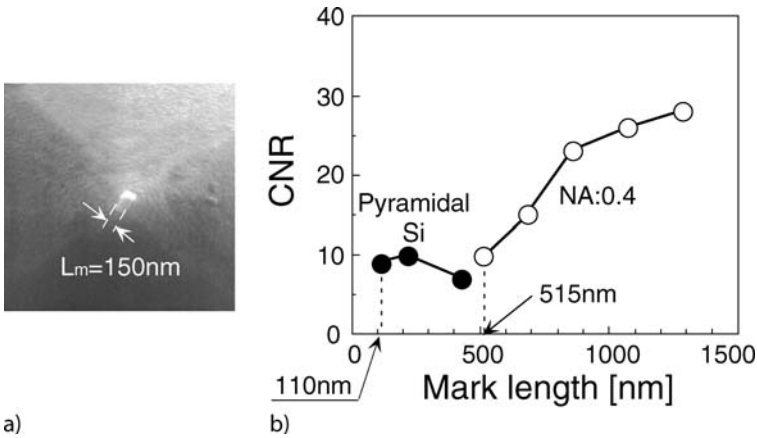


Fig. 16.17. Experimental results of the optical storage and read out. (**a**) Magnified SEM image of the pyramidal silicon probe tip used for the experiment. (**b**) Dependence of the CNR on mark length for the optical near-field (*closed circles*) and propagating light focused by an objective lens (*open circles*)

The dependence of the carrier-to-noise ratio (CNR) on mark length is shown in Fig. 16.17b. In this figure one can see that shorter crystalline phase-changed marks beyond the diffraction limit were recorded and read out by an optical near field generated on the pyramidal silicon probe. The shortest mark length was 110 nm at a CLV of 0.43 m/s, corresponding to a data transmission rate of 2.0 MHz. This is, to our knowledge, the first phase-change recording-reading with a contact slider. Since this slider has 100 elements in the probe array, we expect a 100-fold increase in the data transmission rate by parallel readout. Furthermore, a higher CLV can be expected, since we did not observe any damage on the probe tip or the recording medium after a series of experiments. The constant CNR of the pyramidal silicon probe seen in Fig. 16.17b might be due to the small spot size for recording-reading and the narrow recorded mark width, which are as small as L_m of the pyramidal silicon probe. These results indicate that an increased CNR and decreases mark length will be achieved by means of tracking during readout. Furthermore, it is expected that the recording density can be increased to as high as 1 Tbit/in^2 by optimization of the interference characteristics of the guided modes in the pyramidal silicon probe [34].

16.5 Conclusion

This paper reviewed the recent progress of application of interactions between optical near-fields and nanoscale materials. Photochemical vapor deposition of nanometeric Zn and Al were realized by using an UV optical near-field. Deposition of nanoscale ZnO was also shown. Optical near-field technology offers the opportunity of modifying surfaces and developing new nanostructures that may exhibit a quantum effect due to their extremely small size. Utilizing the very advanced potential of this technology, the concept of nano-photonic IC was proposed. The optical switching operation of a single InGaAs quantum dot was shown to be able to be used for nano-photonic devices. Nano-photonic switching operation utilizing optical near-field interaction was also proposed and related spectroscopy of CuCl quantum dots were demonstrated. High density storage and read-out by optical near-field was also demonstrated. By combining the technique reviewed here with atom manipulation by the optical near-field [5,35], further progress in depositing novel materials and operating more advanced photonic IC can be expected.

Acknowledgements

The author would like to thank Prof. H. Ito, Drs. M. Kourogi, Y. Yamamoto, H. Fukuda (Tokyo Inst. Tech.), Drs. K. Kobayashi, S. Sangu, T. Yatsui, T. Kawazoe, H. Aiyer, V. Polonski, G.H. Lee (ERATO), Drs. T. Saiki and S. Mononobe (KAST) for their collaborations and valuable discussions.

References

1. E.A. Synge: Philos. Mag. **6**, 356 (1928)
2. M. Ohtsu: 'Overview', In: *Near-Field Optics: Principles and Applications*, ed. by X. Zhu, M. Ohtsu (World Scientific, Singapore 2000) pp. 1–8
3. Uma Maheswari, S. Mononobe, K. Yoshida, M. Yoshimoto, M. Ohtsu: Jap. J. Appl. Phys. **38**, 6713 (1999)
4. H. Ito, T. Nakata, K. Sakaki, M. Ohtsu, K.I. Lee, W. Jhe: Phys. Rev. Lett. **76**, 4500 (1996)
5. M. Ohtsu, K. Kobayashi, H. Ito, G.H. Lee: Proc. IEEE **88**, 1499 (2000)
6. M. Ohtsu: P. Soc. Photo–Opt. Inst. **3749**, 478 (1999)
7. E. Maayan, O. Kreinin, G. Bahir, J. Salzman, A. Eyal, R. Beserman: J. Cryst. Growth **135**, 23 (1994)
8. D. Bauerle: Appl. Surf. Sci. **106**, 1 (1996)
9. M. Ohtsu: *Near-Field Nano/Atom Optics and Technology* (Springer-Verlag, Berlin 1998)
10. M. Ohtsu, H. Hori: *Near-Field Nano-Optics* (Kluwer Academic/Plenum Publishers, New York 1999)
11. S. Mononobe, T. Saiki, T. Suzuki, S. Koshihara, M. Ohtsu: Opt. Commun. **146**, 45 (1998)
12. Y. Yamamoto, M. Kourogi, M. Ohtsu, V. Polonski, G.H. Lee: Appl. Phys. Lett. **76**, 2173 (2000)
13. V.V. Polonski, Y. Yamamoto, J.D. White, M. Kourogi, M. Ohtsu: Jap. J. Appl. Phys. **38**, L826 (1999)
14. V.V. Polonski, Y. Yamamoto, M. Kourogi, H. Fukuda, M. Ohtsu: J. Microsc. – Oxford **194**, 545 (1999)
15. D. Ehrlich, R.M. Osgood Jr., T.F. Deutch: J. Vac. Sci. Technol. **21**, 23 (1982)
16. T. Kawazoe, Y. Yamamoto, M. Ohtsu: Appl. Phys. Lett. **79**, 1184 (2001)
17. M. Shimizu, H. Kamei, M. Tanizawa, T. Shiosaki, A. Kawabata: J. Cryst. Growth **89**, 365 (1988)
18. Y. Yamamoto, T. Kawazoe, G.H. Lee, T. Shimizu, M. Kourogi, M. Ohtsu: 'In-situ Lateral Fabrication of Zinc and Aluminum Nanodots by Near Field Optical Chemical Vapor Deposition', In: *Tech. Digest of the Photonic Switching*, (IEEE LEOS, Piscataway 2001) pp. I520–I521
19. G.H. Lee, Y. Yamamoto, M. Kourogi, M. Ohtsu: Thin Solid Films **386**, 117 (2001)
20. T.Y. Ma, G.C. Park, K.W. Kim: Jap. J. Appl. Phys. **35**, 6208 (1996)
21. Z.T. Tang, G.K.L. Wong, P. Yu: Appl. Phys. Lett. **72**, 3270 (1998)
22. D.B. Bagnall, Y.F. Chen, Z. Zhu, T. Yao: Appl. Phys. Lett. **70**, 2230 (1997)
23. T. Matsumoto, M. Ohtsu, K. Matsuda, T. Saiki, H. Saito, K. Nishi: Appl. Phys. Lett. **75**, 3246 (1999)
24. T. Saiki, S. Mononobe, M. Ohtsu, N. Saito, J. Kusano: Appl. Phys. Lett. **68**, 2612 (1996)
25. K. Kobayashi, T. Kawazoe, S. Sangu, J. Lim, M. Ohtsu: 'Theoretical and Experimental Study on a Near-field Optical Nano-switch', In: *Photonics in Switching, OSA Tech. Digest*, (OSA LEOS, Washington, DC 2001) p. 27
26. K. Kobayashi, S. Sangu, H. Ito, M. Ohtsu: Phys. Rev. A **63**, 013806 (2001)
27. N. Sakakura, Y. Masumoto: Phys. Rev. B **56**, 4051 (1997)
28. T. Kawazoe, K. Kobayashi, J. Lim, Y. Narita, M. Ohtsu: 'Verification of Principle for Nano-meter Size Optical Near-field Switch by Using CuCl Quantum Cubes', In: *Tech. Digest of the Photonic Switching*, (IEEE LEOS, Piscataway 2001) pp. I194–I195
29. T. Yastui, M. Kourogi, K. Tsutsui, M. Ohtsu, J. Takahashi: Opt. Lett. **25**, 1279 (2000)
30. H.U. Dangebrink, T. Dziomba, T. Sulzbach, O. Ohlsson, C. Lehrer, L. Frey: J. Microsc. – Oxford **194**, 335 (1999)

31. F. Isshiki, K. Ito, S. Hosaka: Appl. Phys. Lett. **76**, 804 (2000)
32. Y.J. Kim, K. Kurihara, K. Suzuki, M. Nomura, S. Mitsugi, M. Chiba, K. Goto: Jap. J. Appl. Phys. **39**, 1538 (2000)
33. T. Yatsui, M. Kourogi, K. Tsutsui, J. Takahashi, M. Ohtsu: P. Soc. Photo–Opt. Inst. **3791**, 76 (1999)
34. M. Kourogi, T. Yatsui, M. Ohtsu: P. Soc. Photo–Opt. Inst. **3791**, 68 (1999)
35. H. Ito, K. Sakaki, M. Ohtsu, W. Jhe: Appl. Phys. Lett. **70**, 2496 (1997)

17 Optical Trapping of Small Particles

Alexander Rohrbach, Ernst H.K. Stelzer

17.1 Introduction

Kepler postulated in the 17th century that light carries momentum and exerts a force when reflected or refracted. Using this concept, he tried to explain why a comet's tail would always point away from the sun. Later, Sir William Crookes suggested that light pressure was responsible for the rotation of a light mill. The light mill consists of four vanes, which absorb light on one side and reflect it on the other. The vanes form the arms of a vertically supported rotor, which turns with very little friction. The light mill is usually encased in a glass bulb with a high, but not perfect, vacuum. The vanes rotate when light shines on the apparatus. Crookes was wrong, however, in attributing the rotation to the photons' transfer of momentum: in fact it is a heating effect [1]. Kepler was also wrong. The comet's tail is formed by the solar wind, i.e. by ionized particles. Although both Kepler and Crookes were mistaken, they were nevertheless the first to treat the phenomenon of light pressure in physical terms.

In 1873 James Clerk Maxwell provided a quantitative description of optical forces using his electromagnetic theory. Later Lebedev was able to verify these predictions experimentally [2]. However, the calculated forces were so small, that there seemed to be no application for optical pressure. John H. Poynting wrote in 1905: "A very short experience in attempting to measure these light forces is sufficient to make one realize their extreme minuteness—a minuteness which appears to put them beyond consideration in terrestrial affairs." Only in astronomy, in extraterrestrial space where light intensities and distances are huge and the effects of gravity are small, could one imagine any significant influence of optical forces on the movement of small particles.

The invention of the laser changed the scientific world in many ways, also in the field of optical manipulation. The high intensities and high gradients in the electromagnetic field of a laser beam suddenly made it possible to generate forces large enough to accelerate and to trap particles. With lasers particles as small as atoms and as large as whole organisms have been trapped and manipulated in various media.

Arthur Ashkin, considered by many to be the pioneer of optical trapping and optical manipulation, performed his first experiments in the beginning of the 1970s, studying the optical forces on small glass beads. Ashkin's first publication on optical forces described the acceleration of beads in water by a horizontal laser beam [3]. He also constructed a simple trap using two focused, counter-propagating beams:

Török/Kao (Eds.): Optical Imaging and Microscopy, Springer Series in Optical Sciences
Vol. 87 – © Springer-Verlag, Berlin Heidelberg 2007

particles were trapped in between two foci. When the beam was directed upwards, he caused the particles to float in air [4]. This feat, termed optical levitation, was later also performed in vacuum [5] and with various particles. An overview of the historical development of optical traps can be found in reviews by A. Ashkin (see refs. [6] and [7]).

The developments of Ashkin culminated in the invention of the single-beam trap [8]. Whereas in optical levitation gravity is necessary to confine particles, this is not the case for a single-beam trap. In the single-beam trap, strong intensity gradients in a highly focused beam exert a "gradient force" that compensates the "scattering force" (radiation pressure), creating a three-dimensional trap with a single lens. Because of the need for a high numerical aperture lens to strongly focus the laser beam, it was straightforwardly realized in a microscope: a laser beam coupled into a conventional microscope generates an optical trap in the focus of the objective lens. Later called optical tweezers, the single-beam trap is now a highly developed and heavily applied tool in biology, physics and chemistry. Optical tweezers have become the tool of choice for many applications in which a gentle and remotely controllable manipulation is required.

The following sections will cover optical trapping, including photon momentum transfer and the arising trapping forces. Here the physical principles of optical tweezers and how they are applied are outlined in reasonable detail. As an advancement of optical tweezers the development of Photonic Force Microscopy is reviewed and discussed as a high speed tracking system for small volumes. The theory section sketches highly focused fields, particle scattering, interferometric position detection, optical forces and thermal noise. This is followed by a description of a typical experimental setup optimized for 3D trapping and tracking. The section on calibration techniques should give the reader an overview of state-of-the-art procedures. Also in Sect. 17.4, principal benefits and drawbacks of holographic traps are discussed. The last section describes two typical applications of Photonic Force Microscopy in biophysics pointing out how to image interactions instead of structures.

17.2 Optical Trapping

17.2.1 Principles

Photons encode information with their wavelength, their number density, their polarization, their propagation direction, or their phase relative to other photons (coherence). A centuries old list of applications in optics in general and especially in microscopy results from these properties of photons. However, the momentum p connected with the wavelength λ of a photon has been rarely exploited in microscopy, because it was considered as negligibly weak. The momentum, i.e. the force F that the photon exerts for a time period Δt, is

$$p = \hbar \cdot k = h/\lambda = F \cdot \Delta t \tag{17.1}$$

with k-number $k = 2\pi/\lambda$ and with Planck's constant $h = \hbar \cdot 2\pi = 6.626 \cdot 10^{-34}$ Nms. The momentum of a single photon is, for example, $p \approx 6.6 \cdot 10^{-28}$ Ns at a wavelength of $\lambda = 1\,\mu$m in air. With other words, $5 \cdot 10^{16}$ photons, corresponding to an optical power of $P = 10$ mW, would exert a force on a particle of $F = P/c = 30$ picoNewton (pN = 10^{-12} N) due to radiation pressure provided the particle absorbs all photons (c is the speed of light). This force is of the same order as those forces that drive organelles in biological cells. For a dielectric particle, which does not absorb light, but scatters light in different directions, a momentum due to radiation pressure is transferred with a pressure efficiency $Q_{\mathrm{pr}} << 1$. However, this results in an optical force of $F = Q_{\mathrm{pr}} \cdot P \cdot n/c$, which is still in the range of pN. For example, a small glass sphere with a diameter of $\lambda/2$ embedded in water (refractive index $n = 1.33$) has a pressure efficiency $Q_{\mathrm{pr}} = 0.035$, and thus experiences an optical force of $F \approx 1$ pN when all $5 \cdot 10^{16}$ photons hit the sphere [9].

However, to impinge a large number of photons onto a small particle, light must be focused to a small spot where the smallest diameter that can be achieved is $\Delta x \approx \lambda$. Only for coherent laser light, strong intensity gradients arise around the point of maximum intensity. Assuming that a particle consists of a cluster of dipoles, these dipoles oscillate in response to the incident electromagnetic field and emit radiation (the scattered light). In consequence, the dipoles feel a Lorentz force towards the point of the highest intensity if the light frequency ν is below the dipole's resonance frequency ν_0, if $\nu > \nu_0$ the dipole is drawn towards the point of minimum intensity. The force generated as a reaction to the unscattered light is commonly called the gradient force F_{grad}. The force, which reflects the photon's momentum transfer in reaction to the scattered light is called the scattering force F_{scat}. This separation into two forces is justified for particles with diameters not larger than the wavelength, i.e. as long as the field inside the particle has a similar momentum spectrum as the incident field. The gradient force pulls a particle, i.e. the dipoles, into the center of the focus, while the scattering force pushes the particle away from the focus in the direction of light propagation. For this reason, lateral optical forces are much stronger than axial forces due to the strong lateral intensity gradients and the small momentum transfer in the lateral direction. These characteristics are outlined in Fig. 17.7 for a small latex sphere. The trapping position, an equilibrium point for all forces, occurs behind the geometric focus, with the property that any displacement of a particle from this point results in a restoring force.

At this point, we have to distinguish between different particle size ranges. A scatterer with a diameter (2·a) much smaller than the incident wavelength, i.e. $2 \cdot a << \lambda$, is called a Rayleigh scatterer and has a gradient force, which is much stronger than the scattering force. Rayleigh scatterers, such as atoms or small molecules are stably trapped in the center of the focus. Larger particles, which are smaller than or equal to the incident wavelength, in addition experience a scattering force pushing the particle behind the geometric focus until equilibrium of both forces is reached. With increasing particle diameter, the gradient force averages out because different dipoles inside the scatterer experience different forces along different directions. According to EM theory [10] the $\lambda/2$ diameter glass sphere embedded in

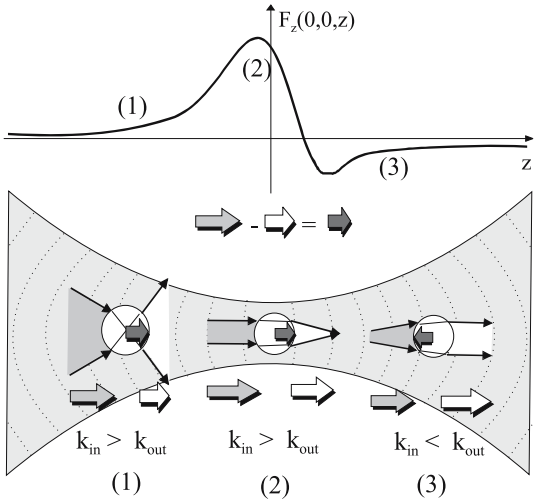


Fig. 17.1. Momentum transfer and optical forces along the optical axis. Three positions (1) (2) (3) of a spherical particle characterize the force profile $F_z(0, 0, z)$. The momentum transferred to the particle is the difference between the incident mean momentum k_{in} and the outgoing mean momentum k_{out} of all photons. This is indicated by gray and white k-vectors, respectively. The difference, indicated by a dark vector within the particle, is positive both in front of and in the geometrical focus, resulting in an optical force that pushes the particle in the direction of light propagation. If $k_{\text{in}} < k_{\text{out}}$, momentum is transferred in backward direction and the particle is pulled back towards the geometrical focus (negative force at position (3)). Stable trapping occurs behind the focus where the optical force is zero ($F_z(0, 0, z) = 0$)

water experiences a maximum backward force of $F_{\text{grad}} + F_{\text{scat}} \approx 2.7\,\text{pN}$ at a transmitted laser power of $P = 30\,\text{mW}$, although a maximum force of $F_{\text{scat}} = 1\,\text{pN}$ pushes the sphere in the opposite direction out of the focus. For particles larger than both the incident wavelength and the dimensions of the focus, the gradient force becomes negligible, while momentum is transferred to the front and the back surfaces of the scatterer. These particles, which are often referred to as Mie-scatterers can be treated with simple ray-optics. They experience approximately the same optical force independently of their diameter. Figure 17.1 illustrates the momentum transferred to a particle in the axial direction as the difference between the mean momentum vector before and after scattering. It is shown, that in front of as well as in the geometric focus, the incident momentum is larger than the outgoing momentum. Only at a certain distance behind the geometric focus is the outgoing mean axial momentum larger than the incident momentum.

17.2.2 Optical Tweezers

Optical tweezers exploit the fact that optical forces in highly focused beams generate a stable three-dimensional optical trap. A particle captured in an optical trap can be moved to different positions in three-dimensional space by moving the trap.

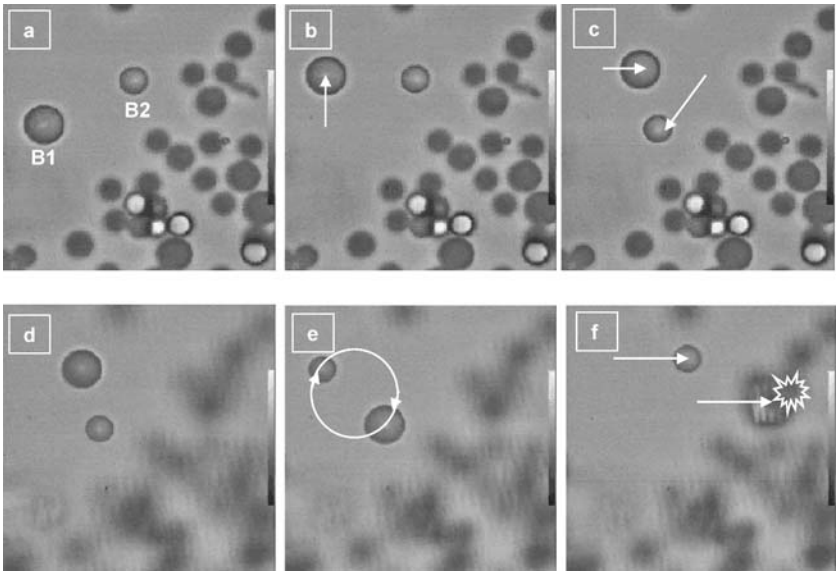


Fig. 17.2. Simultaneous three-dimensional trapping of two beads indicated by B1 and B2. Beads adhering to the coverslip are in a different focal plane and are visible as dark disks. White arrows indicate the directions of the movements of the beads B1 and B2. (**a**) Beads B1 and B2 are trapped by time-multiplexed optical tweezers. (**b**,**c**) Beads B1 and B2 are moved laterally one after the other. (**d**) Beads B1 and B2 are simultaneously moved further away from the coverslip. The beads adhering to the coverslip are out of focus. (**e**) Beads B1 and B2 are simultaneously rotated. (**f**) Beads B1 and B2 are simultaneously moved to the right until bead B1 is pushed out of the trap by a bead attached to the coverslip

The particle can be held at a fixed position by maintaining the position of the trap. Particles undergo small position fluctuations inside the optical trap due to thermal noise. These thermal fluctuations become stronger as the size of the trapped particle or the laser power decrease. However, because of the relatively small position changes due to the particle's Brownian motion, the optical trap can be switched on and off for very short time periods or can be moved to another position to manipulate another particle simultaneously. By time-multiplexing in the 100 Hz range, the un-trapped particle is not able to leave the trapping area. The idea of moving particles in three dimensions with optical tweezers is visualized in Fig. 17.2 by single frames extracted from a movie. Here, two beads are simultaneously trapped and moved laterally and upwards.

The main advantages of optical tweezers are that they are able to exert small, adjustable forces in the pN range and that manipulation of particles is realized without mechanical contact. For comparison, atomic force microscopes [11] use a fine tip that touches the sample, maintains the contact to its probing tip and exerts or measures much higher forces in the μN-range. The first demonstration of optical tweezers in biology was realized by Ashkin et al. in 1987 by trapping viruses and E. coli

bacteria [12,13]. A large number of experiments with this manipulative technique followed, providing new insights especially in cell biology. For example, Block et al. [14] measured the torsional compliance of bacterial flagella. The technique of tweezing was extended to laser cutting ("Laser scissors") of cells and organelles: Greulich and co-workers cut and manipulated pieces of chromosomes for gene isolation [15]; and Berns and co-workers brought cells into contact in order to effect cell fusion by cutting the common wall [16]. In fertility studies, three-dimensional manipulation of live sperm cells [17] and even insertion of selected sperms into eggs [18] were performed with optical tweezers.

Holding a particle in a trap effectively prevents it from moving too far. A particle that tries to leave the optical trap exerts a force which is approximately linear with the displacement and which can be measured. To do this, the trap has to be calibrated, i.e. the optical force profile or the slopes of the force profiles (the force constants κ_x, κ_y, or κ_z , see Fig. 17.7) has to be determined. Calibrated traps allowsed studying the forces exerted by molecular motors that move along microtubules or actin filaments of cells. In such experiments a small bead was used as a handle that is linked to a motor protein. The motor, activated by ATP or GTP, tried to pull the bead out of the trap by moving forward [19–21]. In addition, it was possible to observe the exact position of a moving motor protein: Svoboda et al. measured the motion of a kinesin molecule along a microtubule using an interferometric technique with subnanometer resolution [22]. By using feedback techniques, the precision of force and position measurements could be further increased, which was applied for studying the interaction of myosin-actin systems [23,24]. Stretching macro-molecules such as DNA by exerting optical forces enabled studies of the micro-mechanical properties, but also of the dynamical behavior of DNA polymers [25–27].

Numerous applications grew up during the last years using optical traps in chemistry, polymer and colloidal physics, to measure surface and colloidal forces or the dynamics and the visco-elastic properties of polymer chains. Only a small part is mentioned here (see [28] for a recent review).

17.2.3 Photonic Force Microscopy

For a long time, the thermal motion Δx of a trapped particle seemed to limit the resolution of force measurements due to $\Delta F = \kappa \cdot \Delta x$ and the precision of position measurements. This became especially apparent when an optically trapped particle was used as a probe to scan and image soft surfaces. The first scanning probe microscope was reported by Hertz and co-workers, who captured a 290 nm diameter silica particle, which was scanned around the object at a distance of more than 10λ [29]. The light scattered by the trapped particle was altered by the object, which consisted of thin, freestanding wires. By the early 90s, it became apparent that the key to a high-resolution scanning instrument was the position detection of the trapped particle, when it is also used as a probe.

Several detectors have since been developed, which record and analyze the position fluctuations of a trapped bead. Denk and Webb measured the lateral thermal

motion of microscopic objects with a two-dimensional detector [30]. Their instrument, based on a modified differential interference contrast microscope, was used to investigate the motion of hair bundles. Ghislain and Webb used a trapped stylus (a small glass shard) to scan a soft surface [31]. The probe's axial position changed according to the scanned surface profile and was measured by a photo detector. The detector signal, determined by the intensity of the forward scattered light, was found to be linearly dependent on the particle's displacement within a certain range. At the same time Kawata et al. presented a laser trapped probe scanning microscope [32], which exploited the illumination of the sample with evanescent waves. However, determining the axial position of the probe turned out to be difficult, even by using a spot illumination [33]. Friese et al. [34,35] evaluated the intensity of light reflected from a trapped sphere, which was used to probe three-dimensional surfaces. Florin et al. [36] scanned a trapped fluorescent latex bead across a soft surface. Here, two-photon fluorescence of a 1064 nm Nd:YAG laser was exploited to determine the axial position of the probe, which was displaced by contacting the surface. The scanning could be performed either in a 'rolling mode' or in a 'tapping mode'. This device, which is considered as the first version of the Photonic Force Microscope (PFM) [37], did not yet exploit thermal motion. All the techniques described so far required high laser powers to achieve a good spatial resolution by minimizing thermal noise fluctuations, and, therefore, presented a high risk of photo damage to living specimen.

A major step towards modest illumination strength of the biological sample and towards a precise and flexible three-dimensional position detection of the probe was to take advantage of the coherent nature of the trapping light. The interference of unscattered and forward scattered light produces an interference pattern, which can be recorded with a spatially resolving detector (e.g. a quadrant photodiode) in the back-focal plane of the detection lens [38]. This technique provides not only a lateral position signal similar to that obtained by Denk and Webb [30] and Ghislain and Webb [31], but also encodes the axial position of the particle [39–41] due to the phase anomaly of focused beams [42]. With this detection technique, position displacements of a particle in x, y and z could be tracked at a rate of more than 0.5 MHz with nm spatial precision. Hence, it became possible to perform three-dimensional topography scans around three-dimensional objects by exploiting the Brownian motion of the probe inside the trap [43].

The three basic principles driving a PFM [44], optical forces, thermal fluctuations, and position detection of the probe, are outlined in Fig. 17.3. An objective lens generates a focus with an Airy-disc diameter $\Delta x \approx \lambda$, which is necessary for the generation of strong intensity gradients in all three dimensions. A particle is drawn into the focal region and is trapped stably provided the applied laser power is in the milliwatt range. Due to stochastic collisions with the surrounding fluid, the particle undergoes a random walk (Brownian motion) and thus scans a specific volume within a certain time period. The volume in which these thermal fluctuations are detected is the volume of the optical trap (see random track and trap border outlined in Fig. 17.3). When the laser power and thus the optical forces decrease, the trapping

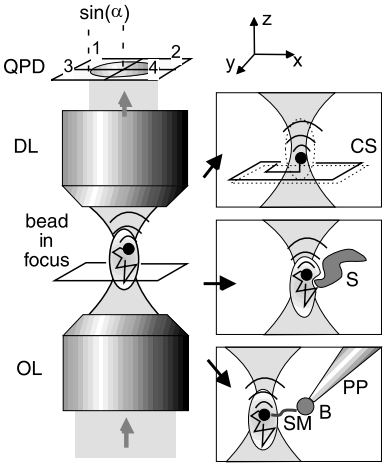


Fig. 17.3. Principles of a Photonic Force Microscope (PFM). A laser beam is tightly focused by a high NA objective lens (OL). The incident light is scattered by the local probe in the focal region. The resulting optical forces generate a three-dimensional optical trap. The local probe explores the trapping volume due to Brownian motion. Light scattered by the probe and unscattered light generate an interference pattern, which is projected by a detection lens (DL) onto a quadrant photodiode (QPD) that is placed in the back focal plane (BFP) of the detection lens. The quadrants are numbered 1 to 4. The detector signals generated by the fluctuating probe are analyzed to determine the interaction with the local environment. Typical experiments are outlined by three boxes on the right: (**a**) The probe is fixed to a coverslip (CS), which is moved with the stage along a specific path to characterize the detection system. (**b**) A structure (S) penetrates and alters the trapping volume of the fluctuating probe disclosing the interaction between structure and probe. (**c**) A single molecule (SM) is tethered between the local probe in the trap and a bead (B) attached to a micro pipette (PP) to measure the visco-elastic properties of the molecule

volume increases. If the optical forces become too weak, the collision forces can kick the particle out of the trap. The position fluctuation, i.e. the rate with which the particle changes its direction, is influenced by the viscous drag γ and the force profile of the optical trap. The optical trap can be calibrated by determining the force profile at known viscous drag γ [45].

The local environment (as outlined in Fig. 17.3) influences the position fluctuations of the trapped probe in a specific manner. By analyzing the altered fluctuations, it is possible to determine the coupling of the probe with external forces. These forces might be due to mechanical linkage or contact with cells, macromolecules or motor proteins [46–50], due to viscosity changes or local flow gradients [51–53], or due to the presence of electrostatic [54,55] or entropic potentials [56]. By moving the trap volume along a grid and repeating the procedure, it is possible to establish complete multi-dimensional maps of the parameters described above [37,43].

17.2.4 3D Tracking with Coherent Light

How is it possible to measure the position changes of the probe inside the optical focus – orders of magnitude smaller than the extensions of the focus? A condensor, which can be another objective lens, collects the unscattered light and the light scattered at the particle in the forward direction (see Fig. 17.3). A quadrant photodiode or a fast camera placed in the back focal plane of the detection lens, records the interference pattern as a function of time and particle position [39,40]. Analyzing the position signals determines the particle's fluctuations and positions in three dimensions at a rate of 500 kHz with an accuracy of better than 5 nm [57]. Alternative methods, such as measuring the power of the back scattered light [34,58], the forward scattered light [59] or exploiting evanescent waves [54,60], provide the same information, but with a lower spatial precision and less flexibility. Another method exploits the strong axial intensity gradient of a two-photon focus that generates a signal that is also approximately linear with the intensity [36]. This method is rather insensitive to phase distortions induced by the specimen, but requires higher laser powers.

The interference intensity of the scattered and the unscattered field, E_i and E_s, at the detector is usually not a ring pattern, but the overlap of smoothly decaying spots: a constant bright spot from the unscattered wave $|E_i|^2$, and a position dependent dim spot produced by the scattered wave $|E_s|^2 + 2\mathrm{Re}\{E_i E_s^*\}$. The interference intensity changes depending on the scatterer. If a position detector consists of at least 4 diodes, the sum of two adjacent detector element intensities is different than the sum of the other two pixel intensities. Applying this technique, lateral displacements of the scatterer from the optical axis can be measured with nm-precision. Surprisingly, axial displacements of the scatterer can be measured when the sum of all 4 pixel intensities is recorded [40]. This only works for coherent scattering processes, where the Gouy phase shift (also phase anomaly) is exploited, which is inherent in every non-plane wave. The Gouy phase shift is a phase shift of π of the electromagnetic field along the region of the maximum intensity of a focused beam and results in a longer wavelength and a shorter momentum vector in the focus. This effect is illustrated in Fig. 17.4, where axial line scans of the electric fields of an incident focused beam and a scattered wave is shown for three different axial positions of a scatterer. The abscissas of the plots are scaled in λ, but the distances between field maxima are larger than λ. Because of this effect, the interference intensity at the detector plane, i.e. the phase difference of the scattered and the unscattered wave depend on the position of the scatterer. The detected intensity is linear over a range of several hundred nanometers with the z-position of the scatterer [57].

17.2.5 Atom Traps

Since the presentation of the first optical trap [61], the field of atom trapping increased in significance in an impressive manner and resulted in several Nobel prizes. Atoms absorb a net momentum h/ν from photons with a rate depending on the frequency ν of the incident field. The deceleration of atoms flying along a direction

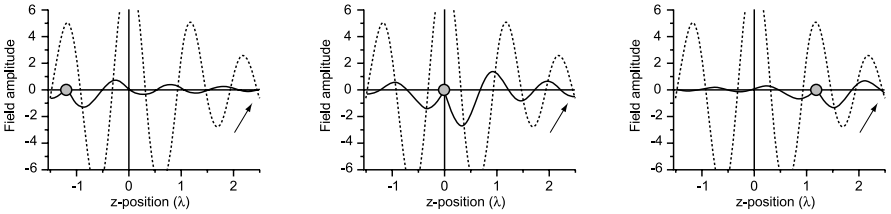


Fig. 17.4. Axial line scans of the electric field around the geometric focus ($z = 0$) illustrate the effect of the Gouy-phase shift on different particle positions. The incident field is indicated by the dotted curve and identical in all three figures. The scattered wave, indicated by the *solid curve*, is emitted at three positions of the particle (*gray spot*) where the amplitude of the incident field is maximal. The interference between the scattered and the unscattered field (see *arrow marker on the right*) depends on the position of the particle

opposite to the momentum transport of the photons was used for optical cooling. With a red detuned laser, i.e. a laser with a frequency $\nu - \Delta\nu$ below the atomic resonance frequency, atoms take up momentum from photons coming from opposite direction, who have an absorption frequency $\nu + \Delta\nu$ due to the Doppler-shift. Using six lasers along three orthogonal and anti-parallel directions it becomes possible to generate 'optical molasses', consisting of atoms cooled down to a few micro Kelvins [62]. By using focused laser light, the gradient forces acting on dipoles are used to realize three-dimensional atom traps [7]. The polarizability α of the atoms can be adjusted by the laser frequency and results in an optical force acting on the atoms towards the point of highest ($\alpha > 0$) or lowest ($\alpha < 0$) intensity of a static light distribution. This simple trap is the foundation for the development of Bose-Einstein-Condensates (BEC) [63], i.e. an atom vapor that is coherent in a single ground state and reveals the fascinating properties inherent to bosons.

17.3 Theory

In order to understand the interaction of photons with a particle and the interaction of the trapped probe with its local environment, one has to appreciate the following points theoretically: the fields in a highly focused beam, the scattering of various particles therein, the interference of the scattered and the unscattered fields for the particle's position detection, the arising trapping forces and the thermal position fluctuations. We want to mention that neither ray optics calculations for trapping forces of particles smaller than or equal to the wavelength provide correct results, nor do Gaussian optics describe highly focused beam correctly [10].

17.3.1 Arbitrary Focused Fields

A flexible and elegant method to calculate arbitrary focused fields is the following approach: we start with a field distribution $\tilde{\mathbf{E}}_0\left(k_x, k_y\right)$ in the back focal plane of an

objective lens. Since we use lasers, this will be an expanded Gaussian beam in the TEM_{00}-mode. This distribution will be altered by passing the aperture stop (usually of a circular shape) and in the focus will deviate from a Gaussian shape. We define an arbitrary focused field as an electromagnetic field $\tilde{\mathbf{E}}_i(k_x, k_y)$ with an arbitrary magnitude and phase across the aperture of the back focal plane of the lens. The field distribution in the focal region of an ideal lens can then be obtained by taking the three-dimensional inverse Fourier transform of the field $\tilde{\mathbf{E}}_i(k_x, k_y)$ [64,65].

$\mathbf{E}(x, y, z)$ can be written as a decomposition of plane waves with directions (k_x, k_y):

$$\begin{aligned}\mathbf{E}_i(x, y, z) &= (2\pi)^{-3} \iiint\limits_{k_\perp \le k_0 \cdot \mathrm{NA}} \tilde{\mathbf{E}}_i(k_x, k_y, k_z) \cdot \exp[-i\mathbf{kr}]\mathrm{d}k_x\mathrm{d}k_y\mathrm{d}k_z \\ &= (2\pi)^{-2} \iint\limits_{k_\perp \le k_0 \cdot \mathrm{NA}} \tilde{\mathbf{E}}_i(k_x, k_y) \cdot \exp\left[-i\left(k_x x + k_y y + \sqrt{k_n^2 - k_x^2 - k_y^2} \cdot z\right)\right] \mathrm{d}k_x\mathrm{d}k_y\end{aligned} \tag{17.2}$$

is called the pupil function and lies on a spherical surface, since in coherent optics the z-component of the k-vector is $k_z = \pm(k_n^2 - k_\perp^2)^{1/2} = k_n \cdot \cos\theta$, $k_\perp = (k_x^2 + k_y^2)^{1/2}$, $k_0 = |\mathbf{k}_0| = 2\pi/\lambda_0 = k_n/n_m$. Therefore, $\tilde{\mathbf{E}}_i(k_x, k_y, k_z)$ is fully described by its two-dimensional projection, the angular momentum representation $\tilde{\mathbf{E}}_i(k_x, k_y)$. $\mathrm{NA} = n_m \cdot \sin\alpha$ is the numerical aperture of the objective lens. $\tilde{\mathbf{E}}_i(k_x, k_y)$ can be further split into a product that describes the field strength $E_0(\mathbf{k})$, the transmission and phase $A(\mathbf{k})$, the polarization $\mathbf{P}(\mathbf{k})$ and the apodization $B(\mathbf{k})$ of the incident electric field at the back focal plane (BFP) such that $\tilde{\mathbf{E}}_i(k_x, k_y) = E_0(k_x, k_y) \cdot A(k_x, k_y) \cdot \mathbf{P}(k_x, k_y) \cdot B(k_x, k_y)$ [10]. The apodization function $B(\mathbf{k}) = \sqrt{(k_n/k_z)} = 1/\sqrt{\cos(\theta)}$ obeys the sine-condition. The polarization function $\mathbf{P}(\mathbf{k}) = (P_x(\mathbf{k}), P_y(\mathbf{k}), P_z(\mathbf{k}))$ describes the components of the electric field vector of a polarized beam that changes its direction from $\mathbf{k} = (0, 0, 1)$ to $\mathbf{k} = (k_x, k_y, k_z)$. A calculation result of the intensity of (17.2) in the yz-plane is shown in Fig. 17.5.

There are other calculation methods that deliver the same results by solving integrals corresponding to that of (17.2) in space domain [66], but become more compli-

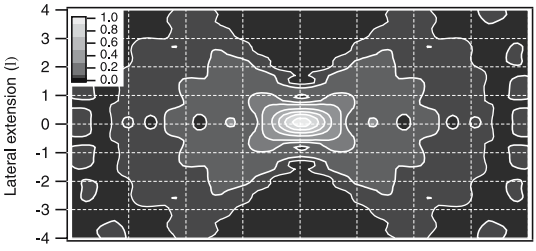


Fig. 17.5. Intensity distribution in the xz-plane ($y = 0$). The scaling is in wavelengths ($\lambda = \lambda_0/n_m$) of the laser in the medium. The intensity in the focus results from a x-polarized Gaussian beam with a two-fold over-illumination of the back focal plane of an ideal lens with $\mathrm{NA} = n_m \cdot 0.9 = 1.2$ in water. Contour levels are non-linearly scaled

cated when modified to account for aberrations in amplitude and phase. A method frequently used to calculate scattering in focused beams or trapping forces starts with a small-angle approximation equation and then corrects with higher orders of a Gaussian beam [67]. This method delivers correct widths of a highly focused beam in the lateral direction, but produces differences of up to 60% for the slope of the axial focus profile in comparison to a realistic, aperture limited focus calculated with our method. The correct description of the electromagnetic fields in the focus is required in order to calculate the correct trapping forces.

17.3.2 Scattering by Focused Fields

It is about a century ago that Rayleigh investigated the scattering of sunlight by point-like particles. His theory was extended by Gans and by Debye for particles of arbitrary shape with sizes of up to the wavelength, but with relatively small refractive indices [68]. At about the same time, Mie derived exact solutions for spherical particles of arbitrary size and refractive index. Their calculations took account of the scattering of a single plane wave by a particle.

However, using the superposition principle it is possible to extend their theories to arbitrary incident fields, which can be decomposed into a spectrum of plane waves with directions (k_x, k_y). The concept is the following: first, the spectrum $\tilde{\mathbf{E}}_s\left(k_x, k_y\right)$, resulting from the scattering of a single plane wave with amplitude E_0 and direction $\mathbf{k}_i = k_n \mathbf{e}_z$ at a sphere is calculated with Mie-theory:

$$\tilde{\mathbf{E}}_s\left(k_x, k_y\right) = E_0 \cdot T_2\left(k_x, k_y\right) \cdot \frac{k_x\, k_z}{k_n\, k_\perp^2} \cdot \begin{pmatrix} k_x \\ k_y \\ -k_\perp^2/k \end{pmatrix} + E_0 \cdot T_1\left(k_x, k_y\right) \cdot \frac{k_y}{k_\perp^2} \cdot \begin{pmatrix} k_y \\ -k_x \\ 0 \end{pmatrix} \quad (17.3)$$

Here T_1 and T_2 are the angle-dependent Mie-scattering functions parallel and perpendicular to the plane of incidence [69]. In a next step, all other scatter spectra $\mathbf{M} \cdot \tilde{\mathbf{E}}_s\left(k_x, k_y\right)$ from the scattering of plane waves with amplitudes $\tilde{\mathbf{E}}_i\left(k_{ix}, k_{iy}\right)$ and directions (k_{ix}, k_{iy}, k_{iz}) are obtained by multiplying Euler rotation matrices $\mathbf{M}\left(k_{ix}, k_{iy}\right)$ [41]. All $\mathbf{M}\left(k_{ix}, k_{iy}\right) \cdot \tilde{\mathbf{E}}_s\left(k_x, k_y\right)$ with $k_{i\perp} < k_0 \cdot \mathrm{NA}$ are superposed. The scattered field $\mathbf{E}_s(\mathbf{r},\mathbf{b})$ at point $\mathbf{r}$ or its angular spectrum $\tilde{\mathbf{E}}_s\left(k_x, k_y, \mathbf{b}\right)$ change with the particle position $\mathbf{b}$ relative to the center of the focus. It simplifies the calculations to leave the particle at its position $(0, 0, 0)$ and to shift the incident beam by $-\mathbf{b}$, (i.e. to modulate the spectrum of the incident field).

17.3.3 Interferometric Position Detection

The interference intensity in the back focal plane (BFP) of the detection lens, i.e. the angular intensity distribution of scattered and unscattered light is determined up to an angle α_D in the BFP of the lens with $\mathrm{NA}_\mathrm{D} = n_m \cdot \sin \alpha_\mathrm{D}$. The interference pattern is projected directly onto a quadrant photodiode (QPD) where the quadrants are numbered 1 to 4. Either an aperture stop or a magnifying lens system regulates the

effective NA_D and hence the contribution of high-angle rays to the position signal. Both detection linearity and sensitivity depend on NA_D [70,71]. The interference pattern is described as follows:

$$I_D\left(k_x, k_y, \mathbf{b}\right) = \frac{\varepsilon_0 c}{2} \cdot \left|\tilde{\mathbf{E}}_i\left(k_x, k_y\right) + \tilde{\mathbf{E}}_s\left(k_x, k_y, \mathbf{b}\right)\right|^2$$
$$= I_i\left(k_x, k_y\right) + I_s\left(k_x, k_y, \mathbf{b}\right) + 2\sqrt{I_i I_s(\mathbf{b})}$$
$$\cdot \cos\left(\phi_i\left(k_x, k_y\right) - \phi_s\left(k_x, k_y, \mathbf{b}\right)\right) \quad (17.4)$$

This intensity is integrated over the area of each quadrant with index $m \in [1,4]$, where the integration limits A, C are either 0 or $k_0 \cdot NA_D$ and B, D are either $-k_0 \cdot NA_D$ or 0, respectively:

$$S_m(\mathbf{b}) = \int_B^A \int_D^C I_D(k_x, k_y, \mathbf{b}) \cdot \mathrm{step}(k_0 \cdot \mathrm{NA_D} - \sqrt{k_x^2 + k_y^2})\mathrm{d}k_x \mathrm{d}k_y \quad (17.5)$$

The distribution $I_D(k_x, k_y, \mathbf{b})$ is delimited in its spatial extent by the diameter of the pupil plane $2 \cdot k_0 \cdot NA_D$, which is expressed by the Heavyside step function. Typically $I_D(k_x, k_y, \mathbf{b})$ is imaged by over-illuminating the QPD 2-3 fold [57].

The three components of the signal $S(\mathbf{b}) = (S_x, S_y, S_z)$ for a particle located at $\mathbf{b}$ are then obtained as follows:

$$\mathbf{S}(\mathbf{b}) = (S_x(\mathbf{b}), S_y(\mathbf{b}), S_z(\mathbf{b})) = ((S_1 - S_2 + S_3 - S_4), (S_1 + S_2 - S_3 - S_4),$$
$$(S_1 + S_2 + S_3 + S_4))/S_0 \quad (17.6)$$

S_0 is a constant value and is obtained by the total intensity incident on the diode when no scatterer is present. The dependency on $\mathbf{b}$ has been omitted for the signals $S_{1,2,3,4}$ to improve readability of (17.6).

For a small particle displacement $\mathbf{b}$ from the center of the focus a linear relationship between $\mathbf{S}(\mathbf{b})$ and $\mathbf{b}$ can be expected, i.e. $\mathbf{S}(\mathbf{b}) = \hat{\mathbf{g}} \cdot (\mathbf{b} - \mathbf{b}_0) \approx (g_x \cdot b_x, g_y \cdot b_y, g_z \cdot (b_z - b_{0z}))$. $\mathbf{b}_0$ indicates the center of the optical trap. The detection sensitivity $\hat{\mathbf{g}}$ is a matrix, which is diagonal over a sufficiently large range of $\mathbf{b}$. This is illustrated by Fig. 17.6 for lateral and axial displacements of a 300 nm latex sphere for a numerical aperture $n \cdot \sin\alpha_D = n \cdot 0.5$ of the detection lens. Higher angular apertures ($\sin\alpha_D > 0.7$) decrease the linear detection range in axial direction. This is due to the fact that a lower NA_D of the condenser lens results in a larger extent of the detection point-spread function, which improves its overlap with the trapping volume. For these reasons, we recommend to adjust the aperture stop of the detection lens to values of $\sin(\alpha) = 0.3$ to 0.5.

Secondly, one can deduce from Fig. 17.6 how strong the coupling between the lateral and the axial position signals is, when particles are displaced off-axis? In Fig. 17.6 the signal distributions $S_x(b_x, 0, b_z)$ and $S_z(b_x, 0, b_z)$ are shown for a (4×1.6) µm range of particle positions $\mathbf{b} = (b_x, 0, b_z)$. The detector signals are displayed by contour lines, while the probability density is shown in a gray level

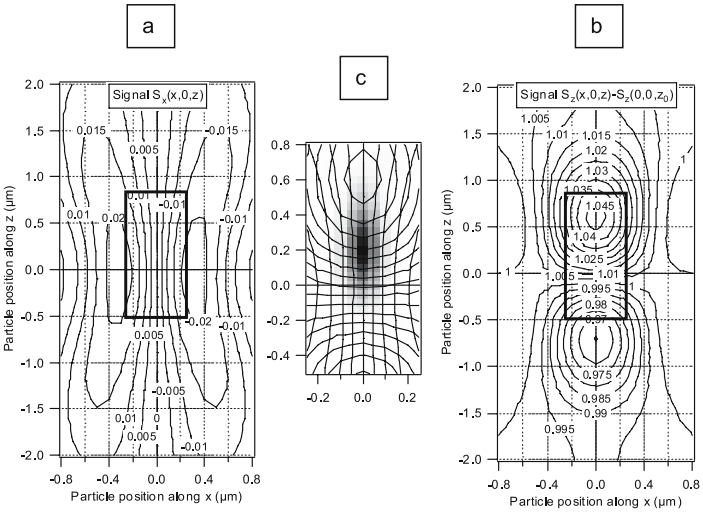


Fig. 17.6. Lateral and axial detector signals for particle positions in the *xz*-plane. The particle is a 300 nm latex sphere. The central regions of (**a**) and (**b**) are magnified and overlaid in an additional contour plot (**c**). The probability density $p_B(x, 0, z)$ of finding the trapped particle at a certain position in the xz-plane is shown by the gray level image in the center plot. A darker gray level indicates a higher probability to find the particle within this volume element. The position signals S_x and S_z are displayed by labeled contour lines. Each line represents a constant signal value at the detector. Vertical contour lines in the S_x-plot indicate a signal that is independent of the axial particle position; horizontal contour lines in the S_z-plot represent a signal that is independent of the lateral particle position (**a**, **b**, **c**)

image (center) for different particle displacements. Here, darker gray levels indicate a higher probability to find the particle at this position. A particle moving along a contour line provides a constant detector signal. Vertical contour lines in the S_x-plot indicate a signal that is independent of the axial particle position; horizontal contour lines in the S_z-plot represent a signal that is independent of the lateral particle position. Magnifications of the central regions of $S_x(b_x, 0, b_z)$ and $S_z(b_x, 0, b_z)$ reveal that nearly all lines are curved. The curvature increases with the displacement from the point $(0, 0, b_{z0})$. The signals become non-linear for small off-axis displacements but are still unique This is not the case for larger displacements. Since nearly all curves are closed, a straight line scan in one direction intersects a contour line twice and the two particle positions deliver the same diode signal.

Interestingly the interferometric tracking technique can also be applied to elongated probes, which undergo fluctuations in position and orientation [72].

Besides reducing the numerical aperture of the detection lens, the linear detection range can also be increased by lowering the numerical aperture of the trapping lens. In this case, the trapping position (the trap center) is shifted due to a stronger relative radiation pressure and the interference pattern is altered advantageously [50].

17.3.4 Trapping forces

Trapping forces are derived from the Lorentz force density $\mathbf{f} = (\mathbf{p}\nabla\mathbf{E}) + \partial\mathbf{p}/\partial t \times \mathbf{B}$ for dipoles, where a linear response on the total electric field $\mathbf{p}(\mathbf{r}, t) = \alpha_0 \varepsilon \mathbf{E}(\mathbf{r}, t)$ is assumed [73]. Higher polarization moments are neglected, which is valid for reasonable laser powers and the specified particle sizes. $\mathbf{E}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ are the total electric and magnetic field, respectively. The polarizability on a volume element V is $\alpha_0 = 3V(m^2 - 1)/(m^2 + 2)$, where $m = n_\mathrm{s}/n_\mathrm{m}$ is the ratio of the refractive indices of the scatterer and the surrounding medium and $\varepsilon = n_\mathrm{m}^2 \cdot \varepsilon_0$ is the electric permittivity [74]. After integrating the time-averaged force density $\langle \mathbf{f}(\mathbf{r}, \mathbf{b}) \rangle = \mathrm{Re}\left\{\alpha_0 \varepsilon \nabla |\mathbf{E}(\mathbf{r}, \mathbf{b})|^2\right\}/4V$ over the volume of the scatterer at position $\mathbf{b}$, one obtains the expression of the optical force without the detour of the Maxwell stress tensor [75]:

$$\begin{aligned}\mathbf{F}(\mathbf{b}) &= \frac{1}{4V}\mathrm{Re}\int_V \alpha_0 \varepsilon \nabla |\mathbf{E}_\mathrm{i} + \mathbf{E}_\mathrm{s}(\mathbf{b})|^2 \mathrm{d}V \\ &= \frac{1}{4V}\mathrm{Re}\int_V \alpha_0 \varepsilon \nabla |\mathbf{E}_i|^2 \mathrm{d}V \\ &\quad + \frac{1}{4V}\mathrm{Re}\oint_{\partial V} \alpha_0 \varepsilon \mathbf{n}\left(\mathbf{E}_\mathrm{i}^* \mathbf{E}_\mathrm{s}(\mathbf{b}) + \mathbf{E}_\mathrm{i}\mathbf{E}_\mathrm{s}^*(\mathbf{b}) + |\mathbf{E}_\mathrm{s}(\mathbf{b})|^2\right) \mathrm{d}A \end{aligned} \tag{17.7}$$

The force is split into two components, the gradient force and the scattering force, resulting from the incident and the scattered field, $\mathbf{E}_\mathrm{i}$ and $\mathbf{E}_\mathrm{s}$. The scattering force (described by the surface integral) results from the extinction and redistribution of momentum. For dielectric particles in the Rayleigh-Gans regime (also known as Born approximation), the following expression for the two force components is valid [74]:

$$\begin{aligned}\mathbf{F}(\mathbf{b}) &= \mathbf{F}_\mathrm{grad}(\mathbf{b}) + \mathbf{F}_\mathrm{sca}(\mathbf{b}) \\ &\approx \frac{\alpha_0 n_m}{2cV}\int_{V(b)} \nabla I_0(\mathbf{r}) d^3 r + \frac{n_m}{kc} I_0(\mathbf{b}) \left(C_\mathrm{ext}(\mathbf{b}) \langle \mathbf{k}_\mathrm{i}(\mathbf{b}) \rangle - C_\mathrm{sca}(\mathbf{b}) \langle \mathbf{k}_\mathrm{s}(\mathbf{b}) \rangle\right)\end{aligned} \tag{17.8}$$

The space-variant gradient ∇I_0 of the incident intensity $I_0(\mathbf{r}) = c\varepsilon/2|\mathbf{E}_\mathrm{i}(\mathbf{r})|^2$ is averaged over the particle volume in real space. With this first-order Born approximation one avoids the elaborate calculation of the electric fields on the surface of the scatterer or the exact internal fields. The second term in (17.8) describes the extinction and redistribution of momentum, given by the cross sections $C_\mathrm{ext} = C_\mathrm{sca} = Q_\mathrm{sca} \cdot \pi (d/2)^2$ together with the mean momentum vectors $\langle \mathbf{k}_\mathrm{i} \rangle$ and $\langle \mathbf{k}_\mathrm{s} \rangle$ of the incident and scattered fields, respectively [10]. They are all functions of the far-field spectra $\tilde{\mathbf{E}}_\mathrm{ext}(k_x, k_y, \mathbf{b}) = \mathrm{FT}\left\{\mathbf{E}_\mathrm{i}(x, y, b_z) \cdot q(x - b_x, y - b_y)\right\}$ for extinction and $\tilde{\mathbf{E}}_\mathrm{s}(k_x, k_y, \mathbf{b})$ for scattering, (FT = Fourier transform). $q(x, y)$ is an extinction function removing the fraction Q_sca from the incident intensity $|\mathbf{E}_\mathrm{i}(x, y, b_z)|^2$ at the

position (b_x, b_y) of the scatterer, $(q(x,y))^2 = 1 - Q_{\mathrm{sca}} \cdot \mathrm{step}(d/2 - r)$, $r = (x^2 + y^2)^{1/2}$, step(r) = Heaviside step function.

Both from (17.7) and (17.8) the optical forces on a Rayleigh scatterer [76–78] can be derived and simplified to

$$\mathbf{F}(\mathbf{b}) = \frac{n_m}{2c} \nabla I_{\mathrm{i}}(\mathbf{b}) \cdot \alpha_0 \cdot V + \frac{n_m}{6\pi c} \cdot I_{\mathrm{i}}(\mathbf{b}) \cdot k_n^3 \langle \mathbf{k}_{\mathrm{i}} \rangle \cdot (\alpha_0 \cdot V)^2 \tag{17.9}$$

with volume $V = 4/3\pi a^3$ (a is the radius of the scatterer), $C_{\mathrm{sca}} = (2/3)^3 \cdot \pi \cdot a^6 \cdot k_n^4 \cdot \alpha^2$ and with momentum transfer $\langle \mathbf{k}_{\mathrm{i}} \rangle$. In the center of the focus, $\langle \mathbf{k}_{\mathrm{i}} \rangle$ is the mean k-vector, which is shortened due to the Gouy-phase shift.

A direct derivation from Maxwell equations leads to the following integral for the optical force:

$$\mathbf{F}(\mathbf{b}) = \mathrm{Re} \iint_A \varepsilon \left(\mathbf{E}^* (\mathbf{n} \cdot \mathbf{E}) - \tfrac{1}{2} \cdot \mathbf{n} |\mathbf{E}|^2 \right) + \left(\mathbf{B}^* (\mathbf{n} \cdot \mathbf{B}) - \tfrac{1}{2} \mathbf{n} |\mathbf{B}|^2 \right) \mathrm{d}A \tag{17.10}$$

The integrand is the Maxwell stress tensor $\hat{T}$ acting on the vector n, which is normal on the surface element dA. The surface integral surrounds the scatterer in an arbitrary manner. The surface should have a large distance to the scatterer so that near fields can be neglected. The electric field $\mathbf{E} = \mathbf{E}_{\mathrm{i}} + \mathbf{E}_{\mathrm{s}}$ as well as the magnetic field $\mathbf{B} = \mathbf{B}_{\mathrm{i}} + \mathbf{B}_{\mathrm{s}}$, are the sums of the respective scattered and the incident fields. All fields must be evaluated on the enclosing surface. The asterix* indicates the complex conjugate field. Although (17.10) delivers the correct optical force for every particle, the integral is less intuitive and rather complicated to solve [79,80].

While (17.10) is universal, but difficult to solve and not descriptive, the two components approach of (17.8) is limited to dielectric particles not larger than the wavelength. However, it is more descriptive and has proven to achieve a quantitative agreement with experiments in all three dimensions [81]. It separately describes the scattering force and the gradient force. The first results from a momentum transfer from the incident photons onto the scatterer while the latter is the dipole force acting on each volume element by drawing it towards the brightest point of the incident intensity. The gradient force, therefore, enables stable trapping, whereas the scattering force pushes the scatterer out of the trap in the direction of light propagation (radiation pressure).

The effect of the two components is illustrated by the axial force profile in Fig. 17.7 left, where the point $F_z(\mathbf{b}_0) = 0$ is behind the center of the focus. The profiles are plotted with the dimensionless trapping efficiency $\mathbf{Q}(\mathbf{b}) = \mathbf{F}(\mathbf{b}) \cdot c/(P \cdot n_m)$, where c/n_m is the speed of light in the immersion medium and P is the total incident optical power P. For particles smaller than the wavelength, the scattering force is always positive and the gradient force is always bipolar. The force constants κ_{i} at the trapping position $\mathbf{b}_0 = (0, 0, b_{z0})$ are obtained according to $\kappa_{\mathrm{i}} = \partial_{\mathrm{i}} F_{\mathrm{i}}(x_{\mathrm{i}})|_{b0} = \partial_{\mathrm{i}} \left[F_{\mathrm{i,grad}}(x_{\mathrm{i}}) + F_{\mathrm{i,sca}}(x_{\mathrm{i}}) \right]\Big|_{b0}$ and are indicated on the right in Fig. 17.7 . Calculations indicate that in lateral direction $\partial_{\mathrm{i}} F_{\mathrm{i,grad}}$ and $\partial_{\mathrm{i}} F_{\mathrm{i,sca}}$ have opposite signs, whereas in axial direction they have the same sign. Therefore, the

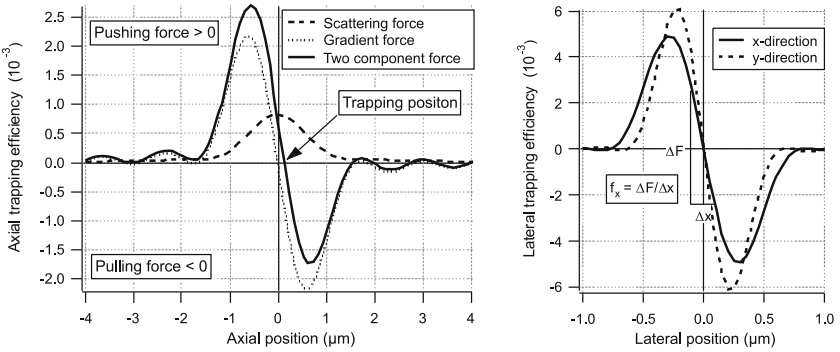


Fig. 17.7. Force profiles calculated with the two component approach for a 216 nm latex bead ($n = 1.57$) in axial and lateral directions. The trapping forces refer to a water immersion lens with NA = 1.2, illuminated with a Gaussian beam with waist $w_0 = 2 \cdot R$ (R = aperture radius), i.e. the aperture is over-illuminated, the wavelength is $\lambda_0 = 1.064\,\mu$m, the field is linearly polarized in x. The axial profile (left) represents the sum of the scattering and the gradient force. The lateral profiles (right) differ due to the field polarization, the ratio of the slopes f_x and f_y is about 60%

scattering force decreases the trap stiffness in lateral direction, but increases the stiffness in axial direction [81].

In Fig. 17.7, a small bead is trapped in water with a water immersion lens and for an incident wave polarized in x. The lateral force profiles on the right hand side clearly show that in the direction of polarization the maximum force and the force constant are weaker. The force constants κ_i are connected in a nonlinear relation to the maximum (backward) force and the depth of the trapping potential. The latter is decisive for the stability of an optical trap. The three force constants κ_i are strongly affected by spherical aberrations, which often occur in trapping experiments due to refractive index mismatch [82].

17.3.5 Thermal Noise

We assume to have a small particle of mass m in a viscous medium where the friction γ damps the particle thus that $\gamma \cdot \dot{x} >> m \cdot \ddot{x}$. In addition to that, the driving force for the particle's motion is assumed to be a random function of time. This situation is described by the Langevin equation [83]:

$$\gamma(\mathbf{r}) \cdot \dot{\mathbf{r}}(t) + \mathbf{F}_{\text{opt}}(\mathbf{r}) + \mathbf{F}_{\text{ext}}(\mathbf{r}, t) = \mathbf{F}_{\text{rand}}(t) \tag{17.11}$$

$\mathbf{F}_{\text{opt}}$ is the optical force in the trap, $\mathbf{F}_{\text{ext}}$ is the external force one wants to measure, $\mathbf{F}_{\text{th}}$ is the thermal random force and $\dot{r}(t)$ is the velocity of the particle. The viscous drag of the immersion fluid (also friction constant) is $\gamma = 6\pi \cdot a \cdot \eta$ (a = sphere radius, η = viscosity) for a spherical particle, which is Stokes' law. The viscous drag increases significantly close to interfaces and $\gamma(\mathbf{r})$ becomes a space-dependent

tensor [52]. However, a few particle diameters away from the surface, γ is constant. The diffusion constant $D(\mathbf{r})$, defined by the Stokes-Einstein-relation,

$$D = k_B T/\gamma \tag{17.12}$$

is related to the strength of the collision forces, which are uncorrelated according to $\langle \mathbf{F}_{th}(t) \cdot \mathbf{F}_{th}(t+\tau)\rangle = 6D\gamma\delta(\tau)$. k_B is the Boltzmann constant, T is the temperature in Kelvin and $\delta(\tau)$ is the delta function. However, due to friction, and even for $\mathbf{F}_{opt} = 0$ and $\mathbf{F}_{ext} = 0$, the resulting small displacements $\mathbf{r}(t)$ of the particle are correlated, i.e. the particle does not undergo a pure random walk. Therefore, the position correlation is linearly decaying with the time period τ according to $\langle \mathbf{r}(t) \cdot \mathbf{r}(t+\tau)\rangle = \text{const.} - 3D \cdot \tau$. In other words, the memory of the particle's motion, which is proportional to $-D$, can be extended by lowering the temperature T or increasing the friction (see (17.12)).

17.4 Experimental Setup and Techniques

17.4.1 Mechanics and optics

Figure 17.8 summarizes the main mechanical and optical components of a typical setup for optical trapping, tracking and imaging. A low noise Nd:YAG-laser (RMS-noise $\approx 0.03\%$) operating in the near infrared ($\lambda_0 = 1064\,nm$) is used for optical trapping and interferometric position detection. The IR-beam passes a first beam-expander and then an acousto-optic modulator (AOM), which, in combination with a reference diode (RD) and an electronic feedback, stabilizes the output power of the laser. Subsequent devices are a beam steering or shaping device (BSD), a second beam expander (Exp) and a dichroic beam splitter (BS). The aperture of the BSD is imaged with a telecentric system onto the back focal plane (BFP) of the objective lens (OL, infrared water immersion, NA = 1.2). The object plane (OP), defined by a standard glass cover slip, can be moved with a xyz-scanner (stage), consisting of a coarse manual xy-translation stage and a fine xyz-piezo scanner. Micromanipulation can also be performed in an open chamber with a patch-pipette (PP). A detection lens (DL, water immersion, NA = 0.3 – 0.9) collects scattered and unscattered light of the trapping laser, which is projected with a lens system (L) and by a dichroic beam splitter onto the InGaAs quadrant photodiode (QPD). Either visible laser light (Argon-Ion laser and a Helium-Neon) in combination with an acousto-optic tunable filter (AOTF) or incoherent light of a Hg-lamp is expanded and is directed via dichroic beam splitters to illuminate the object plane. Fluorescence light excited in the object plane is imaged with the objective lens and the tube lens (TL) by a CCD-camera. The visible light of a halogen lamp is linearly polarized (P) and passes both a mirror (M) and the dichroic beam splitter. A Wollaston prism (WP) near the BFP of the detection lens splits the beam for DIC microscopy. After transmitting the object, a second Wollaston prism behind the objective lens recombines the two beams. After passing a mirror and the DIC analyzer (A), the light is focused by the tube lens onto the CCD-camera.

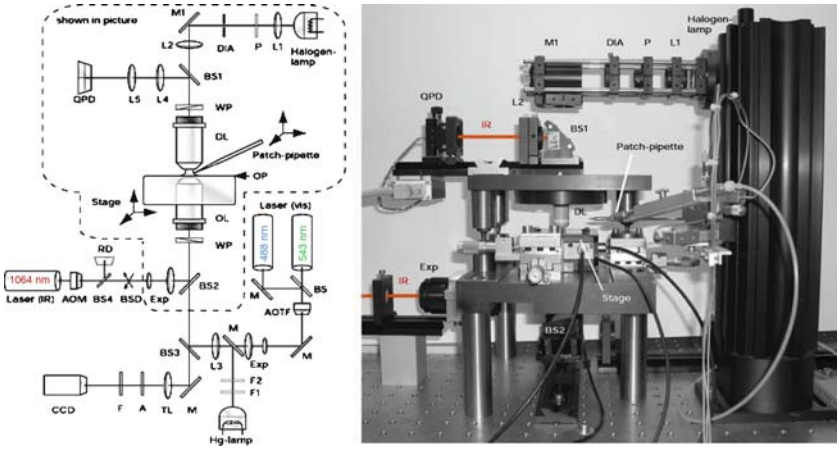


Fig. 17.8. Schematic and photo of a typical setup for optical trapping and tracking. Light paths from four light sources (infrared laser, visible lasers, Hg-lamp and halogene lamp) pass several optical and mechanical elements: acousto-optical modulator (AOM), acousto-optical tunable filter (AOTF), beam steering device (BSD), beam expander (Exp), dichroic beam splitters (BS), reference diode (RD), objective lens (OL), xyz-scan stage (stage), object plane (OP), patch-pipette (PP), detection lens (DL), mirrors (M), lenses (L), quadrant photodiode (QPD), reference diode (RD), polarizer (P), Wollaston prisms (WP), analyzer (A), tube lens (TL), filters (F), CCD-camera (CCD)

The influence of the setup's mechanical vibrations on the position signals can be tested by fixing a bead on the coverslip in the center of the focus. The tiny vibrations of the environment are transferred via the coverslip to the bead. Over a range of seconds, rms-noise <3 nm was measured with the interferometric detection system. By neglecting noise below 2 kHz, the remaining mechanically induced motion signals of the bead were <1 nm in all three dimensions [57].

17.4.2 Lasers and Probes

Trapping wavelengths of $\lambda_0 = 0.8\,\mu\text{m}$ and $\lambda_0 = 1.06\,\mu\text{m}$ were found to be most appropriate because cell damage is relatively low in comparison to that caused by visible light [84]. A wavelength of $\lambda_0 = 800\,\text{nm}$ would result in stronger optical forces, but here is still a lack of suitable laser sources, i.e. lasers that offer a good performance at reasonable prices such as Nd:YAG lasers ($\lambda_0 = 1064\,\text{nm}$). For non-biological or some in-vitro applications shorter laser wavelengths in the visible range enable higher trapping forces at the same laser power. Typically, the particle is spherical, with a diameter between 0.05 and 2 times the incident wavelength and a refractive index sufficiently large in comparison to that of its environment. The immersion medium is typically water, aqueous solution or cytoplasm with indices $n_{\text{med}} = 1.33 - 1.39$. Refractive indices n_{s} of dielectric scatterers are $1.39 - 1.48$ for silica, about $1.57 - 1.60$ for polystyrene (latex) and 3.5 for silicon. For metallic

gold particles, such as gold beads, the complex refractive index is $n_s = 0.4 + 7.4i$ at $\lambda_0 = 1.05\,\mu m$ [85]. The large imaginary part corresponds to a strong absorption of light and leads to a high polarizability of the gold bead, but also to additional radiometric forces, which become significant for gold beads as large as the wavelength. The high polarizability of gold results in stronger light scattering and stronger optical forces. Gold beads down to 20 nm in diameter have been trapped stably [86]. Only small metallic scatterers with a diameter of less than $\lambda/5$ are trapped in the lateral center of the focus, larger metal spheres are trapped off-axis [87,88].

17.4.3 Electronics

Besides all necessary controllers that drive scanners, acousto-optical devices, cameras and lasers, the most interesting electronic component is required for amplifying and acquiring the position data from the quadrant photodiode. An InGaAs pin photodiode with a photosensitivity of 0.67 A/W is used for our setup, avoiding the low-pass filtering effect of Silicon diodes. Pre-amplification of the signals by 20 V/mA (up to 850 kHz) leads to a voltage of 13.4 V per 1 mW laser power at the detector. The output noise of the pre-amplifier is 1 mV. Intensity noise of the diode-pumped infrared laser is less than 0.03% in the range of 10 Hz – 1 MHz. Undulations below 10 Hz are stabilized with an electronic feedback system. In addition, slow drifts in the intensity signals can be subtracted from the QPD signal by simultaneously measuring the laser's output with the reference diode.

The performance limit of the detector due to electronic noise can be estimated by taking account of the scattering efficiency of a weak scatterer (radius $a = 150$ nm, index $n = 1.39$), which is e.g. Qsca (sca subscript) = 0.3% (at a radius $a = 150$ nm and $n = 1.39$). A mean axial sensitivity of $0.25\%/\mu m$ for this scatterer can be estimated (see Table I in [41]). Assuming that 20% of the light in the focus is projected onto the diode, a laser power of 40 mW in the focus will be reduced to 8 mW at the diode. Pre-amplification provides a voltage of $8\,mW \cdot 13.4\,V/mW = 107\,V$. A change of $0.25\%/\mu m$ results in $268\,mV/\mu m$. In our setup, the minimum voltage due to noise is 1 mV, hence the axial resolution limit due to noise is $1\,\mu m/268 = 3.7$ nm. For a scatterer with $a = 108$ nm, $n = 1.57$ at $P = 30$ mW, this value is $1\,\mu m/313 = 3.1$ nm. In the lateral direction, the sensitivity is better by a factor of 2.5 to 4 for all investigated scatterers [57].

The three pre-amplified signals for the x-, y- and z-positions are further amplified by a factor of $2 - 50$, depending on the particle parameters and the trapping laser power. The large DC-part is subtracted electronically from the z-signal. Afterwards, a fast data acquisition PC-card converts simultaneously up to 4 voltages at a rate of up to 5 MHz to 12 bit signals, which are then analyzed as described in the last part of the theory section.

17.4.4 Calibration of Trap and Position Detector

In order to measure external forces $\mathbf{F}_{ext}$ and small particle displacements $\Delta\mathbf{b}$, or, to record complete particle trajectories $\mathbf{b}(t)$ for deriving interaction potentials $W_{ext}(\mathbf{b})$,

it is necessary to calibrate both the optical trap and the position detection system [28,89]. In nearly every application one aims to configure the experiment and the setup such that the force and the position measurements can be performed in the linear regime. More precisely, this means that both the signal **S** at the detector and the force **F** are proportional to the position **b**. We find, $\mathbf{S}(\mathbf{b}) = \hat{\mathbf{g}} \cdot (\mathbf{b} - \mathbf{b}_0) \approx (g_x \cdot b_x, g_y \cdot b_y, g_z \cdot (b_z - b_{0z}))$ and $\mathbf{F}(\mathbf{b}) = \hat{\boldsymbol{\kappa}} \cdot (\mathbf{b} - \mathbf{b}_0) \approx (\kappa_x \cdot b_x, \kappa_y \cdot b_y, \kappa_z \cdot (b_z - b_{0z}))$. $\mathbf{b}_0 = (0, 0, b_{0z})$ is the center of the optical trap. If the displacements are not too large, the tensors $\hat{\mathbf{g}}$ and $\hat{\boldsymbol{\kappa}}$ are diagonal matrices and, therefore, the components of the force constants κ_i do not couple, as well as the components of the position sensitivities $g_i (i = x, y, z)$.

It is likely useful to separate calibration techniques in those that exploit thermal noise and those that do not use the Brownian motion of the particle inside the trap. In the following, we want to present the most important calibration techniques. Therefore, we return to the equation of motion (11) with zero external force $\mathbf{F}_{\text{ext}}$ and disregard its vector character:

$$\gamma \cdot g^{-1} \cdot \dot{S}(t) + \kappa \cdot g^{-1} \cdot S(t) = F_{\text{drive}}(t) \tag{17.13}$$

The force $F_{\text{drive}}(t)$ on the right hand side, driving the particle displacements $b = g^{-1} \cdot S$, can either be the thermal random force $F_{\text{th}}(t)$ or a defined force generated by the user.

Calibration without Thermal Noise The easiest way to calibrate the detection system is to displace the particle by a defined Δb and measure the resulting signal $\Delta S(\Delta b)$. Then the calibration factor, i.e. the sensitivity is $g = \Delta S / \Delta b$. The particle is typically displaced by fixing it to a coverslip and moving the (piezo-electric) stage with the coverslip by Δb. The signals S are recorded by a QPD or by taking two images of the particle with a CCD camera before and after the displacement.

The viscous drag method allows us to calibrate the optical trap, i.e. to determine the trap stiffness κ. If the stage and the fluid medium are moved sinusoidally with the velocity $v = \omega_0 \cdot x_0 \cdot \cos(\omega_0 t)$, then the driving force displacing the trapped particle $F_{\text{drive}}(t) = F_0 \cdot \cos(\omega_0 t)$ is equal to the viscous drag force $v \cdot \gamma$. Equation (17.13) represents a damped driven oscillator with driving frequency ω_0, damping γ and position $b(t)$. From the solution

$$b(t) = \frac{F_0}{\sqrt{\omega_0^2 \gamma^2 + \kappa^2}} \sin(\omega_0 t - \varphi) \; ; \quad \varphi = -\arctan\left(\frac{\kappa}{\omega_0 \gamma}\right) \tag{17.14}$$

it is possible to deduce the desired κ provided γ is known.

Calibration with Thermal Noise If one of the calibration factors κ and g is known, it is possible to determine the other one by the equipartition theorem:

$$\frac{1}{2} k_{\text{B}} T = \frac{1}{2} \kappa \cdot \langle b^2 \rangle \tag{17.15}$$

The energy of $1/2\ k_B T$ is distributed equally in every degree of freedom, which are three translational and three rotational for a stiff particle. $\langle b^2 \rangle$ is the variance of all particle displacements in the trap and can be obtained from a position histogram (see Fig. 17.9b). From this value the stiffness of the trap can be obtained by $\kappa = k_B T / \langle b^2 \rangle$. A great advantage is that it does not depend on the viscous drag γ and calibration is possible also at varying viscosity and at low temporal sampling rates as it is typical for CCD-cameras. If κ is known, the sensitivity g can be obtained via $\left\langle S^2 \right\rangle / g^2 = k_B T / \kappa$, where $\left\langle S^2 \right\rangle$ is the variance of the signal histogram:

$$g = \sqrt{\frac{\langle S^2 \rangle \kappa}{k_B T}} \tag{17.16}$$

The assumption of a linear relationship between trapping force F_{opt} and displacement b entails the assumption of harmonic trapping potentials $W(b) = 1/2\ \kappa \cdot (b - b_0)^2 + \text{const}$. If we further assume that the trapped particle is in thermal equilibrium, then the probability density $p_B(b)$ (or the position histogram) to find a particle at position b obeys Boltzmann statistics and is related to the trapping potential $W(b)$. $W(b)$ and the force in arbitrary direction $\mathbf{F}(b)$ is obtained as follows [9]:

$$W(b) = -k_B T \cdot \ln\left(p_B(b)\right) + W_0 \quad \Leftrightarrow \quad \mathbf{F}(b) = -k_B T \cdot \frac{\nabla(p_B(b))}{p_B(b)} \approx \kappa \cdot b \tag{17.17}$$

The parameter p_0 normalizes the distribution p_B. Figure 17.9a shows the position track $\mathbf{b}(t)$ of the bead. The three-dimensional distribution of positions (Fig. 17.9b) corresponds to the three-dimensional probability density. Tracks at the very edge of the distribution (see Fig. 17.9e) correspond to high thermal energies of the diffusing particle, which can achieve about $8\,k_B T$ within a time period of few seconds. The potential $W(b)$ can also be the sum of the trapping potential $W_{opt}(b)$ and an external potential $W_{ext}(b)$ as pointed out in Sect. 17.5.2.

If a fast position detector is available that samples in the kHz range (typically a QPD or a PSD), a correlation analysis of the fluctuation signals enables an elegant way of determining very precisely the trap stiffnesses κ_i in all three directions $i = x, y, z$ [81]. The position autocorrelation function ACF of a diffusing particle in a harmonic potential $W(b_i) = 1/2\ \kappa_i \cdot (b_i - b_{i0})^2$ decays exponentially (Fig. 17.9d). If the position signal $S_i(b_i) = g_i \cdot b_i$ is linear, one can fit an exponential function to the ACF [45,90]

$$\langle S_i(t) \cdot S_i(t+\tau) \rangle = \left\langle |S_i(0)|^2 \right\rangle \exp(-\tau/\tau_i)\,. \tag{17.18}$$

The autocorrelation time in direction i is $\tau_i = \gamma/\kappa_i$. Brownian dynamics simulations reveal that the sampling time of the position detection should be at least five times the autocorrelation time of the trap or the external potential. The total sampling period T should be at least $2 - 6$ seconds depending on the trap stiffness. However, the reproducibility of finding κ is still improved for times $T > 20\,\text{s}$. Determining trap stiffnesses with the ACF is equivalent to determining trap stiffnesses with the mean square displacement MSD. The ACF is depicted in Fig. 17.9d.

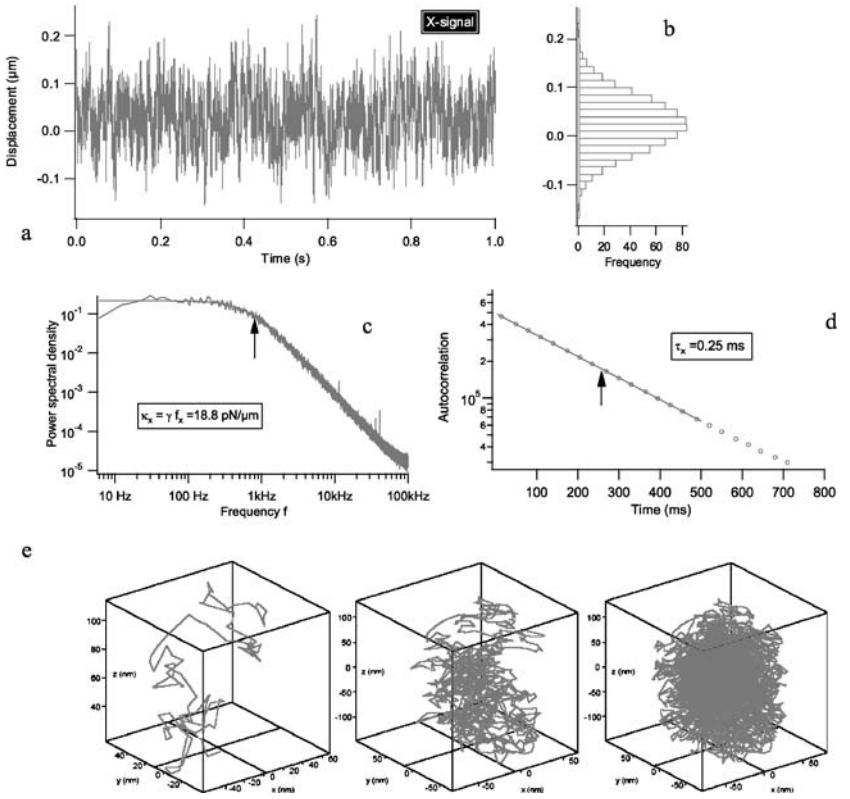


Fig. 17.9. (**a**) Measured positions of a fluctuating particle inside a soft optical trap as a function of time and (**b**) Gausssian shaped position histogram. (**c**) The corresponding power spectral density (PSD) has a linear decay in a double logarithmic plot indicating free diffusion of the particle for high frequencies. The arrow indicates the characteristic frequency ω_x defining the stiffness of the optical trap in this direction at 3 dB. (**d**) The auto-correlation function (ACF) of the same data shows an exponential decay. The arrow indicates the autocorrelation time τ_x of the particle inside the optical trap, i.e. the decay of the normalized ACF to 1/e. (**e**) Position trajectories of a particle inside an optical trap for the first 100, 1000 and 5000 points at a time sampling of dt = 50 µs

The trap stiffness κ_i can also be evaluated by a power spectrum analysis. Performing the Fourier transform and the modulus square of (17.13) provides the force spectral density $|\tilde{F}_{\mathrm{th}}|^2 = \gamma \cdot k_B T/\pi$ in units of N^2·s and the power spectral density (PSD) of the signal component $S_i(t)$:

$$\left|\tilde{S}_i(\omega)\right|^2 = \frac{\gamma \cdot k_B T/\left(\pi \cdot g_i^2\right)}{\kappa_i^2 + \gamma^2\omega^2} = \frac{D/\left(\pi \cdot g_i^2\right)}{\omega_i^2 + \omega^2} \tag{17.19}$$

By fitting a Lorentzian to $\left|\tilde{S}_i(\omega)\right|^2$ in a double logarithmic plot, it is possible to obtain the corner frequency $\omega_i = 2\pi f_i = 2\pi \cdot \gamma \cdot \kappa_i$ at a decay of 3dB. The power spectrum density is depicted in Fig. 17.9c.

17.4.5 Time-Multiplexed and Holographic Optical Traps

In various applications in biology and physics there is need to manipulate or sort several particles in parallel [56,91], to induce many body effects and phase transitions [92,93], to enable particle transport and noise induced phenomena [94], or, to orientate non-spherical structures in addition to a simple translation [95].

Two competing methods take part in a continuing revolution, presenting readers with novel research tools and applications. The first is the rather old technique of time-multiplexed optical tweezers [96], where an optical point trap is translated in lateral direction at rates between 50 Hz – 50 kHz with either scanning mirrors or acousto-optical deflectors (AOD). The other technique uses a phase hologram in the back focal plane of the trapping lens [97], enabling not only a distribution of point traps in 3D space (see Fig. 17.10), but also the generation of nearly arbitrary landscapes of optical energy potentials. At this point an alternative method [98] must be mentioned, which produces optical trap arrays using a generalized phase contrast (GPC) in a plane conjugated to the focal plane. The resulting traps lack of high angular components of the incident beam and produce traps that are only stable in two dimensions. However, by adding a second optical system on the side opposite to the focal plane, the radiation pressures from both directions cancel out and the GPC method allows also 3D trapping [99].

As already mentioned, holographic optical traps (HOTs) allow 3D trapping in 3D space, whereas time-multiplexed traps (TMTs) enable 3D trapping in 2D space (in the focal plane). In addition, TMTs give a continuous kick to the trapped particles while translating to the next location. Especially in Brownian systems and for slow trap displacements, this interferes with establishing thermal equilibrium.

In their infancy, static HOTs were realized by using diffractive optical phase elements, which were produced by printing or reactive ion etching. Only dynamic

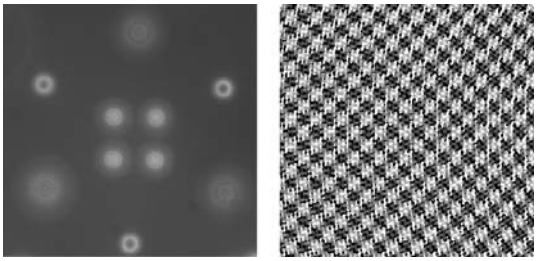


Fig. 17.10. Holographic optical trapping. *Left*: Parallel trapping of ten polystyrene spheres (each 3 μm in diameter) in three different axial planes spaced 5 μm apart. *Right*: Corresponding phase hologram produced by a spatial light modulator (phase range from 0 to 2π indicated in gray levels, pictures by courtesy of C. Schmitz and J. Curtis, University of Heidelberg, Germany)

HOTs [100–102] allowed the manipulation and movement of several particles in parallel. Modern spatial light modulators (SLMs) are able to retard pixel wise an incident wavefront as it is reflected from the SLM. Finer pixelation and increasing phase shifts of the SLMs provide higher diffraction efficiency and the generation of more refined optical traps, which allows trap displacements in the nanometer regime.

Another interesting feature offered by HOTs is the capability to transfer a defined orbital angular momentum from a special beam to the particle [103–105]. With these higher order Laguerre-Gaussian beams which are characterized by their helical wavefronts, toroidal optical traps are formed, called optical vortices. The phase of such beams $\exp\{i(kz + \theta\ell)\}$ has a winding number ℓ called the topological charge which quantifies the helicity of the beam's wavefronts. Since each photon carries orbital angular momentum $\hbar\ell$, the total torque of the beam is proportional to its intensity and can be precisely controlled by ℓ. The orbital angular momentum can exert stronger torques than the corresponding spin angular momentum ($\pm\hbar$), although both can be tuned by increasing the total photon flux by adjusting the power The diameter of the toroidal trap scales linearly with ℓ, and depending on the optics can be many 10's of microns in size [105]. When the toroidal trap is larger in diameter than the trapped particle, particles become confined to the ring. In such cases, the particles can only be maintained in two dimensions pressed up the coverglass where they are driven around the ring by transferred orbital angular momentum. The flexibility of dynamic HOTs makes the creation of optical vortices as well as other types of unique optical manipulation tools (Bessel beams, modulated optical vortices, Matthieu beams, cogwheel traps to name a few) relatively straightforward.

Although the manipulation of several particles in parallel can be accomplished elegantly with HOTs, a so far unsolved problem aggravates their applications in biology and biophysics: fast and precise tracking of several particles in parallel remains restricted to image based systems, which typically are CCD cameras operating at rates below 100 Hz. And, video based position tracking in axial direction remains awkward and less precise. As pointed out in the earlier sections, the steering of optical traps is usually less important and critical than the 3D tracking of the particle. The position of the particles encodes the measured force or the interaction potential. Parallel tracking of several particles is not possible with a single QPD, but would require an array of QPDs placed in a plane conjugated to the focal plane.

A versatile alternative is time chopping, which is automatically provided for TMTs. The optical trap generates an interference pattern in the BFP of the detection lens while it scans over the particle or rests a short time at the particle. The temporally chopped interference signals are recorded with a single QPD and allow fast parallel position tracking of several particles [106,107] even in 3D.

17.5 Applications in Brownian Systems

Photonic Force Microscopy is a special variant of an optical tweezers based microscope. In a PFM the optical trap is used to make unlikely interaction events more

likely, or in other words, it is used to bring two interaction partners together. At least one of them is a Brownian particle, which approaches the other by diffusion. The time until the partners get into contact is greatly reduced by bringing them together with an optical trap. The concept of position and orientation fluctuations is still maintained, but now within a small volume (inside the optical trap) and close to the binding partner. As soon as the partners come into contact, the very precise high-speed tracking system of the PFM records the fluctuation changes due to the mechanical, electrostatic, electrodynamic, or entropic interactions of the binding partners. Therefore it is justified to identify the PFM as a system being able to image interactions – besides a conventional imaging of structures. In this last section we present two examples of Brownian systems. They are related to cell biology and biophysics and describe a long ranging, smooth interaction, and, a short ranging interaction, respectively.

17.5.1 Particle Binding and Uptake by a Living Cell

Macrophages, as a part of our immune system, are cells that take up bacteria and small particles. This process is called phagocytosis. After the initial binding of the particle to the macrophage's cell membrane, the cell engulfs the particle if certain mechano-chemical conditions are met. In the positive case, the particle will be transported (as a so-called phagosome) by various motor proteins along different cytoskeletal filaments from the periphery of the cell to the perinuclear region. During this transport, the phagosome fuses with other organelles to digest the intraphagosomal object (e.g. the bacterium).

The processes of binding, uptake and intracellular transport of 1 μm sized spheres are studied by using photonic force microscopy. With an optical trap, a particle is brought in close proximity to the cell membrane, where the particle binds to either the flat membrane or to a tentacle like structure called protrusion. Upon binding, thermal position fluctuations of the softly trapped particle change in amplitude and time [50]. The fluctuation amplitudes decrease as well as the temporal auto-correlation due to the contact with the membrane. In addition the mean fluctuation value (the mean position) changes, since the cell actively moves the particle with force $\mathrm{F}_{\mathrm{ext}}(\mathbf{r}, t)$ as introduced by (17.11). The position fluctuations are recorded interferometrically with a QPD. The weak trap (laser power $P < 1\,\mathrm{mW}$) does not inhibit the uptake process and after binding the focused IR-light can be used for 3D tracking only. Because of large and active displacements of the particle, an automatic feedback system shifts the stage with cell and particle relative to the trap center. Figure 17.11 illustrates binding and uptake by (a) DIC-images, (b) a schematic and (c) a plot of temporal fluctuation changes. Since the system is hardly reaching thermal equilibrium, a correlation or a PSD analysis over time-windows of $1 - 2$ seconds must be performed (see Sect. 17.4.4). The fluctuations encode information about local rearrangements of the cytoskeleton and about membrane receptors and molecular motors that exert forces on the particle during binding, uptake and intracellular transport.

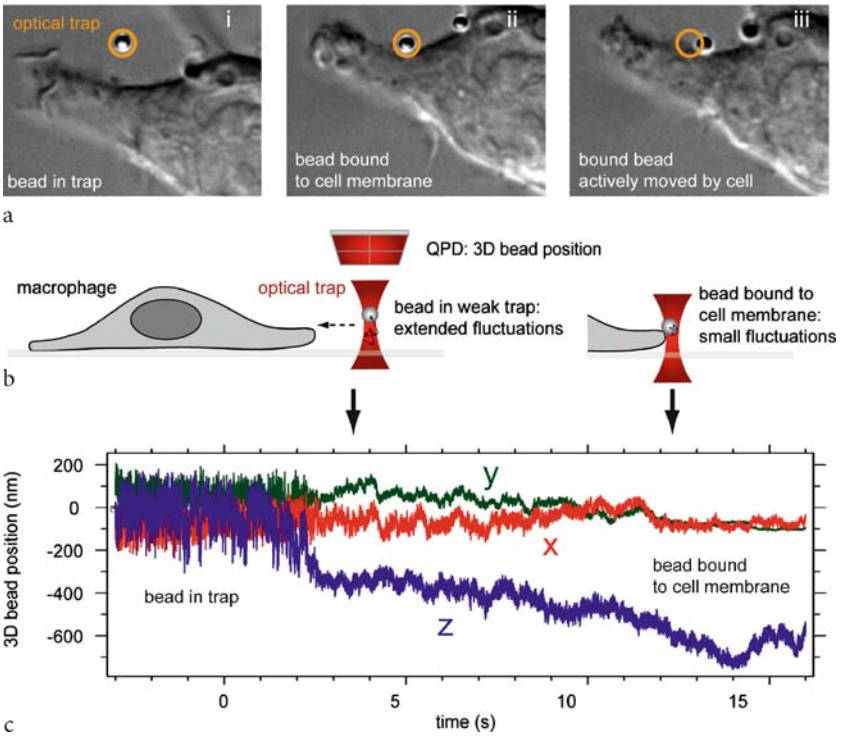


Fig. 17.11. Induction of a binding event and interferometric position tracking. (**a**) DIC image sequence of a trapped particle i) before binding to the membrane of a living cell, ii) during the binding process and iii) during the active response of the macrophage. (**b**) Experimental scheme: the particle is moved close to the cell membrane by an optical trap. The 3D thermal position fluctuations of the particle before and after the binding process are recorded with a QPD. (**c**) Position traces as a functions of time. (Pictures by Holger Kress, EMBL Heidelberg)

17.5.2 Imaging Nano-Mechanical Properties of Single Molecules

Understanding the elasto-mechanical properties of the motor protein Myosin-II helps us to understand how structural changes of this protein influence its molecular function. Depending on the type of Myosin, it can consist of 1 or 2 heads, which are connected via hinges to a stalk-like part. This so-called tail of the molecule consists of subunits, which have specific mechanical properties and functions.

A typical experimental assay is depicted in Fig. 17.12a [47–49]. A rod like macromolecule, e.g., Myosin-II, is attached at one end to an actin filament adhering to the coverslip, and to a bead of 500 nm in diameter at the other. The bead is trapped in 3D in a weak optical trap. The random trajectory of the bead is recorded in 3D with the PFM at a sampling rate of 100 kHz. An estimate of the bead's position autocorrelation time gives $\tau_z = \gamma \cdot \left\langle \Delta b_z^2 \right\rangle / k_B T \approx 3\,\text{ms}$, where γ is the Stokes viscous

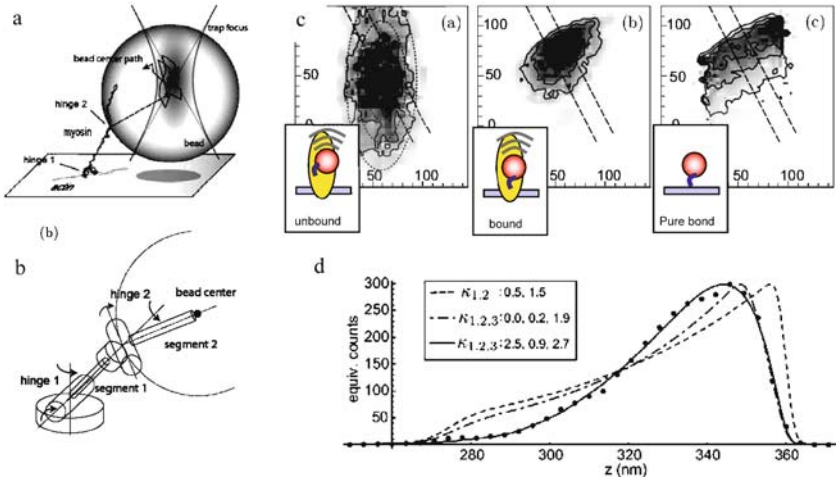


Fig. 17.12. Probing the elasto-mechanical properties of a motor protein. (**a**) An actin-myosin molecular tether, with a short segment of the Brownian bead center trajectory. (**b**) The same molecular bead system represented as a robot arm with two segments. (**c**) Three position histograms of the fluctuating bead before binding of the myosin to the surface, after binding and without the influence of an optical trap. Dark gray values indicate large position counts. (**d**): Line scan through the third histogram (points) and three fits derived from two different robotic models. (Pictures by Nils Becker, EMBL Heidelberg)

drag of the bead at room temperature. The bead position histogram of this trajectory re-presents an effective free energy landscape $W_{\mathrm{mes}}(\mathbf{b}) = W_{\mathrm{mol}}(\mathbf{b}) + W_{\mathrm{opt}}(\mathbf{b})$ for the bead according to (17.17), resulting from both the molecule and the optical trap. The three histograms of the bead center positions $p_{\mathrm{opt}}(\mathbf{b})$, $p_{\mathrm{mes}}(\mathbf{b})$ and $p_{\mathrm{mol}}(\mathbf{b}) \sim p_{\mathrm{mes}}(\mathbf{b})/p_{\mathrm{opt}}(\mathbf{b})$ are shown in Fig. 17.12c from left to right. It is remarkable that the histogram resulting from the pure molecular tether $p_{\mathrm{mol}}(\mathbf{b})$ is symmetric after removing the influence of the optical trap. The radial profile of $p_{\mathrm{mol}}(\mathbf{b})$, plotted with dots in Fig. 17.12d, encodes the visco-elastic information about myosin-II. Therefore, the histogram is an image of the enthalpic and entropic energy landscape of the molecule.

The nano-mechanical properties were identified by best-fits based on robotic models [49]. As depicted in Fig. 17.12b, the macro-molecule is assumed to consist of rigid segments concatenated with hinges. These hinges have angular restall forces with mean stiffnesses κ_i. As a problem of inverse kinematics, the end-point distribution of the sphere center reveals the number, the strengths and the positions of the hinges. Figure 17.12d shows the best fit results, clearly favoring the model with three hinges and angular spring constants $\kappa_{1,2,3}$. The lengths of the segments coincide with those obtained by electron-microscopy. However, our results were obtained under near physiological conditions and by using IR-light at 1.06 μm for position tracking of the tethered probe. The radial interaction profile in Figure 17.12d has a total length of 80 nm with structural details far beyond the optical resolution limit!

17.6 Summary and Outlook

In this disquisition of optical trapping we concentrated on investigating particles smaller than the incident wavelength. These particles with diameters of about 0.1 μm to 1 μm are of special interest for several reasons. Most cell organelles are in that size regime, colloidal suspensions consist of these particles and they are in the size regime equal to or below the optical resolution and therefore offer new imaging properties as local probes. But these particles are also interesting, since they are accompanied with an attractive theory on light focusing, scattering and optical forces, with coincides well with experiments [81]. Using computer simulations, this enables a more effective planning of experiments, of feasibility studies and data analysis, which is of great relevance since the trapping conditions become more critical for smaller particles and environmental forces gain more impact. Especially thermal forces leading to Brownian motion and surface forces play a crucial role, when optical manipulation has to be performed on the nanometer-scale. Control or even exploitation of thermal fluctuations of smallest particles is expected to become a key in nano-technology. Here, optical energy landscapes in combination with noise-enhanced phenomena and self-organization seem to become inseparable. As discussed in Sect. 17.4.5, holographic trapping is the top candidate for exerting unlimited variations of optical torques and forces on several particles of arbitrary shape. However, optical trapping alone is limited in its applications as long as 2D video microscopy is used as an optical control. Only 3D optical tracking of positions and orientations allows observing and interpreting processes correctly. A fast and precise control of all thermal and non-thermal displacements of the particles remains indispensable and still great efforts have to be undertaken, to improve current techniques and to realize new ideas. For probing of even smaller structures and samples relevant in (bio-) nanotechnology, novel probes have to be found, enabling a stronger interaction with light, resulting in more stable traps and more precise position tracking. Plasmon and exciton resonances have to be exploited while using metallic or semi-conductor materials for tiniest probes. A fascinating entanglement of basic research and technical applications on trapping and tracking of small particles is awaiting all of us!

Acknowledgements

The authors thank Jan Huisken and James Jonkman for comments on the manuscript as well as Holger Kress, Nils Becker, Stephan Altmann, and Jennifer Curtis for providing data.

References

1. Maxwell, J.C., “On stresses in rarified gases from inequalities of temperature”. Philos. T. R. Soc. London. 170:231–256 (1879)

2. Lebedev, P., "Untersuchungen über die Druckkräfte des Lichtes". Ann.Physik. **4**: 433 (1901)
3. Ashkin, A., "Acclearation and trapping of particles by radiation pressure". Phys. Rev. Lett. **24**, 4:156–159 (1970)
4. Ashkin, A. and J.M. Dziedzic, "Optical levitation by radiation pressure". Applied Physics Letters. **19**, 8:283–5 (1971)
5. Ashkin, A., "Applications of Laser-Radiation Pressure". Science. **210**, 4474:1081–1088 (1980)
6. Ashkin, A., "History of optical trapping and manipulation of small-neutral particle, atoms, and molecules". Ieee Journal of Selected Topics in Quantum Electronics. **6**, 6:841–856 (2000)
7. Ashkin, A., J.M. Dziedzic, J.E. Bjorkholm, and S. Chu, "Observation of a single-beam gradient force optical trap for dielectric particles". Opt.Letters. **11**, 5:288–290 (1986)
8. Bakker Schut, T.C., G. Hesselink, B.G. de Grooth, and J. Greve, "Experimental and theoretical investigations on the validity of the geometrical optics model for calculating the stability of optical traps". Cytometry. 12:479 (1991)
9. Rohrbach, A., "Switching and measuring a force of 25 femtoNewtons with an optical trap". Opt. Express. **13**, 24:9696–9701 (2005)
10. Rohrbach, A. and E.H.K. Stelzer, "Optical trapping of dielectric particles in arbitrary fields". J. Opt. Soc. Am. A. **18**, 4:839–853 (2001a)
11. Binnig, G. and H. Rohrer, "In touch with atoms". Reviews of Modern Physics. **71**, 2:S324–S330 (1999)
12. Ashkin, A. and J.M. Dziedzic, "Optical trapping and manipulation of viruses and bacteria". Science. **235**, 4795:1517–1520 (1987)
13. Ashkin, A., J.M. Dziedzic, and T. Yamane, "Optical trapping and manipulation of single cells using infrared laser beams". Nature. **330**, 12:769–771 (1987)
14. Block, S.M., D.F. Blair, and H.C. Berg, "Compliance of bacterial flagella measured with optical tweezers". Nature. **338**, 6215:514–518 (1989)
15. Seeger, S., S. Monajembashi, K.J. Hutter, G. Futterman, J. Wolfrum, and K.O. Greulich, "Application of laser optical tweezers in immunology and molecular genetics". Cytometry. **12**, 6:497–504 (1991)
16. Steubing, R.W., S. Cheng, W.H. Wright, Y. Numajiri, and M.W. Berns, "Laser induced cell fusion in combination with optical tweezers: the laser cell fusion trap". Cytometry. **12**, 6:505–510 (1991)
17. Tadir, Y., W.H. Wright, O. Vafa, T. Ord, R.H. Asch, and M.W. Berns, "Micromanipulation of sperm by a laser generated optical trap". Fertil Steril. **52**, 5:870–873 (1989)
18. Schutze, K., A. Clement Sengewald, and A. Ashkin, "Zona drilling and sperm insertion with combined laser microbeam and optical tweezers". Fertil Steril. **61**, 4:783–786 (1994)
19. Block, S.M., L.S. Goldstein, and B.J. Schnapp, "Bead movement by single kinesin molecules studied with optical tweezers". Nature. **348**, 6299:348–352 (1990)
20. Ashkin, A., K. Schutze, J.M. Dziedzic, U. Euteneuer, and M. Schliwa, "Force generation of organelle transport measured in vivo by an infrared laser trap". Nature. **348**, 6299:346–348 (1990)
21. Kuo, S.C. and M.P. Sheetz, "Force of single kinesin molecules measured with optical tweezers". Science. **260**:232–234 (1993)
22. Svoboda, K., C.F. Schmidt, B.J. Schnapp, and S.M. Block, "Direct observation of kinesin stepping by optical trapping interferometry". Nature. **365**, 6448:721–727 (1993)
23. Finer, J.T., R.M. Simmons, and J.A. Spudich, "Single Myosin Steps and Forces Measured with Optical Traps". Biophysical Journal. **66**, 2:A353–A353 (1994 error)

24. Simmons, R.M., J.T. Finer, S. Chu, and J.A. Spudich, "Quantitative measurement of force and displacement using an optical trap". Biophysical Journal. **70**:1813–1822 (1996)
25. Perkins, T.T., D.E. Smith, and S. Chu, "Direct Observation of Tube-Like Motion of a Single Polymer- Chain". Science. **264**, 5160:819–822 (1994)
26. Smith, S.B., Y. Cui, and C. Bustamante, "Overstretching B-DNA: The Elastic Response of Individual Double-Stranded and Single-Stranded DNA Molecules". Science. **271**:795–799 (1996)
27. Wang, M.D., H. Yin, R. Landick, J. Gelles, and S.M. Block, "Stretching DNA with optical tweezers". Biophysical Journal. **72**, 3:1335–1346 (1997)
28. Neuman, K.C. and S.M. Block, "Optical trapping". Rev. Sci. Instr. **75**, 9:2787–2809 (2004)
29. Malmqvist, L. and H.M. Hertz, "Trapped particle optical microscopy". Optics Communications. **94**:19–24 (1992)
30. Denk, W. and W.W. Webb, "Optical measurement of picometer displacements of transparent microscopic objects". Appl. Optics. **29**, 16:2382–2391 (1990)
31. Ghislain, L.P. and W.W. Webb, "Scanning-force microscope based on an optical trap". Optics Letters. **18**, 19:1678–1680 (1993)
32. Kawata, S., Y. Inouye, and T. Sugiura, "Near-Field Scanning Optical Microscope with a Laser Trapped Probe". Japanese Journal of Applied Physics Part 2-Letters. **33**, 12A:L1725–L1727 (1994)
33. Sugiura, T., T. Okada, Y. Inouye, O. Nakamura, and S. Kawata, "Gold-bead scanning near-field optical microscope with laser- force position control". Optics Letters. **22**, 22:1663–1665 (1997)
34. Friese, M.E., H. Rubinsztein-Dunlop, N.R. Heckenberg, and E.W. Dearden, "Determineation of the force constant of a single-beam gradient trap by measurement of backscattered light". Appl. Opt. **35**, 36:7112–7116 (1996)
35. Friese, M.E.J., A.G. Truscott, H. Rubinsztein-Dunlop, and N.R. Heckenberg, "Three-Dimensional Imaging with Optical Tweezers". Appl. Opt. **38**, 31:6597–6603 (1999)
36. Florin, E.-L., J.K.H. Hörber, and E.H.K. Stelzer, "High-resolution axial and lateral position sensing using two-photon excitation of fluorophores by a continuous Nd:YAG laser". Applied Physics Letters. **69**, 4:446–448 (1996)
37. Florin, E.-L., A. Pralle, J.K.H. Hörber, and E.H.K. Stelzer, "Photonic force microscope (PFM) based on optical tweezers and two-photon excitation for biological applications". Journal of Structural Biology. **119**, 2:202–211 (1997)
38. Allersma, M.W., F. Gittes, M.J. deCastro, R.J. Stewart, and C.F. Schmidt, "Two-dimensional tracking of ncd motility by back focal plane interferometry." Biophysical Journal. **74**, 2 Pt1:1074–1085 (1998)
39. Gittes, F. and C.F. Schmidt, "Interference model for back-focal-plane displacement detection in optical tweezers". Opt. Letters. **23**, 1:7–9 (1998)
40. Pralle, A., M. Prummer, E.-L. Florin, E.H.K. Stelzer, and J.K.H. Hörber, "Three-dimensional position tracking for optical tweezers by forward scattered light". Microscopy Research and Techniques. **44**, 5:378–386 (1999)
41. Rohrbach, A. and E.H.K. Stelzer, "Three-dimensional position detection of optically trapped dielectric particles". J. Appl. Phys. **91**, 8:5474–5488 (2002b)
42. Born, M. and E. Wolf, *Principles of optics*. 7 ed. 1999, New York: Cambridge University Press. 497–499.
43. Tischer, C., S. Altmann, S. Fisinger, J.K.H. Hörber, E.H.K. Stelzer, and E.-L. Florin, "Three-dimensional thermal noise imaging". Appl.Phys.Letters. **79**, 23:3878–3880 (2001)

44. Florin, E.-L., H.J.K. Hörber, and E.H.K. Stelzer, *Method for 3D Object-Scanning, United States Patent No US 6,833,923 B2*. 2004: USA.
45. Florin, E.-L., A. Pralle, E.H.K. Stelzer, and J.K.H. Hörber, "Photonic force microscope calibration by thermal noise analysis". Applied Physics A. **66**:S75–S78 (1998)
46. Veigel, C., M.L. Bartoo, D.C.S. White, J.C. Sparrow, and J.E. Molloy, "The stiffness of rabbit skeletal actomyosin cross-bridges determined with an optical tweezers transducer". Biophysical Journal. **75**, 3:1424–1438 (1998)
47. Jeney, S., Ernst H.K. Stelzer, H. Grubmüller, and E.-L. Florin, "Mechanical Properties of Single Motor Molecules Studied by Three-Dimensional Thermal Force Probing in Optical tweezers". ChemPhysChem. **5**:1150–1158 (2004)
48. Scholz, T., S.M. Altmann, C. Tischer, J.-K.H. Hörber, and B. Brenner, "Mechanical properties of single myosin molecules probed with the Photonic Force Microscope". Biophys. J. **88**:360–371 (2005)
49. Becker, N., S. Altmann, T. Scholz, H. Hörber, E. Stelzer, and A. Rohrbach, "Three-dimensional bead position histograms reveal single-molecule nano-mechanics". Phys.Rev.E. **71**, 2:021907 (2005)
50. Kress, H., E.H.K. Stelzer, G. Griffiths, and A. Rohrbach, "Control of relative radiation pressure in optical traps: Application to phagocytic membrane binding studies". Phys. Rev. E: 061927 (2005)
51. Gittes, F., B. Schnurr, P.D. Olmsted, F.C. MacKintosh, and C.F. Schmidt, "Microscopic Viscoelasticity: Shear Moduli of Soft Materials Determined from Thermal Fluctuations". Phys.Rev.Lett. **79**, 17:3286–3289 (1997)
52. Pralle, A., E.-L. Florin, E.H.K. Stelzer, and J.K.H. Hörber, "Local viscosity probed by photonic force microscopy". Applied Physics A. **66**:S71–S73 (1998)
53. Pralle, A., P. Keller, E.-L. Florin, K. Simons, and J.K.H.H. Hörber, "Sphingolipid-cholesterol rafts diffuse as small entities in the plasma membrane of mammalian cells". The Journal of Cell Biology. 148:997–1007 (2000)
54. Clapp, A.R. and R.B. Dickinson, "Direct Measurement of Static and Dynamic Forces between a Colloidal Particle and a Flat Surface Using a Single-Beam Gradient Optical Trap and Evanescent Wave Light Scattering". Langmuir. 17:2182–2191 (2001)
55. Hansen, P.M., J.K. Dreyer, J. Ferkinghoff-Borg, and L. Oddershede, "Novel optical and statistical methods reveal colloid-wall interactions inconsistent with DLVO and Lifshitz theories". Journal of Colloid and Interface Science. 287:561–571 (2005)
56. Crocker, J.C., J.A. Matteo, A.D. Dinsmore, and A.G. Yodh, "Entropic attraction and repulsion in binary colloids probed with a line optical tweezer". Physical Review Letters. **82**, 21:4352–4355 (1999)
57. Rohrbach, A., C. Tischer, D. Neumayer, E.L. Florin, and E.H.K. Stelzer, "Trapping and tracking a local probe with a Photonic Force Microscope". Rev. Sci. Instrum. **75**, 6:2197–2210 (2004b)
58. Huisstede, J.H.G., K.O.v.d. Werf, M.L. Bennink, and V. Subramaniam, "Force detection in optical tweezers using backscattered light". Opt. Express. **13**, 4:1113–1123 (2005)
59. Ghislain, L., N. Switz, and W. Webb, "Measurement of small forces using an optical trap". Review Of Scientific Instruments. **65, N9 (Sep)**: 2762–2768 (1994)
60. Sasaki, K., M. Tsukima, and H. Masuhara, "Three-dimensional potential analysis of radiation pressure exerted on a single microparticle". Applied Physics Letters. **71**, 1:37–39 (1997)
61. Ashkin, A., "Stable radiation-pressure particle traps using alternating light beams." Opt Lett. **9 (10)**:454–456 (1984)

62. Ashkin, A. and J.-M. Dziedzic, "Observation of radiation-pressure trapping of particles by alternating light beams." Phys Rev Lett. **54 (12)**:1245–1248 (1985)
63. Anglin, J.R. and W. Ketterle, "Bose-Einstein condensation of atomic gases". Nature. **416**, 6877:211–218 (2002)
64. McCutchen, C.W., "Generalized aperture and the three-dimensional diffraction image". Journ. Opt. Soc. Am. **54**, 2:240–244 (1964)
65. Goodman, J.W., in *Introduction to Fourier optics*, McGraw-Hill, Editor. 1968: San Francisco. p. 48–56.
66. Török, P., P. Varga, Z. Laczik, and G.R. Booker, "Electromagnetic diffraction of light focused through planar interface between materials of mismatched refraxtive indices: an integral representation". J. Opt. Soc. Am. **12**, 2:325–332 (1995)
67. Barton, J.P. and D.R. Alexander, "Fifth-order corrected electromagnetic field components for a fundamental Gaussian beam". J.Appl.Phys. **66**, 7:2800–2802 (1989b)
68. Bohren, C. and D.R. Huffman, *Absorption and scattering of light by small particles*, in *Absorption and scattering of light by small particles*. 1998, Wiley Science Paperback: New York. p. 158.
69. van de Hulst, H.C., *Light Scattering by Small Particles*. 1957, Leiden: Dover Publications, Inc. 123–125.
70. Rohrbach, A., H. Kress, and E.H.K. Stelzer, "Three-dimensional tracking of small spheres in focused laser beams: influence of the detection angular aperture". Optics Letters. 28:411–413 (2003)
71. Dreyer, J.K., K. Berg-Sorensen, and L. Oddershede, "Improved axial position detection in optical tweezers measurements". Appl. Opt. **43**, 10:1991–1995 (2004)
72. Kress, H., E.H.K. Stelzer, and A. Rohrbach, "Tilt angle dependent 3D-position detection of a trapped cylindrical particle in a focused laser beam". Appl.Phys.Letters. **84**, 18:4271–4273 (2004)
73. Gordon, J.P., "Radiation forces and momenta in dielectric media". Phys. Rev. **8**, 1:14–21 (1973)
74. Rohrbach, A., H. Kress, and E.H.K. Stelzer, "Reply to comment on : Trapping forces, force constants and potential depths for dielectric spheres in the presence of spherical aberrations". Appl. Opt. **43**, 9:1827–1829 (2004)
75. Jackson, J.D., in *Classial Electrodynamics*. 1975, Wiley&Sons: New York.
76. Harada, Y. and T. Asakura, "Radiation forces on a dielectric sphere in the Rayleigh scattering regime". Opt. Comm. 124:529–541 (1996)
77. Tlusty, T., A. Meller, and R. Bar-Ziv, "Optical gradient forces of strongly localized fields". Phys. Rev. Lett. **81**, 8:1738–1741 (1998)
78. Chaumet, P.C. and M. Nieto-Vesperinas, "Time-averaged total force on a dipolar sphere in an electromagnetic field". Optics Letters. **25**, 15:1065–1067 (2000)
79. Kim, J.S. and S.S. Lee, "Scattering of Laser Beams and the Optical Potential Well for a Homogeneous Sphere". J. Opt. Soc. Am. **73**, 3:303–312 (1983)
80. Barton, J.P., D.R. Alexander, and S.A. Schaub, "Theoretical Determination of Net Radiation Force and Torque for a Spherical Particle Illuminated by a Focused Laser Beam". J.Appl.Phys. **66**, 10:4594–4602 (1989a)
81. Rohrbach, A., "Stiffness of optical traps: Quantitative agreement between experiment and electromagnetic theory". Phys. Rev. Lett. 95:168102 (2005)
82. Rohrbach, A. and E.H.K. Stelzer, "Trapping forces, force constants and potential depths for dielectric spheres in the presence of spherical aberrations". Appl. Opt. **41**, 13:2494–2507 (2002a)
83. Reif, F., in *Fundamentals of statistical and thermal physics*. 1985, McGraw-Hill: Auckland. p. 560–565.

84. Neuman, K.C., E.H. Chadd, G.F. Liou, K. Bergman, and S.M. Block, "Characterization of Photodamage to Escherichia coli in Optical Traps". Biophys. J. **77**, 5:2856–2863 (1999)
85. Svoboda, K. and S.M. Block, "Optical trapping of metallic Rayleigh particles". Optics Letters. **19**, 13:930–932 (1994)
86. Hansen, P.M., V.K. Bhatia, N. Harrit, and L. Oddershede, "Expanding the optical trapping range of gold nanoparticles". Nano Letters. **5**, 10:1937–1942 (2005)
87. Sato, S., Y. Harada, and Y. Waseda, "Optical Trapping of Microscopic Metal Particles". Optics Letters. **19**, 22:1807–1809 (1994)
88. Ke, P.C. and M. Gu, "Characterization of trapping force in the presence of spherical aberration". J. Mod. Opt. **45**, 10:2159–2168 (1998)
89. Capitanio, M., G. Romano, R. Ballerini, M. Giuntini, F.S. Pavonea, D. Dunlap, and L. Finzi, "Calibration of optical tweezers with differential interference contrast signals". Rev. Sci. Instr. **73**, 4:1687–1696 (2002)
90. Bar-Ziv, R., A. Meller, T. Tlusty, E. Moses, J. Stavans, and S.A. Safran, "Localized dynamic light scattering: Probing single particle dynamics at the nanoscale". Physical Review Letters. **78**, 1:154–157 (1997)
91. Garces-Chavez, V., D. McGloin, H. Melville, W. Sibbett, and K. Dholakia, "Simultaneous micromanipulation in multiple planes using a self-reconstructing light beam". Nature. **419**, 6903:145–147 (2002)
92. Brunner, M., J. Dobnikar, H.H.v. Grünberg, and C. Bechinger, "Direct measurement of three-body interactions amongst charged colloids". Phys. Rev. Lett. 92:078301 (2004)
93. Mangold, K., P. Leiderer, and C. Bechinger, "Phase transitions of colloidal monolayers in periodic pinning arrays". Physical Review Letters. **90**, 15: - (2003)
94. Harada, T. and K. Yoshikawa, "Fluctuation-response relation in a rocking ratchet". Phys.Rev.E. 69:031113 (2004)
95. Agarwal, R., K. Ladavac, Y. Roichman, G. Yu, C.M. Lieber, and D.G. Grier, "Manipulation and assembly of nanowires with holographic optical traps". Opt. Express. **22**, 13:8906–8912 (2005)
96. Visscher, K., G.J. Brakenhoff, and J.J. Krol, "Micromanipulation by "multiple" optical traps created by a single fast scanning trap integrated with the bilateral confocal scanning laser microscope". Cytometry. **14**, 2:105–114 (1993)
97. Dufresne, E.R. and D.G. Grier, "Optical tweezer arrays and optical substrates created with diffractive optics". Review of Scientific Instruments. **69**, 5:1974–1977 (1998)
98. Eriksen, R., V. Daria, and J. Gluckstad, "Fully dynamic multiple-beam optical tweezers". Opt. Express. 10:597–602 (2002)
99. Rodrigo, P.J., V.R. Daria, and J. Glückstad, "Four-dimensional optical manipulation of colloidal particles". Appl. Phys. Lett. 86:074103 (2005)
100. Reicherter, M., T. Haist, E.U. Wagemann, and H.J. Tiziani, "Optical particle trapping with computer-generated holograms written on a liquid-crystal display". Optics Letters. **24**, 9:608–610 (1999)
101. Liesener, J., M. Reicherter, T. Haist, and H.J. Tiziani, "Multi-functional optical tweezers using computer-generated holograms". Optics Communications. **185**, 1–3:77–82 (2000)
102. Curtis, J.E., B.A. Koss, and D.G. Grier, "Dynamic holographic optical tweezers". Optics Communications. **207**, 1–6:169–175 (2002)
103. He, H., M.E.J. Friese, N.R. Heckenberg, and H. Rubinszteindunlop, "Direct Observation of Transfer of Angular-Momentum to Absorptive Particles from a Laser-Beam with a Phase Singularity". Physical Review Letters. **75**, 5:826–829 (1995)

104. Volke-Sepulveda, K., V. Garces-Chavez, S. Chavez-Cerda, J. Arlt, and K. Dholakia, "Orbital angular momentum of a high-order Bessel light beam". Journal of Optics B-Quantum and Semiclassical Optics. **4**, 2:S82–S89 (2002)
105. Curtis, J.E. and D.G. Grier, "Structure of optical vortices". Physical Review Letters. **90**, 13:133901 -133904 (2003)
106. Nambiar, R. and J.C. Meiners, "Fast position measurements with scanning line optical tweezers". Optics Letters. **27**, 10:836–838 (2002)
107. Neumayer, D., P. Kaiser, E.H.K. Stelzer, and A. Rohrbach, "Interferometric tracking of two particles with dynamic optical traps". Proc. SPIE. **5859**:33–39 (2005)

Index

Springer Series in

OPTICAL SCIENCES

Volume 1

1 **Solid-State Laser Engineering**
By W. Koechner, 5th revised and updated ed. 1999, 472 figs., 55 tabs., XII, 746 pages

Published titles since volume 100

100 **Quantum Interference and Coherence**
Theory and Experiments
By Z. Ficek and S. Swain, 2005, 178 figs., XV, 418 pages

101 **Polarization Optics in Telecommunications**
By J. Damask, 2005, 110 figs., XVI, 528 pages

102 **Lidar**
Range-Resolved Optical Remote Sensing of the Atmosphere
By C. Weitkamp (Ed.), 161 figs., XX, 416 pages

103 **Optical Fiber Fusion Splicing**
By A.D. Yablon, 2005, 137 figs., XIII, 306 pages

104 **Optoelectronics of Molecules and Polymers**
By A. Moliton, 2005, 229 figs., 592 pages

105 **Solid-State Random Lasers**
By M. Noginov, 2005, 131 figs., XII, 238 pages

106 **Coherent Sources of XUV Radiation**
Soft X-Ray Lasers and High-Order Harmonic Generation
By P. Jaeglé, 2006, 332 figs., XIII, 416 pages

107 **Optical Frequency-Modulated Continuous-Wave (FMCW) Interferometry**
By J. Zheng, 2005, 137 figs., XVIII, 254 pages

108 **Laser Resonators and Beam Propagation**
Fundamentals, Advanced Concepts and Applications
By N. Hodgson and H.Weber, 2005, 587 figs., XXVI, 794 pages

109 **Progress in Nano-Electro Optics IV**
Characterization of Nano-Optical Materials and Optical Near-Field Interactions
By M. Ohtsu (Ed.), 2005, 123 figs., XIV, 206 pages

110 **Kramers–Kronig Relations in OpticalMaterials Research**
By V. Lucarini, J.J. Saarinen, K.-E. Peiponen, E.M. Vartiainen, 2005,
37 figs., X, 162 pages

111 **Semiconductor Lasers**
Stability, Instability and Chaos
By J. Ohtsubo, 2005, 169 figs., XII, 438 pages

112 **Photovoltaic Solar Energy Generation**
By A. Goetzberger and V.U. Hoffmann, 2005, 139 figs., XII, 234 pages

113 **Photorefractive Materials and Their Applications 1**
Basic Effects
By P. Günter and J.P. Huignard, 2006, 169 figs., XIV, 421 pages

114 **Photorefractive Materials and Their Applications 2**
Materials
By P. Günter and J.P. Huignard, 2006, 370 figs., XVII, 640 pages

115 **Photorefractive Materials and Their Applications 3**
Applications
By P. Günter and J.P. Huignard, 2007, 316 figs., X, 366 pages

116 **Spatial Filtering Velocimetry**
Fundamentals and Applications
By Y. Aizu and T. Asakura, 2006, 112 figs., XII, 212 pages

Springer Series in

OPTICAL SCIENCES

117 **Progress in Nano-Electro-Optics V**
Nanophotonic Fabrications, Devices, Systems, and Their Theoretical Bases
By M. Ohtsu (Ed.), 2006, 122 figs., XIV, 188 pages

118 **Mid-infrared Semiconductor Optoelectronics**
By A. Krier (Ed.), 2006, 443 figs., XVIII, 751 pages

119 **Optical Interconnects**
The Silicon Approach
By L. Pavesi and G. Guillot (Eds.), 2006, 265 figs., XXII, 389 pages

120 **Relativistic Nonlinear Electrodynamics**
Interaction of Charged Particles with Strong and Super Strong Laser Fields
By H.K. Avetissian, 2006, 23 figs., XIII, 333 pages

121 **Thermal Processes Using Attosecond Laser Pulses**
When Time Matters
By M. Kozlowski and J. Marciak-Kozlowska, 2006, 46 figs., XII, 217 pages

122 **Modeling and Analysis of Transient Processes in Open Resonant Structures**
New Methods and Techniques
By Y.K. Sirenko, N.P. Yashina, and S. Ström, 2007, 110 figs., XIV, 353 pages

123 **Wavelength Filters in Fibre Optics**
By H. Venghaus (Ed.), 2006, 210 figs., XXIV, 454 pages

124 **Light Scattering by Systems of Particles**
Null-Field Method with Discrete Sources: Theory and Programs
By A. Doicu, T. Wriedt, and Y.A. Eremin, 2006, 123 figs., XIII, 324 pages

125 **Electromagnetic and Optical Pulse Propagation 1**
Spectral Representations in Temporally Dispersive Media
By K.E. Oughstun, 2007, 74 figs., XX, 456 pages

126 **QuantumWell Infrared Photodetectors**
Physics and Applications
By H. Schneider and H.C. Liu, 2007, 153 figs., XVI, 250 pages

127 **Integrated Ring Resonators**
The Compendium
By D.G. Rabus, 2007, 243 figs., XVI, 258 pages

128 **High Power Diode Lasers**
Technology and Applications
By F. Bachmann, P. Loosen, and R. Poprawe (Eds.) 2007, approx. 535 figs., VI, 546 pages

129 **Laser Ablation and its Applications**
By C.R. Phipps (Ed.) 2007, approx. 300 figs., XX, 586 pages

130 **Concentrator Photovoltaics**
By A. Luque and V. Andreev (Eds.) 2007, approx. 242 figs., XV, 298 pages

131 **Surface Plasmon Nanophotonics**
By M.L. Brongersma and P.G. Kik (Eds.) 2007, approx. X, 298 pages

132 **Ultrafast Optics V**
By S.Watanabe and K. Midorikawa (Eds.) 2007, approx. 303 figs., XII, 520 pages

133 **Frontiers in Surface Nanophotonics**
Principles and Applications
By D.L. Andrews and Z. Gaburro (Eds.) 2007, approx. 89 figs., X, 250 pages

134 **Strong Field Laser Physics**
By T. Brabec, 2007, approx. 150 figs., XV, 500 pages

135 **Optical Nonlinearities in Chalcogenide Glasses and Their Applications**
By A. Zakery and S.R. Elliott, 2007, approx. 60 figs., IX, 180 pages