HANDBOOK OF ELLIPSOMETRY

Edited by

Harland G. Tompkins
Thin Films Materials Science Consultant
Chandler, Arizona

and

Eugene A. Irene
Department of Chemistry, University of North Carolina
Chapel Hill, North Carolina

William Andrew
publishing

Springer

Copyright © 2005 by William Andrew, Inc.
No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the Publisher.

Cover art © 2005 by Brent Beckley / William Andrew, Inc.

ISBN: 0-8155-1499-9 (William Andrew, Inc.)
ISBN: 3-540-22293-6 (Springer-Verlag GmbH & Co. KG)

Library or Congress Catalog Card Number:
Library of Congress Cataloging-in-Publication Data

Handbook of ellipsometry / Harland G. Tompkins and Eugene A. Irene (eds.).
p. cm.
Includes bibliographical references and index.
ISBN 0-8155-1499-9
1. Ellipsometry. I. Tompkins, Harland G. II. Irene, Eugene A.

QC443.H26 2005
535.5′2—dc22

2004014676

Printed in the United States of America
This book is printed on acid-free paper.

10 9 8 7 6 5 4 3 2 1

Co-published by:

William Andrew Publishing
13 Eaton Avenue
Norwich, NY 13815
1-800-932-7045
www.williamandrew.com
www.knovel.com
(Orders from all locations in North and South America)

Springer-Verlag GmbH & Co. KG
Tiergartenstrasse 17
D-69129 Heidelberg
Germany
www.springeronline.com
(Orders from all locations outside North and South America)

NOTICE

To the best of our knowledge the information in this publication is accurate; however the Publisher does not assume any responsibility or liability for the accuracy or completeness of, or consequences arising from, such information. This book is intended for informational purposes only. Mention of trade names or commercial products does not constitute endorsement or recommendation for their use by the Publisher. Final determination of the suitability of any information or product for any use, and the manner of that use, is the sole responsibility of the user. Anyone intending to rely upon any recommendation of materials or procedures mentioned in this publication should be independently satisfied as to such suitability, and must meet all applicable safety and health standards.

William Andrew, Inc., 13 Eaton Avenue, Norwich, NY 13815 Tel: 607/337/5080 Fax: 607/337/5090

Contributors

Ilsin An
Department of Physics and Materials Research Institute
The Pennsylvania State University
University Park, PA 16802 USA

Hans Arwin
Laboratory of Applied Optics
Department of Physics and Measurement Technology
Linköping University
SE-58183 Linköping, Sweden

Chi Chen
Department of Physics and Materials Research Institute
The Pennsylvania State University
University Park, PA 16802 USA

Robert W. Collins
Professor and NEG Endowed Chair of Silicate and Materials Science
The University of Toledo
Department of Physics and Astronomy
Mail Stop 111
Toledo, OH 43606-3390, USA

Andre S. Ferlauto
Materials Research Laboratory
The Pennsylvania State University
University Park, PA 16802, USA

James N. Hilfiker
J. A. Woollam Co., Inc.
645 M Street, Suite 102
Lincoln, NE 68508, USA

Josef Humlíček
Institute of Condensed Matter Physics, Faculty of Science
Masaryk University Brno
Kotlárská 2, CZ-61137 Brno, Czech Republic

Eugene A. Irene
Department of Chemistry
University of North Carolina
Chapel Hill, NC 27599, USA

Gerald E. Jellison, Jr.
Solid State Division
Oak Ridge National Laboratory
Oak Ridge, TN 37830, USA

Joungchel Lee
Department of Physics and Materials Research Institute
The Pennsylvania State University
University Park, PA 16802, USA

Frank A. Modine
Solid State Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831-6030, USA

Arnulf Röseler
Institut für Spektrochemie und Angewandte Spektroskopie
D-12489 Berlin, Albert-Einstein-Strasse 9, Germany

Mathias Schubert
Fakultät für Physik und Geowissenschaften
Institut für Experimentelle Physik II
Universität Leipzig, 04103, Germany

Harland G. Tompkins
Thin Films Materials Science Consultant
Chandler, AZ, 85224, USA

Juan A. Zapien
Department of Physics and Materials Research Institute
The Pennsylvania State University
University Park, PA 16802, USA

Preface

Light has been used for thickness measurements for a long time. Scientists in the 1700 time frame such as Boyle, Hooke, and Newton, looking at thin transparent layers, observed colors that were the result of interference. Brewster[1] notes that these observations "are of extensive use and may be regarded as presenting us with a micrometer for measuring minute thicknesses of transparent bodies by their colors, when all other methods would be inapplicable." This phenomenon uses the intensity of the light and is the basis for a current metrology technique called "reflectometry," which excels when measuring reproducible films that have thicknesses that are a few thousand angstroms thick. In the early 1800s, the concept of polarized light was developed. In the late 1800s, Drude used the phase shift induced between mutually perpendicular components of polarized light to measure film thicknesses down to a few tens and ones of angstroms. When the mutually perpendicular components of polarized light are out of phase, the light is said to be "elliptically polarized"; and hence the technique that evolved from Drude's early measurements came to be called "ellipsometry." With further development of this technique, the lower limit of thickness has been reduced such that submonolayer coverage can be measured.

For random scientific measurements to evolve into an "analytical technique," there must be a "killer application." This came in the 1960s with the transition from making single transistors out of large "chunks" of silicon or germanium to integrated circuits that used planerization methods with silicon wafers and thin oxide and nitride films. A silicon dioxide film on a polished silicon wafer is the ideal sample for ellipsometry because of the large contrast in the index of refraction of the two materials. This application area is the subject for Chapter 8. A second important development that led directly to the elevation of ellipsometry to an important modern surface and thin film technique is the availability of digital computers, also in the same era of the 1960s. In fact, a modern personal computer is adequate for virtually all the typical ellipsometry calculations and simulations.

For the first three quarters of the 1900s, most ellipsometry was done with only a single wavelength. During the last quarter of the 1900s, spectroscopic ellipsometry has evolved significantly, and single-wavelength ellipsometry has now been relegated to tasks such as routine metrology and fast real-time studies on well-understood samples.

Several international conferences have been held on the subject of ellipsometry, and the technical literature on the topic is voluminous. In

1977, R. M. A. Azzam and N. M. Bashara authored a book[2] titled *Ellipsometry and Polarized Light*, and this work has been the key source to be cited in most technical writing on the subject. *Ellipsometry and Polarized Light* is a scholarly book and not particularly suited to beginners. Editor H. G. Tompkins published books in 1993 and 1999 that were intended for users who were specialists in other fields (e.g., process engineering) but who needed to use ellipsometry.

Considerable progress has been made in the field since *Ellipsometry and Polarized Light* was published, and it appears that a more current scholarly book is needed. Accordingly, a group of authors from the international scientific community were asked to make contributions to the *Handbook of Ellipsometry*. This book is divided into four sections. Chapters 1, 2, and 3 explain the theory of ellipsometry. Chapters 4 through 7 discuss instrumentation. Chapters 8 and 9 are critical reviews of some applications in the field. The last three chapters 10, 11, and 12, deal with emerging areas in ellipsometry.

Finally, the editors would like to acknowledge the enormous effort of the authors in writing the various chapters; of their students and co-workers for contributions, reviews, and other assistance; and of the employers and funding agencies that supported these efforts. The encouragement of the ellipsometry scientific community is also appreciated.

1. D. Brewster, *Treatise on Optics*, Longman, Rees, Orme, Brown, Green, and Taylor, London, (1831) p. 108.
2. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland Publishing Co., Amsterdam (1977).

H. G. Tompkins and E. A. Irene, *November 2004*

Contents

Contributors xiii

Preface xv

Part I. Theory of Ellipsometry

1. Polarized Light and Ellipsometry 3
   - 1.1 A Quick Guide to Ellipsometry 4
     - 1.1.1 Light Waves and Photons 4
     - 1.1.2 Polarization of Light 6
     - 1.1.3 Ellipsometric Configurations 9
     - 1.1.4 Null Ellipsometry 12
     - 1.1.5 Photometric Ellipsometry and Polarimetry 13
   - 1.2 Maxwell and Wave Equations 19
     - 1.2.1 Linear Local Response 20
     - 1.2.2 Linear Non–Local Response 22
     - 1.2.3 Dipole Moment, Susceptibility and Inductions 23
     - 1.2.4 Relationships between Optical Constants 24
     - 1.2.5 Wave Equation for Monochromatic Fields 26
     - 1.2.6 Plane Waves in Isotropic Medium 29
   - 1.3 Representations of Polarization 31
     - 1.3.1 Representation by Ellipsometric Angles ... 32
     - 1.3.2 Special Cases: Linear and Circular Polarization 35

This page has been reformatted by Knovel to provide easier navigation.

1.3.3 Orthogonal Polarization States 37
1.3.4 Representation by Complex Numbers 37
1.3.5 Light Intensity, Detection of Polarization State 40
1.4 Propagation of Polarized Light 45
1.4.1 Jones Vectors 45
1.4.2 Jones Matrices 48
1.4.3 Quantum Mechanical Description, Partial Polarization 53
1.4.4 Stokes Vectors 56
1.4.5 Mueller Matrices 59
1.5 Reflection and Transmission of Polarized Light at Planar Interfaces 67
1.5.1 Matching Plane Waves at a Planar Interface 67
1.5.2 Fresnel Coefficients 72
1.5.3 Special Values of the Angle of Incidence 74
1.5.4 Ratio of Amplitude Reflectivities 76
1.5.5 Propagation Matrices, Stratified Structures 80
1.5.6 Substrate–Film–Ambient System 85
1.6 References 90
2. Optical Physics of Materials 93
2.1 Introduction 93
2.2 Propagation of Light in Solids 102
2.2.1 Optically Isotropic Solids and the Complex Dielectric Function 102
2.2.2 Optically Anisotropic Solids and the Dielectric Tensor 110
2.2.3 Dispersion Relationships 124

This page has been reformatted by Knovel to provide easier navigation.

2.3 Classical Theories of the Optical Properties of Solids 125
2.3.1 Semiconductors and Insulators: the Lorentz Oscillator Model 125
2.3.2 Metals: The Drude Free Electron Model ... 129
2.3.3 Plasmons 132
2.3.4 Optical Sum Rules 136
2.4 Quantum Mechanical Theories of the Optical Properties of Solids 137
2.4.1 Quantum Theory of Absorption and Dispersion 138
2.4.2 Direct Interband Transitions in Solids 146
2.4.3 Band Structure and Critical Points in Solids 150
2.4.4 Indirect Interband Transitions in Solids 153
2.4.5 Intraband Transitions in Metals 157
2.5 Modeling the Optical Properties of Solids 159
2.5.1 Classical Lorentz Oscillator Models 159
2.5.2 Classical Drude Models 172
2.5.3 Generalized Quantum Mechanical Models 178
2.5.4 Specialized Quantum Mechanical Models 207
2.6 Overview and Concluding Remarks 227
Acknowledgments 230
2.7 References and Bibliography 230
2.7.1 Numbered References 230
2.7.2 Bibliography 233
3. Data Analysis for Spectroscopic Ellipsometry 237
3.1 Introduction 237
3.2 Ellipsometry Parameters 239
3.2.1 Calculated Parameters: Jones Matrices ... 240

This page has been reformatted by Knovel to provide easier navigation.

3.2.2 Measured Parameters: Mueller Matrices .. 241
3.2.3 Mueller-Jones Matrices 242
3.3 Calculation of Complex Reflection Coefficients 246
3.3.1 Isotropic, Homogeneous Systems 246
3.3.2 Anisotropic Systems 248
3.3.3 Inhomogeneous Layers 251
3.4 Models for Dielectric Functions 252
3.4.1 Tabulated Data Sets 253
3.4.2 Lorentz Oscillator Model 254
3.4.3 Optical Functions of Amorphous Materials 255
3.4.4 Models for Crystalline Materials 258
3.4.5 Effective Medium Theories 260
3.5 Fitting Models to Data 262
3.5.1 Figures of Merit 263
3.5.2 Errors in Spectroscopic Ellipsometry 265
3.5.3 Convergence Routines 268
3.5.4 An Example: (a-Si_xN_y:H) 271
3.6 Determination of Optical Functions from Spectroscopic Ellipsometry Data 276
3.6.1 Optical Functions from Parameterization .. 278
3.6.2 Newton-Raphson Algorithm 280
3.6.3 Optical Functions of Bulk Isotropic Semiconductors and Insulators 282
3.6.4 Optical Functions of Anisotropic Materials 285
3.6.5 Optical Functions of Thin Films 286
3.7 Depolarization 289
Acknowledgements 293
3.8 Further Reading and References 293
Optics and Ellipsometry 293

This page has been reformatted by Knovel to provide easier navigation.

Data Reduction 294
Numbered References 294

Part II. Instrumentation

4. Optical Components and the Simple PCSA (Polarizer, Compensator, Sample, Analyzer) Ellipsometer 299
   - 4.1 General 299
   - 4.2 The Components 301
     - 4.2.1 Methods of Obtaining Polarized Light 301
     - 4.2.2 Double Refraction 302
     - 4.2.3 Calcite Crystals 303
     - 4.2.4 Polarizers and Analyzers 305
     - 4.2.5 Wollaston Prisms 307
     - 4.2.6 Compensators, Quarter-Wave Plates, and Retarders 308
     - 4.2.7 Photoelastic Modulators 316
     - 4.2.8 Monochromators 317
     - 4.2.9 Goniometers 321
   - 4.3 Ellipsometer Component Configurations 322
     - 4.3.1 Early Null Ellipsometer Configurations 322
     - 4.3.2 Photometric Ellipsometer Configurations 323
     - 4.3.3 Spectroscopic Ellipsometers 324
     - 4.3.4 Other Configurations 326
   - 4.4 References 327
5. Rotating Polarizer and Analyzer Ellipsometry 329
   - 5.1 Introduction 329
   - 5.2 Comparison of Ellipsometers 333
   - 5.3 Instrumentation Issues 343
     - 5.3.1 Optical Configuration 343
     - 5.3.2 Optical Components and Spectral Range 345
     - 5.3.3 Alignment 351

This page has been reformatted by Knovel to provide easier navigation.

5.3.4 Electronic Design and Components 356
5.4 Data Reduction for the Rotating Polarizer and Analyzer Ellipsometers 364
5.4.1 Ideal $PXSA_r$ Configuration 364
5.4.2 Errors in the $PXSA_r$ Configuration 371
5.4.3 P_rXSA Configuration 378
5.5 Precision Considerations 386
5.6 Calibration Procedures 392
5.6.1 Ideal Rotating Polarizer and Analyzer Ellipsometers 394
5.6.2 Detecting and Correcting Errors in Calibration 407
5.6.3 Detecting and Correcting Compensator Errors 423
5.7 Summary: Recent and Future Directions 425
5.8 References 429
6. Polarization Modulation Ellipsometry 433
6.1 Introduction 433
6.2 The Photoelastic Modulator (PEM) 436
6.2.1 General Description and Historical Perspective 436
6.2.2 Mathematical Description of a PEM 440
6.2.3 Stokes Vector Descriptions of the PSG and PSA 442
6.3 Experimental Configurations of Polarization Modulation Ellipsometers 446
6.3.1 Polarization Modulation Ellipsometry (PME) with Analog Data Acquisition 446
6.3.2 Phase Modulated Ellipsometry (PME) with Digital Data Acquisition 447
6.3.3 Two-Channel Spectroscopic Polarization Modulation Ellipsometer (2-C SPME) 449

This page has been reformatted by Knovel to provide easier navigation.

6.3.4 Two-Modulator Generalized Ellipsometer (2-MGE) 450
6.4 Light Intensity through a Polarization Modulation Ellipsometer 452
6.4.1 Mueller Matrices for Various Samples 452
6.4.2 Intensity for a Standard PME 455
6.4.3 Intensity for the 2-Modulator Generalized Ellipsometer (2-MGE) 457
6.5 Waveform Analysis 461
6.5.1 Basis Function 463
6.5.2 Phase-Sensitive Detection 465
6.5.3 Digital Waveform Analysis 466
6.5.4 Two-Modulator Systems 467
6.6 Calibration Procedures 469
6.6.1 One-Modulator PMEs 470
6.6.2 Two-Modulator PMEs 472
6.7 Errors 474
6.7.1 General Discussion 474
6.7.2 Systematic Errors of PMEs 475
6.8 Further Reading and References 479
6.8.1 Further Reading 479
6.8.2 Numbered References 479
7. Multichannel Ellipsometry 481
7.1 Introduction 481
7.2 Overview of Instrumentation 483
7.2.1 Self-Compensating Designs 483
7.2.2 Rotating-Element Designs 487
7.2.3 Phase-Modulation Designs 491
7.2.4 Design Comparisons 493
7.2.5 Errors Unique to Multichannel Detection Systems 497
7.3 Rotating-Element Designs 502

This page has been reformatted by Knovel to provide easier navigation.

7.3.1 Rotating Polarizer 502
7.3.2 Single Rotating Compensator 523
7.3.3 Dual Rotating Compensator 546
7.4 Concluding Remarks 562
Acknowledgements 564
7.5 References 564

Part III. Critical Reviews of Some Applications

8. SiO_2 Films 569
8.1 Introduction 569
8.1.1 Preeminence of SiO_2 in Microelectronics: the Ellipsometry Connection 569
8.1.2 Electronic Passivation 570
8.1.3 Properties of SiO_2 Films 571
8.2 Historical Perspective – Prior to 1970 578
8.3 Modern Studies – Since 1970 585
8.3.1 Thick SiO_2 Films 585
8.3.2 Thin SiO_2 Films 599
8.3.3 Recent Results on Ultra Thin SiO_2 Films and the Si-SiO_2 Interface 619
8.4 Conclusions 632
Acknowledgements 633
8.5 References 633
9. Theory and Application of Generalized Ellipsometry 637
9.1 Introduction 637
9.2 The Generalized Ellipsometry Concept 638
9.2.1 Comments on Notations in GE 638
9.2.2 The Optical Jones Matrix 640
9.2.3 The Generalized Ellipsometry Parameters 643
9.2.4 Generalized Ellipsometry Acquisition Techniques 647

This page has been reformatted by Knovel to provide easier navigation.

9.3 Theory of Generalized Ellipsometry 650
9.3.1 Birefringence in Stratified Media 650
9.3.2 4 x 4 Maxwell's Equations in Matrix Form 652
9.3.3 Transmission and Reflection GE 656
9.4 Special Generalized Ellipsometry Solutions 657
9.4.1 Biaxial Films (Symmetrically Dielectric Materials) 657
9.4.2 Bi-Biaxial or Magneto-Optical Films (Non-Symmetrically Dielectric Materials) 661
9.4.3 Chiral Biaxial Films (Axially Twisted Symmetrically Dielectric Materials) 663
9.4.4 Isotropic Dielectric Films 669
9.4.5 Further Solutions: [1 1 1] Superlattice Ordering in III-V Compounds (CuPt-Ordering) 671
9.5 Strategies in Generalized Ellipsometry 675
9.5.1 Data Acquisition Strategies for Anisotropic Samples 676
9.5.2 Strategies for Treatment of Sample Backside Effects 679
9.5.3 Model Strategies 682
9.6 Generalized Ellipsometry Applications 683
9.6.1 Anisotropic Bulk Materials 684
9.6.2 Anisotropic Films 693
9.7 Conclusions 710
Acknowledgements 710
9.8 Further Reading and References 711
9.8.1 General Reading 711
9.8.2 Numbered References 712

This page has been reformatted by Knovel to provide easier navigation.

Part IV. Emerging Areas in Ellipsometry

10. VUV Ellipsometry 721
10.1 Introduction 721
10.2 Historical Review of Short Wavelength Ellipsometry 722
10.2.1 BESSY Ellipsometer 722
10.2.2 EUV Ellipsometer 724
10.3 VUV Ellipsometry Today 726
10.3.1 Current VUV Instrumentation 726
10.4 Importance of VUV Ellipsometry 732
10.5 Survey of Applications 737
10.5.1 Lithography 740
10.5.2 Gate Dielectrics 748
10.5.3 High-Energy Optical Constants 749
10.6 Future of VUV Ellipsometry 757
10.7 Acknowledgments 757
10.8 References 757
11. Spectroscopic Infrared Ellipsometry 763
11.1 Experimental Tools 763
11.1.1 Two Kinds of Instruments 763
11.1.2 Optical Equipment for the Infrared-Ellipsometry 768
11.1.3 The Degree of Polarization 771
11.1.4 Linearity of the Detection System 775
11.1.5 Infrared Synchrotron Radiation 775
11.2 Applications 776
11.2.1 Optics of Absorbing Media 776
11.2.2 Vibration Modes – the Concept of Weak and Strong Oscillators 778
11.2.3 Inversion of Infrared Ellipsometric Measurements 781

This page has been reformatted by Knovel to provide easier navigation.

11.2.4 Anisotropy Features in the Infrared Ellipsometric Spectra 786
11.3 References 797
12. Ellipsometry in Life Sciences 799
Poem and Dedication 799
12.1 Introduction 800
12.2 Historical Background 802
12.3 The Interfaces under Study 802
12.4 From Optics to Biology 804
12.4.1 The Unique Possibilities 804
12.4.2 Verification of Ellipsometric Results 805
12.5 Methodology for Data Evaluation – from Ψ and Δ to Biologically Related Parameters 806
12.5.1 A Thin Biolayer on a Flat Ideal Substrate 806
12.5.2 A Thick Biolayer on a Flat Ideal Substrate 817
12.5.3 Adsorption of Biomolecules into Porous Structures 817
12.5.4 Surface Roughness 819
12.5.5 Use of Dispersion Models 820
12.5.6 Anisotropy 820
12.6 Methodology – Experimental 821
12.6.1 Instrumentation 821
12.6.2 Cell Designs 822
12.6.3 *In situ* Considerations for Biological Interfaces 824
12.6.4 Some Model Surfaces 825
12.6.5 Studies on Real Biological Surfaces 827
12.6.6 Complementary and Independent Information 828
12.6.7 Experimental Design 828

This page has been reformatted by Knovel to provide easier navigation.

12.7 Applications 829
12.7.1 Introduction 829
12.7.2 Adsorption of Biomolecules to Model Surfaces 830
12.7.3 Spectroscopy 839
12.7.4 Imaging 841
12.7.5 Biological Surfaces 843
12.7.6 Biosensors Based on Ellipsometric Readout 844
12.7.7 Engineering Applications 845
12.8 Outlook 846
Acknowledgements 847
12.9 References 847

Index **857**

This page has been reformatted by Knovel to provide easier navigation.

Theory of Ellipsometry

1 Polarized Light and Ellipsometry

Josef Humlíček

Institute of Condensed Matter Physics, Faculty of Science
Masaryk University Brno, Czech Republic

Polarization is a fundamental property of light. It has been treated comprehensively in many texts such as Born and Wolf[1], Brosseau[2], Chipman[3], and Saleh and Teich[4]. In this chapter, we summarize the basic theoretical knowledge on the properties of polarized light on two levels of complexity.

First, we collect in Section 1.1 information needed by "plug-and-play" users of commercial equipment, who are interested mostly in some of the applications of ellipsometry. As the precision and accuracy of state-of-the-art ellipsometers increase, detection of partially polarized light becomes increasingly important; we have included a short discussion of its basics in the quick-guide section.

Second, selected topics are treated in more detail in the following sections. The phenomenological background, ranging from Maxwell equations to plane electromagnetic waves, is covered in Section 1.2. It also provides the link of field amplitudes to measurable quantities (intensities of light waves), which is essential in any ellipsometric setup. The microscopic origin of optical response is discussed by Collins and Ferlauto in Chapter 2. The properties of general elliptic polarization are described using several representations in Section 1.3. Each of the representations is frequently encountered in ellipsometric literature; we discuss their use and provide mutual links for reference. In addition, we explain principal ways of measuring the parameters of polarized light in photometric ellipsometers. Photometric setups are dealt with in more detail by Tompkins, Jellison, and Roeseler elsewhere in this volume. Section 1.4 is devoted to a systematic description of the propagation of purely and partially polarized light through optical systems. We aim at selected optical components and try to explain the ways of predicting the polarization phenomena. A number of additional items is discussed by Jellison and Modine in Chapters 3 and 6. The closing section of this chapter, 1.5, deals with the simplest isotropic surfaces and homogeneous and isotropic film systems. We provide a detailed mathematical description and several illustrations of a typical behavior of polarized light at planar interfaces. The more complex case of anisotropic materials is treated by Schubert in Chapter 9.

1.1 A Quick Guide to Ellipsometry

1.1.1 Light Waves and Photons

Light is a transverse electromagnetic wave with frequencies ranging from about 3×10^{11} to 3×10^{16} Hz. The corresponding wavelengths in vacuum are from about 1 mm to 0.01 μm. In the range of optical frequencies, as well as below it (microwaves and radio waves), matter behaves as a continuum. Above the optical range, starting with soft X-rays, the discrete atomic structure of matter becomes important even for the phenomenology of radiation–matter interaction, since it leads to diffraction.

The light wave in a medium is linked to the induced motion of the electric charges of electrons and atomic nuclei, constituting macroscopic electric currents. On the other hand, it is mostly safe to neglect the induced magnetization, since the motion of magnetic moments of electrons and nuclei is too slow to follow the rapid optical oscillations. Thus, the most important quantity describing the light wave is the vector of its electric–field intensity, $\vec{E}$. The existence of preferential directions of the action that the electric field in the wave exerts on electric charges in matter is the reason for the importance of its *polarization*. Early experiments with the optical behavior of calcite crystals led to the discovery of the polarization of light by Malus in 1808, well before the classical electromagnetic theory had been established. The spatial and temporal dependence of the electric field $\vec{E}(\vec{r},t)$ in a uniform, isotropic medium of the complex permittivity ε, is described by the wave equation[1,5]

$$\left[\nabla^2 - \frac{\varepsilon(\omega)}{c^2}\frac{\partial^2}{\partial t^2}\right]\vec{E}(\vec{r},t) = 0, \tag{1.1}$$

where ω is the angular frequency of the light wave, and c is the light velocity in vacuum. A useful solution of the wave equation is the monochromatic plane wave propagating along the z–axis of an orthogonal coordinate system,

$$\vec{E}(z,t) = \mathrm{Re}\left\{\begin{bmatrix} E_x \\ E_y \end{bmatrix} e^{i(k_z z - \omega t)}\right\}. \tag{1.2}$$

Here E_x and E_y are the complex amplitudes of $\vec{E}$ along the x– and y–axes; they can be arranged conveniently in the 2×1 column vector. The symbols $\mathrm{Re}\{f\}$ and $\mathrm{Im}\{f\}$ mean real and imaginary parts of a complex quantity f, respectively; for example, the x component of the field intensity is $E_x(z,t) = \mathrm{Re}\{E_x\}\cos(k_z z - \omega t) - \mathrm{Im}\{E_x\}\sin(k_z z - \omega t)$. We use the complex form of most of the equations, since they are simple and transparent.

However, care should be taken in adopting a number of different possible conventions[6].

In order to satisfy the wave equation, Eq. (1.1), the nonvanishing component of the propagation vector of the plane wave of Eq. (1.2) assumes the values given by the dispersion equation

$$k_z = \frac{\omega}{c}\sqrt{\varepsilon} = \frac{\omega}{c}N, \tag{1.3}$$

where $N = n + ik = \sqrt{\varepsilon}$ is the complex refractive index of the medium. By choosing the plane–wave solution, Eq. (1.2), with the time dependence of $\exp(-i\omega t)$, we are adopting the standard physics convention, where the imaginary parts of ε and N are positive. Another convention, in which the two imaginary parts are negative (usually preferred in optics) is easily recognizable, since all expressions containing ε and N become complex conjugate.

The *intensity* of a light wave is its energy crossing a unit area perpendicular to the direction of propagation per unit time. It is given by the magnitude of the time average of the Poynting vector, $\vec{E}(t) \times \vec{H}(t)$, where $\vec{H}$ is the intensity of the magnetic field. In SI units, the intensity of the plane wave of Eq. (1.2) in free space is

$$I \equiv |\langle \vec{E}(t) \times \vec{H}(t) \rangle| = \frac{c\varepsilon_0}{2}(|E_x|^2 + |E_y|^2), \tag{1.4}$$

where ε_0 is the permittivity of vacuum. The signals measured by detectors in ellipsometric setups are proportional to the intensity, i.e., to the *squared modulus* of the complex amplitude of the electric intensity. Since $c\varepsilon_0 \approx 2.654 \times 10^{-3}$ A/V, a wave with the electric field amplitude of 1 V/m has the intensity of about 1.3 mW/m^2. At low light intensities used in ellipsometric measurements, the response of materials to the light wave is linear.

The understanding of the light–matter interaction requires, as a rule, quantum mechanical description. A monochromatic wave of the angular frequency ω carries energy in the quanta of $\hbar\omega$, where $\hbar = 1.05457266 \times 10^{-34}$ Js is the Planck constant. To convert the wavelength λ to the photon energy $\hbar\omega$ in the practical units of electron–volts, we use the following relation,

$$\hbar\omega[\mathrm{eV}] = \frac{1.23985}{\lambda\,[\mu\mathrm{m}]}. \tag{1.5}$$

The photon energy ranges from about 1 meV in the far–infrared to about 100 eV in the vacuum ultraviolet. For example, the green light of a mercury discharge lamp has the vacuum wavelength of 0.5461 μm, corresponding

to the photon energy of 2.270 eV; a collimated light beam with the intensity of 1.3 mW/m^2 (mentioned above as having the electric field intensity of 1 V/m in the classical picture) consists of the flux of 3.6×10^{15} photons per second per m^2. A striking consequence of the quantized nature of the light wave is the onset of sensitivity of photoconductive or photovoltaic detectors. For example, silicon–based detectors provide signals only for photon energies above the bandgap (about 1.1 eV); a light beam of lower photon energy is not absorbed, irrespective of its intensity.

The second fingerprint of the quantum nature of a light wave (i.e., the flux of individual photons) is the superposition of possible polarization states of the photons, described by the quantum–mechanical probability amplitudes. This is formally identical with coherent superposition of the complex amplitudes of plane waves in the classical picture. The advantage of the quantum mechanical description becomes obvious especially in the treatment of unpolarized or partially polarized light.

1.1.2 Polarization of Light

The most general polarization of a monochromatic light wave, Eq. (1.2), is *elliptic*. Since the wave is transverse, the endpoint of the electric–field intensity vector precesses along an elliptic trajectory in any plane perpendicular to the direction of propagation. One revolution is achieved in the very short time interval of $2\pi/\omega$. The time evolution can be viewed as a superposition of harmonic vibrations along two perpendicular axes. If the two vibrations are shifted in phase, the resulting trajectory is elliptic. An example of the polarization ellipse is shown in Fig. 1.1. The wave is assumed to propagate along the z axis of the right–handed cartesian coordinate system x–y–z. The amplitudes of the electric field in the x and y directions are denoted by X and Y, respectively; both of them are real, non–negative quantities. The time dependence of the vector $\vec{E}(t)$ of Eq. (1.2) in the plane $z = 0$ can be written down conveniently using the following complex form,

$$\vec{E}(t) = \begin{bmatrix} E_x(t) \\ E_y(t) \end{bmatrix} = \mathrm{Re}\left\{ \begin{bmatrix} Xe^{i\Delta} \\ Y \end{bmatrix} e^{i\omega(t-t_o)} \right\}. \tag{1.6}$$

At the initial time $t = t_0$, the y component is at its maximum as indicated by the dashed arrow in Fig. 1.1; the maximum value of the x component is reached after the time interval of Δ/ω (dotted arrow). Note that the angle between the dashed and dotted vectors in Fig. 1.1 is related to the *relative phase* Δ of the vibrations along x– and y–directions. For positive values of Δ, the sense of precession is clockwise, and the polarization is

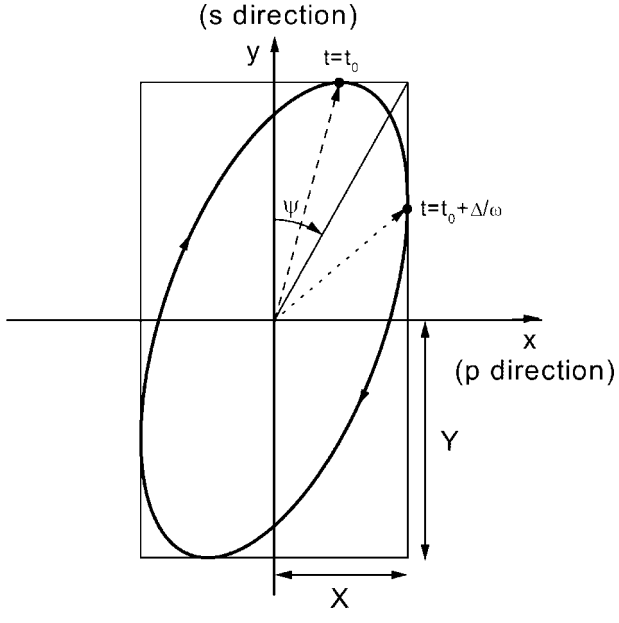


Figure 1.1 Polarization ellipse, described by the ellipsometric angles ψ and Δ. The polarized wave propagates in the positive sense of the z axis, which points towards the reader.

called right–handed. For negative values of Δ, the polarization ellipse is described counter–clockwise; it is called left–handed polarization. The values of Δ are usually limited to the interval from zero to 2π, or from $-\pi$ to π. The nomenclature regarding the sense of rotation is also not unique[6].

Besides the phase shift Δ, the state of elliptic polarization is determined by the amplitudes X and Y. More precisely, only *relative amplitude* X/Y is relevant in ellipsometric measurements; multiplying both X and Y by a common constant changes merely the light intensity. The relative amplitude can be expressed with the help of the angle ψ shown in Fig. 1.1; it is given by $\tan\psi = X/Y$, varying from zero to $\pi/2$. Consequently, the elliptic polarization of Eq. (1.6) can be represented by the *Jones vector*

$$\begin{bmatrix} \sin\psi e^{i\Delta} \\ \cos\psi \end{bmatrix}, \tag{1.7}$$

which is determined by the two real angles ψ and Δ. An important notion in dealing with polarization is that of the *orthogonality* of polarization states. Two states are called orthogonal, when their Jones vectors are

orthogonal in the usual sense of vector algebra. The polarization orthogonal to that of Eq. (1.7) therefore is

$$\begin{bmatrix} -\cos\psi e^{i\Delta} \\ \sin\psi \end{bmatrix} = \begin{bmatrix} \sin(\pi/2 - \psi) e^{i(\Delta \pm \pi)} \\ \cos(\pi/2 - \psi) \end{bmatrix}. \tag{1.8}$$

Simple special cases of a general elliptic polarization are:

- Linear polarization, for which $\Delta = 0$ or π. This is equivalent to the vibrations along x and y either in–phase, or with the opposite phase. The polarization ellipse collapses into the linear segment connecting the upper–right and lower–left corners of the rectangle in Fig. 1.1 for $\Delta = 0$, or the upper–left and lower–right corners for $\Delta = \pi$. The angle between this direction and the x axis is called azimuth; it varies between zero and π and defines uniquely the state of linear polarization. Two linear polarizations with their azimuths at the right angle are orthogonal. Linearly polarized light is produced by *linear polarizers*; these are devices that reject (either absorb or deflect)[7] the unwanted orthogonal linear polarization. The angular setting of the polarizer determines the azimuth of the resulting light polarization.
- Circular polarization, for which $\psi = \pi/4$, and $\Delta = \pi/2$ or $-\pi/2$; the light is called right– or left–circularly polarized, respectively. The x– and y–components have the same magnitude and are $\pm\pi/2$ off phase. The polarization ellipse becomes a circle, with the two possibilities of the sense of precession. Right– and left–circular polarizations are mutually orthogonal. Circular polarization can easily be produced by a linear polarizer followed by a *retarder*. A *quarter–wave* retarder shifts the phase of the light polarized linearly along its fast axis by $\pi/2$ with respect to the orthogonal linear polarization along its slow axis. If a linear polarizer is set to $\pm\pi/4$ with respect to the fast axis of a quarter–wave retarder, the two circularly polarized states result.

In the quantum–mechanical picture, the general state of elliptic polarization of Eqs. (1.6) and (1.7) is a superposition of the state of linear polarization along x with the complex amplitude $\sin\psi \exp i\Delta$, and the state of linear polarization along y with the complex amplitude $\cos\psi$. The squared moduli of the two amplitudes, $\sin^2\psi$ and $\cos^2\psi$, are the probabilities of the photon being transmitted through ideal linear polarizers oriented along x and y, respectively. At the same time, it is a superposition of any other

pair of orthogonal polarizations, e.g., the right– and left–circular ones. Such a state is called *pure*; the predictions of the behavior of a light beam composed of identical photons are probabilistic, governed by the laws of quantum mechanics. The unavoidable fluctuations become observable whenever the flux is weak enough. An additional level of uncertainty arises when the polarization state of any photon in a light beam results from random processes. The best known case is the completely random polarization of the light emitted from thermal sources; ideally, any possible pure state of polarization is found with the same probability of one half. In other words, the intensity of the *completely unpolarized light* is halved by any lossless ideal polarizer (linear, circular, or elliptic), irrespective of the polarizer settings.

A state of *partial polarization* is called *mixed* in quantum mechanics[8]. It results from a probabilistic preference of a certain pure (elliptic, in general) polarization state. This preference is most naturally quantified by the probability w_χ of finding the specified pure state labeled χ. Thus, instead of the pair (ψ,Δ) of two real quantities defining any pure state, the mixture is given by the three real numbers (ψ,Δ,w_χ). The mixed state becomes pure for $w_\chi = 1$ or $w_\chi = 0$, i.e., when an observation of the polarization leads with certainty to the specific pure state or to the pure state orthogonal to it, respectively. Another convenient measure of the preference towards a certain pure state is the *degree of polarization* P_χ, related to the probability w_χ by

$$P_\chi = 2w_\chi - 1, \; w_\chi = \frac{1 + P_\chi}{2}. \tag{1.9}$$

It is usually introduced as the measure of incoherent addition of classical waves. The intensity of a partially polarized beam is decomposed into the fraction P_χ of the totally polarized light and the fraction $1 - P_\chi$ of the totally unpolarized light[1]. Except for the laser sources, the light entering an ellipsometer is usually almost totally unpolarized. After passing the first polarizer, the degree of polarization is typically fairly high, unless some of the optical elements or the measured sample exercise a depolarizing action. The properties of partial polarization states are discussed in more detail in Section 1.4.

1.1.3 Ellipsometric Configurations

Ellipsometric technique is based on a suitable manipulation of the polarization state by auxiliary polarizing elements and measured sample. The basic PCSA configuration of an ellipsometer is shown in Fig. 1.2, consisting of a

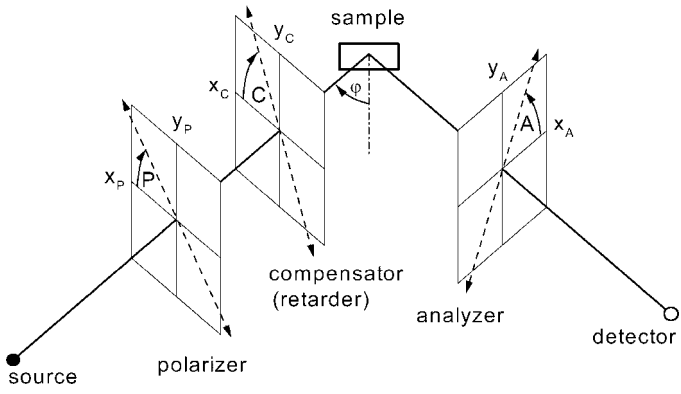


Figure 1.2 Polarizer–compensator–sample–analyzer (PCSA) configuration of an ellipsometer.

light source, linear polarizer (P), retarder (called also compensator, C), sample (S), linear polarizer (called analyzer, A), and detector. The arm with the source, polarizer and retarder prepares a known polarization state of light incident on the sample. The arm with the analyzer and detector is used to detect the change of polarization produced by the sample. An alternative PSCA scheme results from moving the compensator between the sample and analyzer.

A planar sample is assumed in Fig. 1.2, with the angle of incidence denoted by φ. The beams incident on, and reflected from, the sample, lie in the *plane of incidence*, which contains also the normal to the sample surface. The oblique incidence ($\varphi > 0$) is typical, since it leads to pronounced changes of the polarization of incident light upon reflection from the sample. The reason is a pronounced difference in the behavior of the two basis linear polarizations, parallel to the plane of incidence (*p*–polarization) and perpendicular to it (*s*–polarization; from *senkrecht*, German for perpendicular). In isotropic materials, currents induced by the light wave flow in the direction of the electric field. The currents are different when driven along the sample surface or interfaces within a layered sample for *s*–polarization, or partly perpendicular to the surface (*p*–polarization). For anisotropic samples, measurable changes of polarization state can occur even at normal incidence, $\varphi = 0$, since the induced currents need not be parallel to the electric field.

In the scheme of Fig. 1.2, the directions of linearly polarized light transmitted by the polarizer and analyzer are defined by the angles P and A, respectively. The azimuth of the fast axis of the compensator is C. All of these angles are measured from the local x–axes lying in the plane of incidence, with the positive values for the clockwise rotations when looking in

the positive z–direction (i.e., using the right–hand rule). The dependence of the field amplitude seen by the detector on the relevant parameters can be easily calculated by performing the following steps:

(i) Assume the complex amplitude E_P of the linearly polarized wave transmitted by the polarizer. Note that this amplitude will be independent of the azimuth P for an unpolarized or circularly polarized wave incident on the polarizer.

(ii) The wave incident on the compensator is a superposition of the component aligned along the fast axis, having the complex amplitude $E_P\cos(P - C)$, and the orthogonal component along the slow axis, having the amplitude $E_P\sin(P - C)$. The transmitted wave is modified by multiplying the two components by the complex transmittances t_{fast} and t_{slow}, respectively. Consequently, the field amplitudes transmitted by the compensator along its fast and slow axes are

$$E_{fast} = t_{fast}E_P\cos(P - C), \quad E_{slow} = t_{slow}E_P\sin(P - C). \tag{1.10}$$

(iii) The wave incident on the sample is a superposition of the component parallel and perpendicular to the plane of incidence, i.e., p– and s–polarizations. The corresponding complex amplitudes are $E_{fast}\cos C - E_{slow}\sin C$, and $E_{fast}\sin C + E_{slow}\cos C$; the reflected wave is modified by multiplying the two components by the complex reflectivities r_p and r_s, respectively. Consequently, the amplitudes of p– and s–polarized components reflected from the sample are

$$\begin{aligned} E_{rp} &= r_P(E_{fast}\cos C - E_{slow}\sin C), \\ E_{rs} &= r_s(E_{fast}\sin C + E_{slow}\cos C). \end{aligned} \tag{1.11}$$

(iv) Finally, the wave transmitted by the analyzer results from the addition of incident, p– and s–polarized components, projected onto the direction of the azimuth A. Substituting the intermediate amplitudes of Eqs. (1.10) and (1.11), the result is

$$\begin{aligned} E_A &= E_{rp}\cos A + E_{rs}\sin A \\ &= E_P\{r_p\cos A[t_{fast}\cos(P - C)\cos C - t_{slow}\sin(P - C)\sin C] \\ &\quad + r_s\sin A[t_{slow}\sin(P - C)\cos C + t_{fast}\cos(P - C)\sin C]\}. \end{aligned} \tag{1.12}$$

According to Eq. (1.4), the light intensity measured by the detector is proportional to $|E_A|^2$.

A simpler ellipsometric scheme is the polarizer–sample–analyzer (PSA) configuration, resulting by removing the compensator from PCSA. Note that the compensator is effectively removed from the system by aligning either of its axes parallel to the azimuth of the polarizer. The field amplitude on the detector results from Eq. (1.12) by inserting $t_{fast} = t_{slow} = 1$ as

$$\begin{aligned} E_A &= E_P(r_p \cos P \cos A + r_s \sin P \sin A) \\ &= E_P r_s(\rho \cos P \cos A + \sin P \sin A). \end{aligned} \tag{1.13}$$

We have denoted by ρ the complex reflectance ratio,

$$\rho = \frac{r_p}{r_s} = \tan\psi e^{i\Delta}, \tag{1.14}$$

which is usually expressed in terms of the angles ψ and Δ defined here.

The field amplitudes are usually calculated using Jones vectors and matrices in properly chosen coordinate systems. For example, the calculation for the PSA configuration can be summarized as follows,

$$E_A\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}1 & 0\\0 & 0\end{bmatrix}\begin{bmatrix}\cos A & \sin A\\ -\sin A & \cos A\end{bmatrix}\begin{bmatrix}r_p & 0\\0 & r_s\end{bmatrix}\begin{bmatrix}\cos P\\ \sin P\end{bmatrix}E_P. \tag{1.15}$$

Here, the column vector on the right–hand side is the normalized Jones vector of the linearly polarized light transmitted by the polarizer, in the x–y coordinate system of the source arm. This vector is multiplied from the left by three matrices. The first of them is the (diagonal) Jones matrix of the sample, consisting of the amplitude reflectivities for the p– and s–polarization. The next matrix multiplication rotates the x–y coordinate system of the detection arm by the angle $-A$. The first coordinate is then aligned with the transmission axis of the analyzer, allowing us to use the Jones matrix of the analyzer in its eigenpolarization system. The relation of the scalar field amplitudes E_A and E_P resulting from Eq. (1.15) is, of course, the same as that of Eq. (1.13). The advantages of the matrix description of the propagation of polarized light become even more obvious for imperfect (partially polarizing) optical elements. A convenient scheme uses *Stokes vectors and matrices*.

1.1.4 Null Ellipsometry

The PCSA configuration can be used to determine fairly simply the two ellipsometric parameters of the sample. The procedure consists of finding component settings for extinguishing the light at the detector. The azimuths of the polarizer, P_o, and compensator, C_o, are adjusted so as the

resulting elliptically polarized light is reflected from the sample as linearly polarized. The reflected beam is then extinguished by a suitably adjusted analyzer azimuth, A_o. Setting $E_A = 0$ in Eq. (1.12), we obtain a linear equation for the complex reflectance ratio. Its solution reads

$$\rho = \frac{\tau_c \tan(P_0 - C_0) + \tan C_0}{\tan(P_0 - C_0)\tan C_0 - 1} \tan A_0. \tag{1.16}$$

We have denoted by τ_c the complex transmittance ratio of the compensator,

$$\tau_c = \frac{t_{slow}}{t_{fast}} = \tan \psi_c e^{i\Delta_c}, \tag{1.17}$$

which can be expressed in terms of the angles ψ_c and Δ_c.

A significant simplification results for a quarter–wave ($\tau_c = i$) compensator set at the azimuth of $C = \pi/4$. Namely, the output intensity of Eq. (1.12) is extinguished for two settings P_1, A_1 and P_2, A_2 related by $P_2 = \pm P_1$, $A_2 = -A_1$. In addition, these azimuths are straightforwardly related to the measured ellipsometric angles

$$\psi = A_1, \quad \Delta = 2P_1 + \pi/2, \tag{1.18}$$

for $A_1 \geq 0$. Using this very simple scheme with a laser source, high–quality polarizing elements and mountings, high–precision and accuracy measurements are achievable even with a manual operation of the instrument. However, the operation is slow and tedious. The usual way of calibrating the instrument consists of performing multiple–zone measurements[9], which determines and corrects the unavoidable errors in the zero positions of P, C and A and the retardation angle of the compensator. Essential ingredients of this procedure are measurements performed at the second setting of the compensator at $C = -\pi/4$.

1.1.5 Photometric Ellipsometry and Polarimetry

Photometric ellipsometry is based on measurements of intensity for a number of suitably chosen settings of the optical components influencing the polarization state of light. Unlike the nulling variant, it is well suited for spectroscopic studies in broad spectral intervals, increasing tremendously the amount of information obtainable from the measurements. Advances in instrumentation transformed ellipsometry from a cumbersome laboratory tool to a widespread, efficient technique for both basic research

and applications [10,11]. Of course, a very important issue in any photometric system is the sensitivity, linearity, response time, and polarization–dependence of light detectors.

The PSA configuration provides basic possibilities for photometric measurements. The light intensities are measured for several properly chosen azimuths of the polarizer and analyzer. Since the analyzed state of polarization is independent of absolute intensities, one of them can be used as a reference for the measurement of relative values. Consequently, at least three independent intensities are required to determine the two real ellipsometric parameters of Eq. (1.14). Equation (1.13) gives, for a fixed azimuth P, the intensity transmitted by the analyzer as

$$I(A) = I(P)|r_s|^2 \cos^2 P(\tan^2\psi \cos^2 A + \tan^2 P \sin^2 A + 2\tan\psi \cos\Delta \tan P \cos A \sin A). \tag{1.19}$$

Taking the intensity for $A = \pi/2$ as the reference, the ellipsometric angle ψ is obtained from the relative intensity measured for $A = 0$ as

$$\tan\psi = |\tan P| \sqrt{I(0)/I(\pi/2)}. \tag{1.20}$$

The third intensity can be measured for $A = \pi/4$; used in Eq. (1.19), it provides the following explicit result for the ellipsometric angle Δ,

$$\cos\Delta = \mathrm{sgn}(P)\,[2I(\pi/4) - I(0) - I(\pi/2)]/\left[2\sqrt{I(0)I(\pi/2)}\right]. \tag{1.21}$$

The obvious limitation here is the impossibility of distinguishing the sign of Δ, which is confined to an interval of the length of π. The azimuth P of the polarizer is the disposable quantity in this scheme; it should be chosen properly to minimize measurement errors. Equation (1.20) suggests the favorable setting of $P \approx \psi$, for which $I(0) \approx I(\pi/2)$. In any case, the values of P close to either zero or $\pm\pi/2$ should be avoided, since they lead to a rapid loss of sensitivity. It should be emphasized that a possible polarization sensitivity of the detector has been neglected; if present, it can be measured and accounted for using a sample with known parameters, or straight–through operation with no sample.

The number of measured intensities has to be increased when the measurement should also correct the results for unknown zero positions P_o and A_o of the polarizer and analyzer, respectively. The two values are defined as the readings P_r and A_r of the angular scales for the polarization exactly parallel to the plane of incidence, i.e., for the zero actual values P and A in $P_r = P + P_o$ and $A_r = A + A_o$. Alternatively, P_o and A_o represent

minus the actual azimuths for zero readings of the two scales. Note that a tilt of the sample in Fig. 1.2 along its axis lying in the plane of incidence changes the zero positions. The three intensities of Eqs. (1.20) and (1.21) are actually equal to $I(-A_o)$, $I(-A_o + \pi/4)$ and $I(-A_o + \pi/2)$. Averaging of the results obtained for the intensities measured with the readings $P + P_o$ and $-(P + P_o)$ at the polarizer scale provides the correction to the first order in small P_o and A_o. In addition, it specifies the zero positions of both polarizer and analyzer.

The PSA scheme with the fixed polarizer is usually operated with a larger number of intensity measurements at different analyzer azimuths. The most common way is to use equidistantly spaced points covering a half–rotation (rotating–analyzer-ellipsometer, RAE)[12]. The intensity of Eq. (1.19) can be rewritten as

$$I(A) = \frac{I(P)|r_s|^2 \cos^2 P}{2(\tan^2\psi + \tan^2 P)}(1 + \alpha\cos 2A + \beta \sin 2A), \qquad (1.22)$$

where

$$\alpha = \frac{\tan^2\psi - \tan^2 P}{\tan^2\psi + \tan^2 P}, \quad \beta = \frac{2\tan P \tan\psi \cos\Delta}{\tan^2\psi + \tan^2 P}. \qquad (1.23)$$

The α and β are cosine and sine Fourier coefficients of $I(2A)$, normalized to the constant background. They result from the discrete Fourier transform of the measured intensities. Once computed, they are used in the following relations – equivalent, in principle, with Eqs. (1.20) and (1.21) –

$$\tan\psi = |\tan P|\sqrt{\frac{1+\alpha}{1-\alpha}}, \quad \cos\Delta = \mathrm{sgn}(P)\frac{\beta}{\sqrt{1-\alpha^2}}. \qquad (1.24)$$

The use of a larger set of analyzer positions reduces noise and allows the detection of spurious effects, such as nonlinearity of detection system. Note that both Fourier coefficients vanish for $P = \pm\psi$ and $\Delta = \pm\pi/2$; these are the conditions for the circularly polarized light in front of the analyzer.

The photometric data can also be collected for a fixed position of the analyzer and several azimuths of the polarizer (rotating–polarizer-ellipsometry, RPE). This scheme suppresses possible polarization sensitivity of the detector. However, the source has to be either unpolarized or circularly polarized for a simple treatment identical to the RAE case above; note that the amplitude of Eq. (1.13) is symmetric with respect to

the azimuths P and A for the field E_P independent of P. The RPE mode should be preferred to RAE for a weaker polarization of the source compared with the polarization sensitivity of the detector.

Although the PSA scheme is very efficient in many circumstances, the inclusion of a compensator in the PCSA or PSCA configurations offers substantial advantages. First, the polarization dependence of both source and detector can be suppressed by fixing P and A; the necessary number of intensity readings can be obtained by changing the azimuth and/or phase shift of the compensator. Second, the ellipsometric phase shift Δ can be measured with improved precision and accuracy, in the interval from zero to 2π. Third, partial polarization can be detected and quantified.

Assume a sample illuminated by a linearly polarized light from an ideal polarizer at the azimuth P in the PSA scheme. With no depolarization effects, the reflected light is in a pure elliptic polarization state χ showing up in the Fourier coefficients of Eqs. (1.23). A partly polarized state results when the state χ appears in the reflected beam with the probability $w_\chi < 1$, and the orthogonal pure state χ' with the complementary probability of $1 - w_\chi$. This can occur due to either imperfections of the sample, such as a spread of its complex reflectance ratio across the illuminated surface, or the imperfections of the incoming beam other than its polarization state, such as a spread of the angles of incidence or wavelengths. The total flux behind the analyzer at the azimuth A is composed of two parts,

$$\begin{aligned} \tilde{I}(A) \sim\ & w_\chi(1 + \alpha\cos 2A + \beta\sin 2A) \\ & + (1 - w_\chi)(1 + \alpha'\cos 2A + \beta'\sin 2A). \end{aligned} \tag{1.25}$$

Here α' and β' are the Fourier coefficients that would be observed if the reflected light had the polarization χ', i.e., for $w_\chi = 0$; it can be easily verified that $\alpha' = -\alpha$ and $\beta' = -\beta$. Consequently, the measured intensity

$$\begin{aligned} \tilde{I}(A) &\sim 1 + (2w_\chi - 1)\alpha\cos 2A + (2w_\chi - 1)\beta'\sin 2A \\ &= 1 + \tilde{\alpha}\cos 2A + \tilde{\beta}\sin 2A \end{aligned} \tag{1.26}$$

is of the same form as for the pure elliptic polarization in Eq. (1.22). However, the cosine and sine Fourier coefficients are reduced in magnitude by multiplying with the degree of polarization $P_\chi = 2w_\chi - 1$, see Eq. (1.9),

$$\tilde{\alpha} = P_\chi \frac{\tan^2\psi - \tan^2 P}{\tan^2\psi + \tan^2 P}, \quad \tilde{\beta} = P_\chi \frac{2\tan P \tan\psi \cos\Delta}{\tan^2\psi + \tan^2 P}. \tag{1.27}$$

Both Fourier coefficients vanish for the totally unpolarized light reflected from the sample, when $w_\chi = 1/2$, $P_\chi = 0$. A rotating ideal linear polarizer cannot distinguish between unpolarized and circularly polarized light.

The two real Fourier coefficients obtained from the intensities measured in the PSA configuration do not allow obtaining the three real parameters of a partial polarization. Using a compensator in the PCSA or PSCA configuration can solve the problem. A simple possibility consists of adding the measurements of the Fourier coefficients with rotating analyzer for the axis of the compensator parallel or perpendicular to the plane of incidence. Using the notation of Eqs. (1.14) and (1.17), the intensities calculated above are modified by simply taking the product

$$\rho\tau_c = \tan\psi \tan\tau_c \, e^{i(\Delta+\Delta_c)} \tag{1.28}$$

in place of the complex reflectance ratio ρ of the sample. Accordingly, the Fourier coefficients

$$\tilde{\alpha_c} = P_\chi \frac{\tan^2\psi \tan^2\psi_c - \tan^2 P}{\tan^2\psi \tan^2\psi_c + \tan^2 P}, \quad \tilde{\beta}_c = P_\chi \frac{2\tan P \tan\psi \tan^2\psi_c \cos(\Delta + \Delta_c)}{\tan^2\psi \tan^2\psi_c + \tan^2 P} \tag{1.29}$$

are measured with the properly aligned compensator, in addition to those of Eqs. (1.27). With given values of the relative attenuation of the compensator, $\tan\psi_c$, and its phase shift, Δ_c, Eqs. (1.27) and (1.29) can be solved for the parameters of the sample, ψ and Δ, and the degree of polarization, P_χ. Techniques used to measure partially polarized light are sometimes called *polarimetry*.

The desired quantities are very simply extracted for an ideal quarter-wave retarder ($\tan\psi_c = 1$, $\Delta_c = -\pi/2$). Since Eqs. (1.27) and Eqs. (1.29) simplify as

$$\tilde{\alpha} = \tilde{\alpha_c} = P_\chi \frac{\tan^2\psi - \tan^2 P}{\tan^2\psi + \tan^2 P}, \quad \tilde{\beta} = P_\chi \frac{2\tan P \tan\psi \cos\Delta}{\tan^2\psi + \tan^2 P}, \quad \tilde{\beta}_c = P_\chi \frac{2\tan P \tan\psi \sin\Delta}{\tan^2\psi + \tan^2 P}, \tag{1.30}$$

we obtain the following explicit results for the sought quantities:

$$\cos\Delta = \mathrm{sgn}(P) \frac{\tilde{\beta}}{\sqrt{\tilde{\beta}^2 + \tilde{\beta}_c^2}}, \quad \sin\Delta = \mathrm{sgn}(P) \frac{\tilde{\beta}_c}{\sqrt{\tilde{\beta}^2 + \tilde{\beta}_c^2}}, \quad \tan\Delta = \frac{\tilde{\beta}_c}{\tilde{\beta}}, \tag{1.31}$$

$$\tan\psi = |\tan P| \frac{\tilde{\alpha} + \sqrt{\tilde{\alpha}^2 + \tilde{\beta}^2 + \tilde{\beta}_c^2}}{\sqrt{\tilde{\beta}^2 + \tilde{\beta}_c^2}}, \tag{1.32}$$

$$P_\chi = \frac{\tilde{\alpha}^2 + \tilde{\beta}^2 + \tilde{\beta}_c^2 + \tilde{\alpha}\sqrt{\tilde{\alpha}^2 + \tilde{\beta}^2 + \tilde{\beta}_c^2}}{\tilde{\alpha}\sqrt{\tilde{\alpha}^2 + \tilde{\beta}^2 + \tilde{\beta}_c^2}}. \tag{1.33}$$

Note that Eqs. (1.31) determine the Δ angle in the zero to 2π range, and improve the precision for the values close to zero or π. It should be emphasized that the degree of polarization depends on both the sample (which provides the incoherent sum of polarization states upon reflection), and the polarizer azimuth P. For $P = 0$ or $\pi/2$ the reflected light is pure, linearly polarized p or s wave, respectively, for any isotropic sample; the sine Fourier coefficients vanish, $\tilde{\beta} = \tilde{\beta}_C = 0$; the cosine coefficients are unity, $\tilde{\alpha} = \tilde{\alpha}_C = 1$, resulting in the unit degree of polarization, $P_\chi = 1$. Of course, these two eigendirections are unusable for measuring the sample properties. At the preferable setting of $|P| = \psi$, we have $\tilde{\alpha} = \tilde{\alpha}_C = 0$, leading to

$$P_\chi = \frac{\tilde{\beta}^2 + \tilde{\beta}_c^2}{\sqrt{\tilde{\beta}^2 + \tilde{\beta}_c^2}}. \tag{1.34}$$

In this case, we can easily follow the consequences of limiting the measurement to the part without the quarter–wave retarder, and assuming no depolarization:

- the reported value of ψ obtained from Eq. (1.24) is correct, $\psi = |P|$, since the cosine Fourier coefficient α is zero;
- the reported value Δ_r obtained from Eq. (1.24) is in error, since the sine Fourier coefficient β is modified by the degree of polarization in the second of Eqs. (1.30); its relation to the true value of Δ is given by

$$\cos\Delta_r = P_\chi \cos\Delta; \tag{1.35}$$

  the decrease of P_χ from unity to zero is misinterpreted as the tendency towards circular polarization, $\Delta \to \pm\pi/2$.

A finite spectral resolution and/or an imperfectly collimated beam can produce depolarization effects for perfect samples, since their response to the polarization state of the light beam depends on the photon energy and angle of incidence. On the other hand, a nonuniformity in composition across the illuminated sample surface is a potential source of depolarization effects, since it can lead to the incoherent addition of reflected

waves for an incident ideal monochromatic plane wave. Another obvious source of depolarization is a spread of film thicknesses in layered samples. As the precision of the ellipsometric technique increases, it becomes necessary to treat properly the possible partial polarization (see, e.g., the recent work[13]). We deal with partially polarized light in more detail in Section 1.4.

1.2 Maxwell and Wave Equations

Phenomenological description of the optical behavior of any material involves a few characteristics, called *response functions*. They are introduced conveniently within the framework of Maxwell equations and *material relations*; the latter are sometimes called *constitutive relations*.

The macroscopic electromagnetic field is described by the intensities $\vec{E}$ and $\vec{H}$ of the electric and magnetic components, respectively. In a non–magnetic medium, they obey Maxwell's equations (in SI units)

$$\vec{\nabla} \times \vec{H} = \varepsilon_0 \frac{\partial \vec{E}}{\partial t} + \vec{j}, \vec{\nabla} \times \vec{E} = -\mu_0 \frac{\partial \vec{H}}{\partial t}, \tag{1.36}$$

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho}{\varepsilon_0}, \quad \vec{\nabla} \cdot \vec{H} = 0, \tag{1.37}$$

satisfied at any position $\vec{r}$ and time t. Here ρ and $\vec{j}$ are the macroscopic charge and current densities; they are usually divided into *external* and *induced* parts, $\rho = \rho_{ext} + \rho_{ind}$, $\vec{j} = \vec{j}_{ext} + \vec{j}_{ind}$. In Eqs. (1.36) and (1.37), ε_0 is permittivity of vacuum, μ_0 is permeability of vacuum:

$$\varepsilon_0 = \frac{10^7}{4\pi c^2}\,\mathrm{Fm}^{-1}, \quad \mu_0 = 4\pi 10^{-7}\,\mathrm{Hm}^{-1}, \quad \varepsilon_0\mu_0 = \frac{1}{c^2}, \tag{1.38}$$

where c is the velocity of light in vacuum.

The macroscopic picture involves mean values of quantities varying rather strongly on the atomic scale. For example, we are dealing with the spatial average of the microscopic electric field intensity, $\vec{e}$, over some macroscopic volume V:

$$\vec{E}(\vec{r},t) = \frac{1}{V}\int_V \vec{e}(\vec{r}\,',t)\, d\vec{r}\,'. \tag{1.39}$$

V is centered around $\vec{r}$; it has to be large enough to smooth out the microscopic variations of $\vec{e}$, but small compared to the wavelength of light. For crystals, the volume for averaging is just one unit cell.

1.2.1 Linear Local Response

We assume no external charges and currents, i.e., the densities in Eqs. (1.36) and (1.37) are due to the electromagnetic field itself. Moreover, the macroscopic average of the charge density ρ_{ind} vanishes in transverse fields. Consequently, the only relevant material property is that relating the current density, $\vec{j} = \vec{j}_{ind}$, to the field intensities. We shall assume a linear dependence, which is valid for weak fields. This is a direct generalization of the simplest material equations for quasistationary fields, which are usually written down separately for conductors and insulators, respectively:

$$\vec{j}(\vec{r},t) = \hat{\sigma}_0 \vec{E}(\vec{r},t), \tag{1.40}$$

$$\vec{j}(\vec{r},t) = (\hat{\varepsilon}_0 - 1)\,\varepsilon_0 \frac{\partial \vec{E}(\vec{r},t)}{\partial t}. \tag{1.41}$$

In the case of a conductor containing free charge carriers, $\hat{\sigma}_0$ in Eq. (1.40) is the real tensor of static (i.e., *"dc"*) conductivity. The case of an insulator with no free charges is slightly more complex. The induced current density of Eq. (1.41) is due to the motion of bound charges, vanishing in the *dc* limit. This *dielectric displacement* is proportional to the rate of temporal change of $\vec{E}$; $\hat{\varepsilon}_0$ in Eq. (1.41) is the real tensor of (dimensionless) static permittivity. The bound charges in a conductor also contribute to the induced current; it is impossible to distinguish it from the contribution of free carriers on the macroscopic level at optical frequencies.

The needed generalization of the material equations (1.40) and (1.41) retains their linearity. Moreover, for most substances, their local character can also be preserved. The substantial change, however, consists in replacing the synchronous response by the weaker requirement of causality – the induced current depends on the electric field only *before* the time of observation. The linear and local material equation generalizing Eq. (1.40) is therefore

$$\vec{j}(\vec{r},t) = \int_{-\infty}^{t} \hat{f}(t-t')\,\vec{E}(\vec{r},t')dt' = \int_{0}^{\infty} \hat{f}(t')\,\vec{E}(\vec{r},t-t')dt'. \tag{1.42}$$

Consequently, the optical response of a homogeneous material is described by a real tensor function $\hat{f}(t)$ with the dimension of conductivity/time, defined for $t > 0$. It represents the current density produced by a δ–function pulse of the electric field. The usual approximation of optics is the *monochromatic* field of circular frequency ω,

$$\vec{E}_\omega(\vec{r},t) = \vec{E}_\omega(\vec{r})\exp(-i\omega t). \tag{1.43}$$

Inserting it into Eq. (1.42), we obtain the sought generalizations of Eqs. (1.40) and (1.41) for conductor and insulator, respectively:

$$\vec{j}_\omega(\vec{r},t) = \left[\int_0^\infty \hat{f}(t')\,\exp(i\omega t')dt'\right]\vec{E}_\omega(\vec{r},t), \tag{1.44}$$

$$\vec{j}_\omega(\vec{r},t) = \left[\frac{i}{\omega\varepsilon_o}\int_0^\infty \hat{f}(t')\,\exp(i\omega t')dt'\right]\varepsilon_o\,\frac{\partial\vec{E}_\omega(\vec{r},t)}{\partial t}. \tag{1.45}$$

For the monochromatic field, these material equations are of the same form as Eqs. (1.40) and (1.41), with the complex, frequency–dependent conductivity tensor

$$\hat{\sigma}(\omega) = \int_0^\infty \hat{f}(t')\,\exp(i\omega t')dt', \tag{1.46}$$

and the complex, frequency–dependent tensor of permittivity (usually called *dielectric function*, or sometimes also *dielectric constant*)

$$\hat{\varepsilon}(\omega) = 1 + \frac{i}{\omega\varepsilon_o}\int_0^\infty \hat{f}(t')\,\exp(i\omega t')dt' = 1 + \frac{i}{\omega\varepsilon_o}\hat{\sigma}(\omega). \tag{1.47}$$

The dependence of the conductivity and permittivity on frequency is called *dispersion*.

The real and imaginary parts of $\hat{\sigma}(\omega)$ are the cosine and sine Fourier transforms of the real tensor function $\hat{f}(t)$, respectively; the real and imaginary parts of $\hat{\varepsilon}(\omega)$ are very simply related to them. The following symmetry property

$$\hat{\varepsilon}(-\omega) = \hat{\varepsilon}^*(\omega), \quad \hat{\sigma}(-\omega) = \hat{\sigma}^*(\omega), \tag{1.48}$$

where the asterisk denotes the complex conjugation, follows directly from the definitions of Eqs. (1.46) and (1.47). We can use interchangeably each of $\sigma(\omega)$ and $\varepsilon(\omega)$ to characterize the linear response. They differ, however, rather substantially in the behavior at low frequencies. For a conductor, the complex conductivity reaches the real *dc* value of σ_0 at $\omega = 0$, while the complex permittivity diverges as $i\sigma_0/(\omega\varepsilon_0)$ for $\omega \to 0$. For an insulator, the complex permittivity reaches the real static limit of $\hat{\varepsilon}_0$ at $\omega = 0$, while the complex conductivity goes to zero as $-i\omega\varepsilon_0(\hat{\varepsilon}_0 - 1)$ for $\omega \to 0$. Thus, the static dielectric behavior is described by the slope of the frequency dependence of $\sigma(\omega)$.

1.2.2 Linear Non–Local Response

In a further generalization of the starting material equations (1.40) and (1.41), the assumption of the induced current density, $\vec{j}(\vec{r})$, to be proportional to the field strength at the same point $\vec{r}$ is relaxed. This case should occur when the average path traveled by the charge carriers is not negligible with respect to the spatial variations of the optical field. In addition to the harmonic time–dependence of Eq. (1.43), we assume also the harmonic *spatial* distribution of the electric field intensity:

$$\vec{E}_{\omega\vec{k}}(\vec{r},t) = \vec{E}_{\omega\vec{k},0} \exp(-i\omega t - i\vec{k} \cdot \vec{r}). \tag{1.49}$$

For a real wavevector $\vec{k}$ of the magnitude k, the field pattern is repeated at the distance of the wavelength, $\lambda = 2\pi/k$. When the average path of the charge carriers is not small compared with λ, the induced current is influenced by the fields at different locations. In analogy with Eq. (1.42), the linear response becomes

$$\vec{j}(\vec{r},t) = \int dt' \int d\vec{r}'\, \hat{\phi}(t', \vec{r}')\vec{E}(\vec{r} - \vec{r}', t - t'). \tag{1.50}$$

The space–time extent of the integration has to warrant both causality and finite speed of the electromagnetic interaction. When inserting the material equation (1.50) into Eq. (1.49) we obtain

$$\vec{j}_{\omega\vec{k}}(\vec{r},t) = \left[\int dt' \int d\vec{r}'\hat{\phi}(t',\vec{r}') \exp(i\omega t' + i\vec{k} \cdot \vec{r}')\right] \vec{E}_{\omega\vec{k}}(\vec{r},t). \tag{1.51}$$

Thus, we arrive at the ω– and $\vec{k}$–dependent complex tensor of conductivity

$$\hat{\sigma}(\omega,\vec{k}) = \int dt' \int d\vec{r}'\hat{\phi}(t',\vec{r}') \exp(i\omega t' + i\vec{k} \cdot \vec{r}'). \tag{1.52}$$

By a similar reasoning as in the previous subsection, we obtain also the ω– and $\vec{k}$–dependent complex tensor of permittivity, $\hat{\varepsilon}(\omega,\vec{k})$. The dependence of conductivity and permittivity on the wavevector is called *spatial dispersion*. The local response is recovered when the function $\hat{\phi}(t',\vec{r}')$ in Eq. (1.50) becomes factorized and its spatial dependence is δ–like, $\hat{\phi}(t',\vec{r}') = \hat{f}(t - t') \cdot \delta(\vec{r} - \vec{r}')$. This is equivalent to taking the $\vec{k} = 0$ values of σ and ε, which is also called *long–wavelength* limit. The non–negligible spatial dispersion occurs rather rarely; the usual way of reporting response function includes

only their frequency dependence. However, the spectacular phenomenon of optical activity (i.e., of rotating the direction of polarization of a linearly polarized wave traversing the material) can be linked to the non–local nature of the optical response[17].

1.2.3 Dipole Moment, Susceptibility and Inductions

Macroscopic *polarization* $\vec{P}$ is the spatial average of the induced dipole moment per unit volume. The linear response of Eq. (1.42) or (1.50) implies a linear dependence of $\vec{P}$ on the macroscopic field strength; with the harmonic temporal and spatial dependence $\exp(-i\omega t - i\vec{k}\cdot\vec{r})$ of $\vec{P}$ and $\vec{E}$, this relation is

$$\vec{P}_{\omega\vec{k}} = \hat{\chi}(\omega,\vec{k})\,\varepsilon_0\vec{E}_{\omega\vec{k}}. \tag{1.53}$$

The dimensionless tensor function $\hat{\chi}(\omega,\vec{k})$ is called electric susceptibility. The electric induction, $\vec{D}$, is also a linear function of the intensity $\vec{E}$,

$$\vec{D} \equiv \varepsilon_0\vec{E} + \vec{P} = (1+\hat{\chi})\varepsilon_0\vec{E} = \hat{\varepsilon}\varepsilon_0\vec{E}, \tag{1.54}$$

where $\hat{\varepsilon} = 1 + \hat{\chi}$ is the (relative) permittivity. The polarization and induced current are proportional in monochromatic fields,

$$\vec{P}_{\omega\vec{k}} = i\vec{j}_{\omega\vec{k}}/\omega, \quad \vec{j}_{\omega\vec{k}} = -i\omega\vec{P}_{\omega\vec{k}}. \tag{1.55}$$

Macroscopic magnetization $\vec{M}$ is the spatial average of the induced magnetic moment per unit volume. The assumption of nonmagnetic materials means vanishing $\vec{M}$, and magnetic induction preserves its vacuum value, $\vec{B} \equiv \mu_0\vec{H} + \vec{M} = \mu_0\vec{H}$.

For monochromatic transverse fields in the long–wavelength limit, Maxwell equations (1.36) and (1.37) and the material equation (1.47) combine into the following set

$$\vec{\nabla}\times\vec{B}_\omega(\vec{r}) = -i\omega\frac{\hat{\varepsilon}(\omega)}{c^2}\vec{E}_\omega(\vec{r}),\ \vec{\nabla}\times\vec{E}_\omega(\vec{r}) = i\omega\vec{B}_\omega(\vec{r}), \tag{1.56}$$

$$\vec{\nabla}\cdot\vec{E}_\omega(\vec{r}) = 0, \quad \vec{\nabla}\cdot\vec{B}_\omega(\vec{r}) = 0, \tag{1.57}$$

providing the spatial dependence of the field quantities in a homogeneous material of the dielectric function $\hat{\varepsilon}$.

1.2.4 Relationships Between Optical Constants

It is customary to introduce the *complex refractive index N* as the square root of the (scalar) dielectric function, i.e., $\varepsilon = N^2$. The relationships between optical constants valid at any frequency ω are listed in Table 1-1, using the most common notation for their real (σ_1, ε_1, n) and imaginary (σ_2, ε_2, k) parts. We have added the line with negative inverse of the dielectric function; its imaginary part is called *loss function*, since it is related to the energy loss of a longitudinal wave in the material. It can be directly measured using the electron energy loss spectroscopy (EELS), and compared with the data obtained from optical measurements.

Causality of the linear material equations implies the validity of a number of so called Kramers–Kronig relations. They can be derived rigorously from the analytic behavior of $\sigma(\omega)$ and its functions, like $\varepsilon(\omega)$, $N(\omega)$, and others, in the upper half–plane of the *complex* values of Ω. Using $\Omega = \omega + i\gamma$, $\gamma > 0$, in Eq. (1.46), the complex conductivity becomes

$$\sigma(\Omega) = \sigma(\omega,\gamma) = \int_0^\infty f(t') \exp(i\omega t') \exp(-\gamma t') dt'. \qquad (1.58)$$

This complex–valued function has obviously no poles in the upper half–plane of Ω for any "physically reasonable" response function f, being

Table 1-1. Relationship between Optical Constants.

Optical Constant (symbol)	**Real part**	**Imaginary part**
conductivity ($\sigma = \sigma_1 + i\sigma_2$)	$\sigma_1 = \omega\varepsilon_0\varepsilon_2$	$\sigma_2 = -\omega\varepsilon_0(\varepsilon_1 - 1)$
dielectric function ($\varepsilon = \varepsilon_1 + i\varepsilon_2$)	$\varepsilon_1 = 1 - \sigma_2/(\omega\varepsilon_0)$	$\varepsilon_2 = \sigma_1/(\omega\varepsilon_0)$
	$\varepsilon_1 = n^2 - k^2$	$\varepsilon_2 = 2nk$
refractive index ($N = n + ik$)	$n = \sqrt{(\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2})/2}$	$k = \sqrt{(-\varepsilon_1 + \sqrt{\varepsilon_1^2 + \varepsilon_2^2})/2}$
	$n = \varepsilon_2/(2k)$	$k = \varepsilon_2/(2n)$
negative inverse of dielectric function ($-\varepsilon^{-1}$)	$-\varepsilon_1/(\varepsilon_1^2 + \varepsilon_2^2)$	$\varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2)$

limited and continuous including its derivative. Moreover, $\sigma(\Omega)/(\Omega - \omega)$ decays fast enough for $|\Omega| \to \infty$ to ensure the vanishing value of the contour integral over the semicircle at infinity. This is essentially due to the dominance of inertial effects at high frequencies over the restoring and/or dissipative forces[15]. Consequently, the use of the Cauchy integral formula for a suitable closed contour in the complex Ω plane[16] leads to the following integral relations linking the real and imaginary parts of the complex conductivity:

$$\sigma_1(\omega) = \frac{1}{\pi} \fint_{-\infty}^{\infty} \frac{\sigma_2(\Omega)}{\Omega - \omega} \mathrm{d}\Omega = \frac{2}{\pi} \fint_0^{\infty} \frac{\Omega \sigma_2(\Omega)}{\Omega^2 - \omega^2} \mathrm{d}\Omega, \tag{1.59}$$

$$\sigma_2(\omega) = \frac{-1}{\pi} \fint_{-\infty}^{\infty} \frac{\sigma_1(\Omega)}{\Omega - \omega} \mathrm{d}\Omega = \frac{-2\omega}{\pi} \fint_0^{\infty} \frac{\sigma_1(\Omega)}{\Omega^2 - \omega^2} \mathrm{d}\Omega, \tag{1.60}$$

Since the integrands diverge for $\omega = \Omega$, the Cauchy principal value has to be taken. This is indicated by the integral sign with a line through it.

The equivalent integral relation for the dielectric function is

$$\varepsilon_1(\omega) - 1 = \frac{2}{\pi} \fint_0^{\infty} \frac{\Omega \varepsilon_2(\Omega)}{\Omega^2 - \omega^2} \mathrm{d}\Omega, \quad \varepsilon_2(\omega) = \frac{\sigma_0}{\omega \varepsilon_0} - \frac{2\omega}{\pi} \fint_0^{\infty} \frac{\varepsilon_1(\Omega) - 1}{\Omega^2 - \omega^2} \mathrm{d}\Omega. \tag{1.61}$$

Here σ_0 is the dc conductivity, which vanishes for insulators. For the refractive index, the Kramers–Kronig relations become

$$n(\omega) - 1 = \frac{2}{\pi} \fint_0^{\infty} \frac{\Omega k(\Omega)}{\Omega^2 - \omega^2} \mathrm{d}\Omega, \; k(\omega) = \frac{-2\omega}{\pi} \fint_0^{\infty} \frac{n(\Omega) - 1}{\Omega^2 - \omega^2} \mathrm{d}\Omega. \tag{1.62}$$

Note that there is no pole for metals at the origin $\omega = 0$ in the refractive index. Instead, the low–frequency behavior is described by a square–root singularity

$$N(\omega) \approx \sqrt{\frac{i\sigma_o}{\omega \varepsilon_o}} = (1 + i) \sqrt{\frac{\sigma_o}{2\varepsilon_o}} \frac{1}{\sqrt{\omega}}. \tag{1.63}$$

The Kramers–Kronig relations are linked with a number of *sum rules*. Using the first of Eqs. (1.61) for $\omega \to \infty$ gives the asymptotic behavior

$$\lim_{\omega \to \infty} \varepsilon_1(\omega) = 1 - \frac{\omega_p^2}{\omega^2} + \cdots, \tag{1.64}$$

where

$$\omega_p^2 = \frac{2}{\pi}\int_0^\infty \Omega\,\varepsilon_2(\Omega)d\Omega = \frac{2}{\pi\varepsilon_o}\int_0^\infty \sigma_1(\Omega)d\Omega. \tag{1.65}$$

Here ω_p denotes the plasma frequency $\sqrt{4\pi Ne^2/m}$, where N is the total electron density, e and m are electron charge and mass, respectively (in CGS units). Equation (1.65) is the well–known f sum rule. A necessary condition for the integrals in Eqs. (1.65) to exist is that ε_2 and σ_1 fall off faster than ω^{-2} and ω^{-1}, respectively. Consequently, the high–frequency limit of the second of Eqs. (1.61) gives the asymptotic behavior

$$\lim_{\omega\to\infty}\varepsilon_2(\omega) = \frac{\sigma_0}{\omega\varepsilon_0} - \frac{2}{\pi\omega}\int_0^\infty [\varepsilon_1(\Omega) - 1]d\Omega + \cdots, \tag{1.66}$$

where

$$\sigma_0 = \frac{2\varepsilon_0}{\pi}\int_0^\infty [\varepsilon_1(\Omega) - 1]d\Omega, \tag{1.67}$$

since the term proportional to ω^{-1} has to vanish. For any insulator (i.e., for $\sigma_0 = 0$), the deviation of ε_1 from its vacuum value of 1 averaged over all frequencies is zero. This is a sum rule concerning dispersive processes, accompanying the f sum rule of Eq. (1.65) for absorptive processes. Equation (1.67) can also be taken to be a direct consequence of the Kramers–Kronig relation for σ_1, Eq. (1.59), since $\sigma_0 = \sigma_1(0)$ and $\sigma_2(\Omega)/\Omega = \varepsilon_0[\varepsilon_1(\Omega) - 1]$. The basic sum rules involving the complex refractive index are

$$\omega_p^2 = \frac{4}{\pi}\int_0^\infty \Omega\,k(\Omega)d\Omega, \int_0^\infty [n(\Omega) - 1]d\Omega = 0, \tag{1.68}$$

as obtained from the high–frequency expression $N(\omega) = 1 - \omega_p^2/(2\omega^2) + \cdots$.

1.2.5 Wave Equation for Monochromatic Fields

For a monochromatic optical field of angular frequency ω, the spatial distribution of the intensities, $\vec{E}$ and $\vec{H}$, can be found from the *wave equation*, supplemented by boundary conditions. Maxwell equations (1.56) and (1.57), valid in the long–wavelength limit, can be transformed into a single differential equation of the second order in the following way. Applying the

curl operation to the second of Eqs. (1.56) we find

$$\vec{\nabla} \times [\vec{\nabla} \times \vec{E}_\omega(\vec{r})] = \vec{\nabla} \cdot [\vec{\nabla} \cdot \vec{E}_\omega(\vec{r})] - (\vec{\nabla} \cdot \vec{\nabla})\vec{E}_\omega(\vec{r}) = i\omega\vec{\nabla} \times \vec{B}_\omega(\vec{r}). \tag{1.69}$$

Since the divergence of $\vec{E}_\omega$ vanishes, and the right–hand side of Eq. (1.69) is given by the first of Eqs. (1.56), we have the wave equation for the electric field intensity:

$$\left[\Delta + \frac{\omega^2}{c^2}\hat{\varepsilon}(\omega)\right]\vec{E}_\omega(r) = 0. \tag{1.70}$$

Here Δ is the Laplace operator, $\Delta = \vec{\nabla} \cdot \vec{\nabla} = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$.

Elementary solutions of the wave equations can be sought in the form of the plane waves

$$\vec{E}_\omega(\vec{r}) = \vec{E}_0 \exp(i\vec{k} \cdot \vec{r}) = \vec{E}_0 \exp[i(k_x x + k_y y + k_z z)], \tag{1.71}$$

and a general solution as a superposition of these plane waves. The amplitude $\vec{E}_0$ has cartesian components E_{0x}, E_{0y}, and E_{0z}. The wave equation (1.70) transforms into a set of three homogeneous linear equations for the three complex components of the amplitude,

$$\left[-\vec{k} \cdot \vec{k} + \frac{\omega^2}{c^2}\hat{\varepsilon}(\omega)\right]\begin{bmatrix} E_{0x} \\ E_{0y} \\ E_{0z} \end{bmatrix} = 0. \tag{1.72}$$

In terms of the cartesian components of the tensor $\hat{\varepsilon}$, the wave equation can be written as

$$\begin{bmatrix} \varepsilon_{xx} - D & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} - D & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} - D \end{bmatrix}\begin{bmatrix} E_{0x} \\ E_{0y} \\ E_{0z} \end{bmatrix} = 0, \tag{1.73}$$

where

$$D = \frac{c^2}{\omega^2}(k_x^2 + k_y^2 + k_z^2). \tag{1.74}$$

A symmetric tensor $\hat{\varepsilon}$ is assumed; we do not include possible optical activity of the material, which can be described phenomenologically by introducing an antisymmetric imaginary part of the matrix $\hat{\varepsilon}$[17]. The symmetry of the dielectric tensor of a crystal with no optical activity results

from energy–flow considerations[18]. A non–zero solution of Eqs. (1.73) exists if the determinant of the matrix of the system vanishes.

Consider, without loss of generality, a wave of Eq. (1.71) propagating in the z direction, i.e., $k_x = k_y = 0$. Since $\vec{k} \cdot \vec{E}_0 = 0$, the wave is polarized in the x–y plane, $E_{0z} = 0$. The wave equation (1.73) reduces to the following pair of homogeneous linear equations for the non–zero amplitudes,

$$\begin{bmatrix} \varepsilon_{xx} - \dfrac{c^2}{\omega^2}k_z^2 & \varepsilon_{xy} \\ \varepsilon_{xy} & \varepsilon_{yy} - \dfrac{c^2}{\omega^2}k_z^2 \end{bmatrix} \begin{bmatrix} E_{0x} \\ E_{0y} \end{bmatrix} = 0. \tag{1.75}$$

Non–zero solutions exist for the vanishing determinant of its matrix, which is a quadratic equation for k_z^2 with the roots

$$k_{z(1,2)}^2 = \frac{\omega^2}{c^2} \frac{\varepsilon_{xx} + \varepsilon_{yy} \pm \sqrt{(\varepsilon_{xx} - \varepsilon_{yy})^2 + 4\varepsilon_{xy}^2}}{2}. \tag{1.76}$$

Thus, two linearly polarized waves can propagate along an arbitrary direction, with the magnitude of their wavevectors given by the last equation. This leads to the phenomenon of *birefringence* in anisotropic materials. The directions of the linear polarizations can easily be found from Eq. (1.75).

When the two roots of Eq. (1.76) coincide, the wave equation (1.75) is satisfied by any amplitudes E_{0x} and E_{0y}. Consequently, waves of any polarization can propagate along the z–direction, which is called *optical axis*. The multiple root of (1.76) occurs for $\varepsilon_{xx} = \varepsilon_{yy}$, $\varepsilon_{xy} = 0$ in our coordinate system. This has a simple geometrical interpretation in terms of the ellipsoid associated with the tensor $\hat{\varepsilon}$,

$$\begin{aligned} [x \quad y \quad z] \begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \varepsilon_{xx}x^2 + \varepsilon_{yy}y^2 + \varepsilon_{zz}z^2 + 2\varepsilon_{xy}xy \\ &\quad + 2\varepsilon_{xz}xz + 2\varepsilon_{yz}yz = 1. \end{aligned} \tag{1.77}$$

Namely, its cross–section perpendicular to z is circular for $\varepsilon_{xx} = \varepsilon_{yy}$, $\varepsilon_{xy} = 0$. A light wave polarized in the x–y plane "sees" the same dielectric function irrespective of the direction of its vibrations.

In general, either two, one, or any, direction of propagation has the property of optical axis for a given (symmetric) dielectric tensor $\hat{\varepsilon}$. This leads to the following classification of the optical anisotropy:

- Biaxial
  Two optical axes exist for the dielectric tensor with mutually different values along the principal axes of the ellipsoid of Eq. (1.77). Occurs for crystals of triclinic, monoclinic, and orthorhombic structure. The directions of optical axes may vary with frequency ω.
- Uniaxial
  A single optical axis exists for the dielectric tensor having rotational symmetry of its ellipsoid of Eq. (1.77). Occurs for crystals of rhombohedral, tetragonal, and hexagonal structure. The optical axis coincides with the high–symmetry axis of the crystal.
- Isotropic
  The ellipsoid of Eq. (1.77) is a sphere. all of the directions of propagation are equivalent. Occurs for cubic crystals or amorphous materials.

1.2.6 Plane Waves in Isotropic Medium

An isotropic material is described by its complex permittivity, $\varepsilon = \varepsilon_1 + i\varepsilon_2$, or the complex refractive index, $N = n + ik = \sqrt{\varepsilon}$, see Section 1.2.4. The monochromatic plane wave of Eq. (1.71) is characterized by its wavevector $\vec{k} = \vec{k}_1 + i\vec{k}_2$; the real and imaginary parts, $\vec{k}_1$ and $\vec{k}_2$, need not be parallel. Let us denote the angle between them by θ, and their magnitudes by k_1 and k_2. A plane wave is called *homogeneous* for $\theta = 0$; otherwise it is *inhomogeneous*. Examples are given in Section 1.5.1. The plane wave is a product of a harmonic vibration traveling in the direction of $\vec{k}_1$, and an exponential damping in the direction of $\vec{k}_2$,

$$\vec{E}_\omega(\vec{r}) = \vec{E}_0 \exp(i\vec{k}_1 \cdot \vec{r}) \exp(-\vec{k}_2 \cdot \vec{r}). \tag{1.78}$$

The plane wave (1.71) solves the wave equation (1.72) for

$$\vec{k} \cdot \vec{k} = k_1^2 - k_2^2 + i2k_1k_2 \cos\theta = \frac{\omega^2}{c^2}(\varepsilon_1 + i\varepsilon_2). \tag{1.79}$$

Equating the real and imaginary parts, we obtain the magnitudes of the wavevector as

$$k_1 = \frac{\omega}{c}\sqrt{\frac{\varepsilon_1 + \sqrt{\varepsilon_1^2 + (\varepsilon_2/\cos\theta)^2}}{2}}, \quad k_2 = \frac{\omega}{c}\sqrt{\frac{-\varepsilon_1 + \sqrt{\varepsilon_1^2 + (\varepsilon_2/\cos\theta)^2}}{2}}. \tag{1.80}$$

For a homogeneous wave, we obtain the following simple relations of the magnitudes of the wavevector and the real and imaginary parts of the refractive index,

$$k_1 = \frac{\omega}{c} n, \quad k_2 = \frac{\omega}{c} k. \tag{1.81}$$

In a non–absorbing medium of $k = \varepsilon_2 = 0$, the plane wave is either undamped ($k_2 = 0$), or evanescent ($\theta = \pi/2$, the harmonic spatial dependence along k_1 is damped exponentially in a perpendicular direction). The latter case occurs for total reflection of obliquely incident light, see Section 1.5.1.

Let us resolve the amplitude $\vec{E}_0$ of the plane wave of Eq. (1.78) into its real and imaginary parts, $\vec{E}_0 = \vec{E}_{01} + i\vec{E}_{02}$. Similarly, the amplitude of the magnetic intensity is written as $\vec{H}_0 = \vec{H}_{01} + i\vec{H}_{02}$. The transversality of the fields, following from Maxwell equations, can be expressed as the condition of vanishing of the scalar products of the amplitudes and wavevector,

$$\vec{k} \cdot \vec{E}_0 = 0, \quad \vec{k} \cdot \vec{H}_0 = 0, \quad \vec{E}_0 \cdot \vec{H}_0 = 0. \tag{1.82}$$

Written for the real and imaginary parts of the vectors,

$$\begin{aligned}
&\vec{k}_1 \cdot \vec{E}_{01} - \vec{k}_2 \cdot \vec{E}_{02} = 0, && \vec{k}_2 \cdot \vec{E}_{01} + \vec{k}_1 \cdot \vec{E}_{02} = 0,\\
&\vec{k}_1 \cdot \vec{H}_{01} - \vec{k}_2 \cdot \vec{H}_{02} = 0, && \vec{k}_2 \cdot \vec{H}_{01} + \vec{k}_1 \cdot \vec{H}_{02} = 0,\\
&\vec{E}_{01} \cdot \vec{H}_{01} - \vec{E}_{02} \cdot \vec{H}_{02} = 0, && \vec{H}_{02} \cdot \vec{E}_{01} + \vec{H}_{01} \cdot \vec{E}_{02} = 0.
\end{aligned} \tag{1.83}$$

Two simple possibilities of alignment of the vectors are the following:

- Transverse electric (TE) mode
  The real and imaginary parts of the amplitude $\vec{E}_0$ are parallel. From Eqs. (1.83) we find that $\vec{k}_1 \cdot \vec{E}_{01} = \vec{k}_2 \cdot \vec{E}_{01} = \vec{H}_{01} \cdot \vec{E}_{01} = \vec{H}_{02} \cdot \vec{E}_{01} = 0$, and the same result for $\vec{E}_{01}$ replaced by $\vec{E}_{02}$. Consequently, the four vectors $\vec{k}_1$, $\vec{k}_2$, $\vec{H}_{01}$, and $\vec{H}_{02}$ lie in the plane perpendicular to the electric intensity $\vec{E}_{01}$.
- Transverse magnetic (TM) mode
  The real and imaginary parts of the amplitude $\vec{H}_0$ are parallel. The four vectors $\vec{k}_1$, $\vec{k}_2$, $\vec{E}_{01}$, and $\vec{E}_{02}$ lie in the plane perpendicular to the magnetic intensity $\vec{H}_{01}$.

Note that the distinction between TE and TM modes is relevant for inhomogeneous waves only. Any plane wave is a superposition of the TE and TM modes.

The electric and magnetic intensities of an optical wave are observed indirectly via the energy flux registered by a detector. The energy density and energy flux density are

$$w = \frac{1}{2}\,(\vec{E}\cdot\vec{D} + \vec{H}\cdot\vec{B}) \quad \text{and} \quad \vec{S} = \vec{E}\times\vec{H}, \tag{1.84}$$

respectively. A light wave loses its energy when traveling through an absorbing medium. The rate of the energy density loss is

$$Q = \langle \vec{j}\cdot\vec{E}\rangle, \tag{1.85}$$

where $\vec{j}$ is the induced current density, and the brackets denote the time average. The real parts of $\vec{j}$ and $\vec{E}$ have to be inserted into Eq. (1.85), which is analogous to the Joule heat production in a conductor. Assume a monochromatic plane wave with the real magnitude $E_0 \cos \omega t$. The real part of the induced current density obtained from Eqs. (1.44) through (1.47) is $\sigma_1 E_0 \cos(\omega t) = \omega\varepsilon_0\varepsilon_2 E_0 \cos(\omega t)$. Thus, the absorbed energy per unit time,

$$Q = \sigma_1\langle E_0^2 \cos^2(\omega t)\rangle = \sigma_1 \frac{E_0^2}{2} = \omega\varepsilon_o\varepsilon_2 \frac{E_0^2}{2}, \tag{1.86}$$

is measured by the real part, σ_1, of the complex conductivity, and the imaginary part, ε_2, of the complex dielectric function.

1.3 Representations of Polarization

In this section, we deal with several representations of pure polarization states. We start with the description using real quantities, and provide links with the concise expressions using complex numbers.

Assume a monochromatic light wave traveling along the positive direction of the z axis of a Cartesian, right–handed coordinate system x–y–z. Its electric vector is perpendicular to z, with the components along x and y given by

$$E_x(t) = X\cos(-\omega t + \delta_x), \quad E_y(t) = Y\cos(-\omega t + \delta_y) \tag{1.87}$$

Here ω is the angular frequency, X and Y are the amplitudes of the x and y components, δ_x and δ_y are their absolute phases at the initial time $t = 0$. The endpoint of the intensity vector performs a fast periodic movement

along an *elliptic* trajectory limited to the intervals $\pm|X|$ and $\pm|Y|$ in the x and y directions, respectively. The equation of the ellipse results from Eqs. (1.87) by excluding ωt. The latter can be rewritten as

$$\begin{aligned} Y E_x \sin\delta_y - X E_y \sin\delta_x &= -XY \cos(\omega t) \sin(\delta_x - \delta_y), \\ Y E_x \cos\delta_y - X E_y \cos\delta_x &= XY \sin(\omega t) \sin(\delta_x - \delta_y). \end{aligned} \tag{1.88}$$

Squaring and adding them gives the equation of the ellipse,

$$Y^2E_x^2 + X^2E_y^2 - 2XYE_xE_y \cos(\delta_x - \delta_y) = X^2Y^2 \sin^2(\delta_x - \delta_y). \tag{1.89}$$

1.3.1 Representation by Ellipsometric Angles

The most general planar harmonic vibration at a given point in space is specified completely by the four real parameters X, Y, δ_x, δ_y, linked to the chosen x–y frame. However, the sum $X^2 + Y^2$ is proportional to the intensity of the wave, which is irrelevant in the description of the state of polarization. Further, the absolute phases δ_x, δ_y are not observable in ellipsometric studies, due to the very short period of oscillations, $2\pi/\omega$. Thus, we are left with only two real parameters characterizing the general elliptic polarization of Eqs. (1.87). Since a suitable choice of the absolute phases guarantees the possibility to choose non–negative values of X and Y, we can take $X = E_0 \sin\psi_{xy}$, $Y = E_0 \cos\psi_{xy}$ with $E_0 > 0$ and the angle ψ_{xy} limited to the first quadrant. Note that $\tan\psi_{xy}$ is the amplitude of the vibrations along x measured relative to the amplitude along y. The second real parameter of the elliptically polarized wave is the relative *phase shift* of the x–axis vibration with respect to the y–axis vibration, called Δ_{xy}. In summary, the two angles describing the elliptic polarization are

$$\tan\psi_{xy} = X/Y,\ \psi_{xy} \in \langle 0, \pi/2 \rangle,\ \Delta_{xy} = \delta_x - \delta_y,\ \Delta_{xy} \in \langle -\pi, \pi \rangle. \tag{1.90}$$

Using the two angles and the amplitude E_0, the ellipse of Eq. (1.89) is

$$\frac{E_x^2}{E_0^2} + \tan^2\psi_{xy} \frac{E_y^2}{E_0^2} - 2\tan\psi_{xy} \frac{E_x}{E_0} \frac{E_y}{E_0} \cos\Delta_{xy} = \sin^2\psi_{xy} \sin^2\Delta_{xy}. \tag{1.91}$$

The pair of angles (ψ,Δ) is very useful in the treatment of polarized light; it actually became a symbol of ellipsometry. It should be remembered that both of these angles depend on the coordinate system chosen in the plane of polarization. We stress this point by using the subscripts xy in the corresponding symbols.

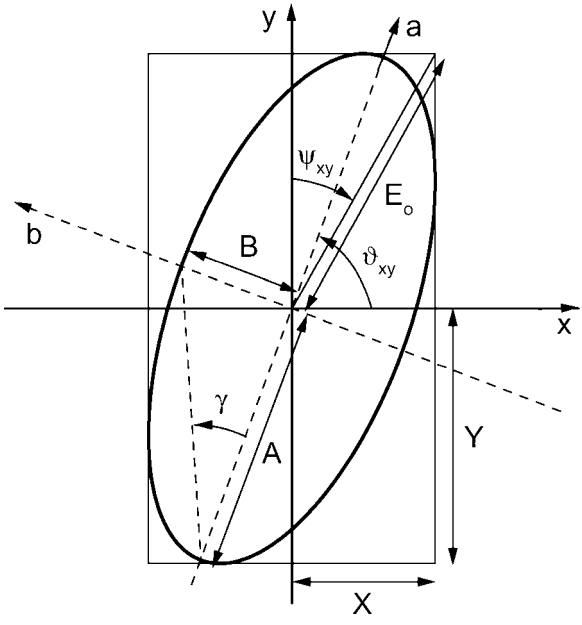


Figure 1.3 The ellipse of polarization and its parameters.

An example of the polarization ellipse is shown in Fig. 1.3. The *sense* of the motion along the ellipse, which is present in Eqs. (1.87) and absent in Eqs. (1.88) and (1.90), *can be observed* in suitably arranged ellipsometric measurements. It depends on the sign of the phase shift Δ_{xy}. Namely, the ellipse is traversed clockwise when looking against the direction of propagation (i.e., against the positive direction of the z axis) for $\Delta_{xy} > 0$; the polarization is then called *right-handed*. For $\Delta_{xy} < 0$ the sense of the motion is counter–clockwise, denoted as *left-handed* polarization.

A simpler description of the geometrical form of the polarization ellipse is obtained in the Cartesian coordinate system oriented along its axes, see Fig. 1.1. Let a be the coordinate along the major axis of the length $2A$, b along the minor axis of the length $2B$, $0 \leq B \leq A \leq E_0$. The angle ϑ_{xy} between the positive directions of the x and a axes is called *azimuth*. It defines the orientation of the ellipse and can be confined to the interval $\langle -\pi/2, \pi/2 \rangle$. In the coordinate system a–b, the components of the electric intensity vector are

$$E_a = E_x \cos\vartheta_{xy} + E_y \sin\vartheta_{xy}, \quad E_b = -E_x \sin\vartheta_{xy} + E_y \cos\vartheta_{xy}, \quad (1.92)$$

and the ellipse of Eq. (1.89) is

$$B^2E_a^2 + A^2E_b^2 = A^2B^2. \tag{1.93}$$

The azimuth can be complemented by the ratio B/A, which is independent of the coordinate system. It can be treated conveniently by introducing the *ellipticity angle* γ,

$$\tan\gamma = \pm B/A, \quad \gamma \in \langle -\pi/4, \pi/4 \rangle, \tag{1.94}$$

assuming positive and negative values of γ for right–handed and left–handed polarizations, respectively. The dimensionless value of $\tan\gamma$ is called *ellipticity* e; it is confined to the interval $\langle -1,1 \rangle$. A comparison of the equations of the polarization ellipse in both coordinate systems leads to the following relations for its parameters:

$$\begin{aligned} &X^2 = A^2\cos^2\vartheta_{xy} + B^2\sin^2\vartheta_{xy}, \quad Y^2 = A^2\sin^2\vartheta_{xy} + B^2\cos^2\vartheta_{xy}, \\ &X^2 + Y^2 = A^2 + B^2 = E_0^2, \quad X^2 - Y^2 = (A^2 - B^2)\cos 2\vartheta_{xy}, \\ &XY|\sin\Delta_{xy}| = AB, \quad 2XY\cos\Delta_{xy} = (A^2 - B^2)\cos 2\vartheta_{xy}. \end{aligned} \tag{1.95}$$

The lengths of the major and minor axes are

$$\begin{aligned} A &= E_0\sqrt{\frac{1 + \sqrt{1 - \sin^2 2\psi_{xy}\sin^2\Delta_{xy}}}{2}}, \\ B &= E_0\sqrt{\frac{1 - \sqrt{1 - \sin^2 2\psi_{xy}\sin^2\Delta_{xy}}}{2}}. \end{aligned} \tag{1.96}$$

From Eqs. (1.95) and (1.94), trigonometric functions of the azimuth and ellipticity angles can be expressed concisely in terms of the ellipsometric angles in the x–y coordinate system as

$$\tan 2\vartheta_{xy} = -\tan 2\psi_{xy}\cos\Delta_{xy}, \quad \sin 2\gamma = \sin 2\psi_{xy}\sin\Delta_{xy}. \tag{1.97}$$

Conversely,

$$\cos 2\psi_{xy} = -\cos 2\gamma\cos 2\vartheta_{xy}, \quad \tan\Delta_{xy} = \frac{\tan 2\gamma}{\sin 2\vartheta_{xy}}. \tag{1.98}$$

The following explicit expressions are useful in conversions of the ellipsometric angles,

$$\psi_{xy} = \arctan\sqrt{\frac{1 + \tan^2\vartheta_{xy}\tan^2\gamma}{\tan^2\vartheta_{xy} + \tan^2\gamma}}, \qquad \psi_{xy}\in\langle 0,\pi/2\rangle,$$

$$\Delta_{xy} = 2\arctan\frac{\sqrt{1 - \cos^2 2\gamma\cos^2 2\vartheta_{xy}} - \sin 2\vartheta_{xy}\cos 2\gamma}{\sin 2\gamma}, \quad \Delta_{xy}\in\langle -\pi,\pi\rangle,$$

$$\theta_{xy} = \arctan\frac{\sin 2\psi_{xy}\cos\Delta_{xy}}{\sqrt{1 - \sin^2 2\psi_{xy}\sin^2\Delta_{xy}} - \cos 2\psi_{xy}}, \qquad \theta_{xy}\in\langle -\pi/2,\pi/2\rangle,$$

$$\gamma = \frac{1}{2}\arctan\frac{\sin 2\psi_{xy}\sin\Delta_{xy}}{\sqrt{1 - \sin^2 2\psi_{xy}\sin^2\Delta_{xy}}}, \qquad \gamma\in\langle -\pi/4,\pi/4\rangle. \tag{1.99}$$

The two pairs of angles, (ψ_{xy},Δ_{xy}) and (ϑ_{xy},γ), can be used interchangeably to describe the polarization ellipse. The benefit of using the second pair consists primarily of the fact that the ellipticity angle γ describes a polarization state independently of the coordinate system. Note that in the privileged coordinate system *a*–*b*, the ellipsometric angles defined according to Eqs. (1.90) become

$$\psi_{ab} = \frac{\pi}{2} - |\gamma|,\ \psi_{ab}\in\left\langle\frac{\pi}{4},\frac{\pi}{2}\right\rangle,\ \Delta_{ab} = \pm\frac{\pi}{2}.$$

The first of these equations results from the definition $\tan\psi_{ab} = A/B = 1/|\tan\gamma|$. The positive and negative values of Δ_{ab} correspond to the right– and left–handed polarizations, respectively.

1.3.2 Special Cases: Linear and Circular Polarization

Two special cases of polarization states of light are *linear* and *circular* polarization. They occur when the parameter independent of the coordinate system, ellipticity e, assumes the values of zero and ± 1, respectively. These states are readily identified in the series of elliptic polarizations shown in Fig. 1.4.

For linear polarization, the length of the minor axis is $B = 0$. The ellipse collapses into a linear segment of the length $2A = 2E_0$ which has no handedness. Different linearly polarized states are distinguished by their azimuth ϑ_{xy} in a given coordinate system. Except for the alignment of the linear segment parallel to x or y, the relative phase shift of the x and y components has to be either zero or π, i.e., the two components vibrate either in–phase or with the opposite phase. The latter case

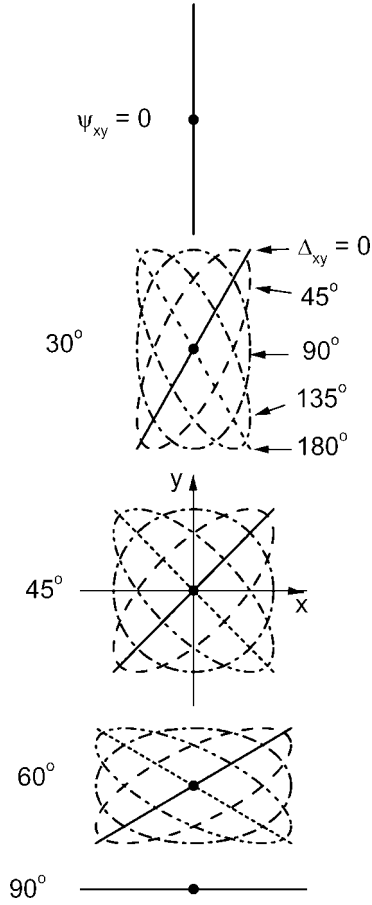


Figure 1.4 Different states of polarization characterized by the ellipsometric angles ψ and Δ (0 . . . solid lines; ±45° . . . dashed lines; ±90° . . . dash–dotted lines; ±135° . . . dash–dot–dotted lines; ±180° . . . dotted lines) in the *x*–*y* coordinate system, with the origin marked by full circles. The sense of precession along the ellipses is clockwise and counter–clockwise for Δ_{xy} positive and negative, respectively.

is usually called "half–wavelength", or "$\lambda/2$" shift. For $\vartheta_{xy} = 0$ or $\pi/2$, the relative phase shift becomes irrelevant, since one of the components vanishes.

The two circular polarizations have indeterminate azimuth, since the distinction between the major and minor axes is lost, $A = B = E_{0'}/\sqrt{2}$. The ellipsometric angles ψ and Δ are the same in any coordinate system,

namely, $\psi = \pi/4$ and $\Delta = \pm\pi/2$. In other words, the circularly polarized wave results from the superposition of two components of equal strength aligned along any two orthogonal axes. Further, the two (linearly polarized) components are shifted in phase by either $\pi/2$, producing the right–circular polarization, or by $-\pi/2$, resulting in the left–circular polarization. The magnitude of this phase shift is called "quarter–wavelength", or "$\lambda/4$".

1.3.3 Orthogonal Polarization States

A state of linear polarization, described completely by its azimuth ϑ_{xy}, is *orthogonal* to the state of linear polarization having the azimuth $\vartheta_{xy} \pm \pi/2$. In order to confine the range of azimuths to $\langle 0,\pi\rangle$, the plus sign is taken for $\vartheta_{xy} < \pi/2$ and the minus sign otherwise. The orthogonality here has the usual meaning: the two linear segments of the electric field vibrations form a right angle. It should be noted that this property is independent of the coordinate system. Any of the linear polarization shown in Fig. 1.4 has its orthogonal counterpart there.

The concept of orthogonality can be extended to the states of arbitrary elliptic polarization. We call two states mutually orthogonal when they are characterized by (i) orthogonal azimuths, and (ii) ellipticity angles of the same magnitude and opposite sign, i.e., (ϑ_{xy},γ) and $(\vartheta_{xy} \pm \pi/2,-\gamma)$. The corresponding ellipses of polarization have their major and minor axes mutually orthogonal, the minor/major axis ratio is the same, and they are traversed in the opposite sense. Any of the polarization states shown in Fig. 1.4 has its orthogonal counterpart there; however, we have to distinguish the handedness in finding orthogonal pairs of non-zero ellipticity. The state orthogonal to (ψ_{xy},Δ_{xy}) is $(\pi/2 - \psi_{xy}, \Delta_{xy} - \pi)$ for $\Delta_{xy} > 0$, and $(\pi/2 - \psi_{xy}, \pi + \Delta_{xy})$ for $\Delta_{xy} < 0$. The only two circularly polarized states are orthogonal to each other.

1.3.4 Representation by Complex Numbers

Two real parameters needed to represent a general state of elliptic polarization can be joined conveniently in a single *complex* number. One of the simplest possibilities uses the ellipsometric angles ψ and Δ in the Cartesian x–y coordinate system to form the complex value of χ_{xy} as

$$\chi_{xy} = \chi_r + i\chi_i = \frac{Y \exp(-i\delta_y)}{X \exp(-i\delta_x)} = \frac{1}{\tan\psi_{xy}} \exp(i\Delta_{xy}). \qquad (1.100)$$

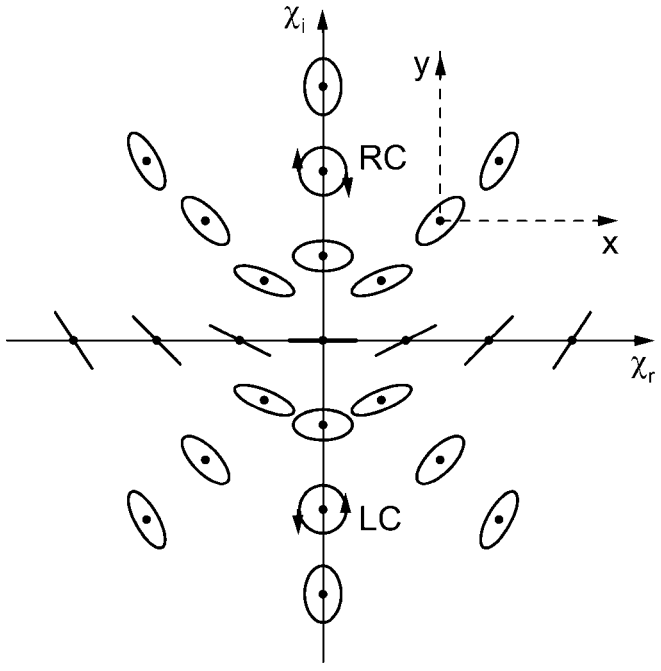


Figure 1.5 States of polarization represented by points (full circles) in the complex plane $\chi_{xy} = \chi_r + i\chi_i$. In the whole upper half–plane, the sense of precession is clockwise; $\chi_{xy} = +i$ represents the right–handed circular polarization (RC). In the whole lower half–plane, the sense of precession is counter–clockwise; $\chi_{xy} = -i$ represents the left–handed circular polarization (LC). At the origin, $\chi_{xy} = 0$, the linear polarization is aligned along x, at infinity, $|\chi_{xy}| \to \infty$, the linear polarization is along y.

With the two angles limited to the ranges of Eqs. (1.90), there is one–to–one correspondence between the elliptic polarization state and the value of χ_{xy} in the complex plane. Note that the values of real and imaginary parts, χ_r and χ_i, of χ_{xy} depend on the choice of the x–y coordinate system; we have omitted the xy subscript here in order to simplify notation. Expressed in terms of the azimuth and ellipticity angles of Eqs. (1.97), the complex quantity χ_{xy} is

$$\chi_{xy} = \frac{\tan\vartheta_{xy} + i\tan\gamma}{1 - i\tan\vartheta_{xy}\tan\gamma}. \tag{1.101}$$

This representation provides a convenient overview of all possible states of elliptic polarization, which is shown in Fig. 1.5:

- The origin, $\chi_{xy} = 0$, represents the linear polarization parallel to the x axis, since $\psi_{xy} \to \pi/2$ and Δ_{xy} is irrelevant in Eq. (1.100). Similarly, the point at infinity of the complex plane, $|\chi_{xy}| = \sqrt{\chi_r^2 + \chi_i^2} = \infty$, represents the linear polarization along the y axis, since $\psi_{xy} = 0$ and Δ_{xy} is irrelevant in Eq. (1.100).
- All states of linear polarization (vanishing ellipticity, $\gamma = 0$) correspond to points on the real axis, with $\chi_r = \tan\vartheta_{xy} = \pm 1/\tan\psi_{xy}$. On the positive and negative parts of the real axis, the azimuths are from the intervals $(0,\pi/2)$ and $(\pi/2,\pi)$, respectively.
- The points on the imaginary axis, $\chi_r = 0$, represent the polarizations with the major and minor axes of the ellipse aligned along x and y directions, with $\chi_i = \pm 1/\tan\psi_{xy}$. The major axis is along the x direction for $\chi_i \in (-1,1)$; in this case, $\chi_i = \tan\gamma$ as obtained from Eq. (1.101) with $\vartheta_{xy} = 0$. The major axis is along the y direction for $|\chi_i| > 1$; in this case, $\chi_i = 1/\tan\gamma$ as obtained from Eq. (1.101) with $\vartheta_{xy} = \pi/2$.
- The two points $\chi_{xy} = \pm i$ on the imaginary axis represent the right– and left–handed circular polarizations, respectively.

According to the definition of Eq. (1.100), the states with a constant angle ψ_{xy} lie on a circle with the center in the origin, and the diameter of $1/\tan\psi_{xy}$. The states with a constant phase shift Δ_{xy} form a straight line starting at origin and forming the angle of Δ_{xy} with the positive real axis.

The complex representation of the states with a constant azimuth and ellipticity angles is determined by inverting Eq. (1.101). Since

$$\chi_r = \frac{\tan\vartheta_{xy}(1 - \tan^2\gamma)}{1 + \tan^2\vartheta_{xy}\tan^2\gamma}, \quad \chi_i = \frac{\tan\gamma(1 + \tan^2\vartheta_{xy})}{1 + \tan^2\vartheta_{xy}\tan^2\gamma}, \tag{1.102}$$

we obtain

$$\tan 2\vartheta_{xy} = \frac{2\chi_r}{1 - \chi_r^2 - \chi_i^2}, \quad \sin 2\gamma = \frac{2\chi_i}{1 + \chi_r^2 + \chi_i^2}. \tag{1.103}$$

The two equations can be rearranged into

$$(\chi_r + 1/\tan 2\vartheta_{xy})^2 + \chi_i^2 = 1/\sin^2 2\vartheta_{xy}, \quad \chi_r^2 + (\chi_i - 1/\sin 2\gamma)^2 = 1/\tan^2 2\gamma. \tag{1.104}$$

The first of Eqs. (1.104) describes a circle with the center on the real axis at $-1/\tan 2\vartheta_{xy}$, having the radius of $1/\sin 2|\vartheta_{xy}|$. Any such circle passes through the points of the complex plane representing the circular

polarizations, $\pm i$, due to the indeterminate azimuth of these states. The second of Eqs. (1.104) describes a circle with the center on the imaginary axis at $1/\sin 2\gamma$, having the radius of $1/\tan 2|\gamma|$.

The pairs of orthogonal polarization states are easily recognized in the complex–number representation. Recalling the definition of the orthogonality (orthogonal azimuths, and ellipticity angles of the same magnitude and opposite sign), we have the state orthogonal to that of Eqs. (1.100) and (1.101),

$$\tilde{\chi}_{xy} = \frac{-1 - i\tan\vartheta_{xy}\tan\gamma}{\tan\vartheta_{xy} - i\tan\gamma} = -\tan\psi_{xy}\exp(i\Delta_{xy}) = -\frac{1}{\chi^*_{xy}}, \qquad (1.105)$$

where the asterisk denotes the complex conjugation. The condition $\chi^*_{xy}\cdot\tilde{\chi}_{xy} = -1$ is an extremely concise expression of the fact that the two polarization states, χ_{xy} and $\tilde{\chi}_{xy}$, are orthogonal.

Let us note that there are several equivalent possibilities for choosing the complex–number representation of the polarization states. For example, we can take the complex plane $\tilde{\chi}_{xy}$ of Eq. (1.105) with the obvious modifications, such as the origin $\tilde{\chi}_{xy} = 0$ representing the linear polarization along y instead of x. Another equivalent representation is the complex number

$$\rho_{xy} = \tan\psi_{xy}\exp(i\Delta_{xy}), \qquad (1.106)$$

differing from $\tilde{\chi}_{xy}$ by the opposite sign. This is formally the same as adding π to the phase shift Δ_{xy}, or, inverting the sense of one of the x or y axes in Eq. (1.87). Accordingly, the upper half–plane, $\mathrm{Im}(\rho_{xy}) > 0$, represents left–handed states, and the lower half–plane right–handed states.

1.3.5 Light Intensity, Detection of Polarization State

Detectors used in ellipsometric setups provide signals proportional to the incoming *energy flux* of a light wave, called its *intensity* I. This is the time–average $\langle\vec{S}(t)\rangle$ of the rapidly oscillating Poynting vector $\vec{S}(t)$, which is the vector product of the instantaneous intensities of electric and magnetic fields, $\vec{S}(t) = \vec{E}(t)\times\vec{H}(t)$. The SI unit of $\vec{S}$ is $\mathrm{W/m^2}$, resulting from the units V/m and A/m of $\vec{E}$ and $\vec{H}$, respectively. Since the vectors $\vec{E}$ and $\vec{H}$ are mutually perpendicular and their magnitudes are proportional, we can take the time–averaged square of the magnitude of $\vec{E}$ as a convenient quantity proportional to the energy flux.

There is usually no interest in absolute light intensities in ellipsometric measurements. This justifies the use of *relative* intensity I/I_0,

taken with respect to a suitable reference value I_0. Here we will use the time average of the square of $E_0 \cos \omega t$,

$$I_0 \equiv \langle E_0^2 \cos^2 \omega t \rangle = \frac{\omega}{\pi} \int_0^{\pi/\omega} E_0^2 \cos^2 \omega t \, dt = \frac{E_0^2}{2}. \qquad (1.107)$$

This is representative of a monochromatic linearly polarized light wave of the amplitude E_0 and arbitrary azimuth. It can easily be proved that the general elliptic polarization of Eqs. (1.87), with $X = E_0 \sin \psi_{xy}$ and $Y = E_0 \cos \psi_{xy}$, gives the same value,

$$I \equiv \langle X^2 \cos^2 (\omega t + \delta_x) + Y^2 \cos^2 (\omega t + \delta_y) \rangle = \frac{E_0^2}{2}. \qquad (1.108)$$

In other words, the relative intensity of any elliptically polarized light having the amplitude E_0 is unity.

A state of linear polarization can be very easily detected by observing its intensity after passing a *linear polarizer*. This is a device transmitting or rejecting incident linearly polarized light, for the direction of the polarization parallel or perpendicular to the transmission axis of the polarizer, respectively. An *ideal* linear polarizer passes freely the light of parallel polarization, while it stops completely the crossed polarization. The azimuth ϑ_{xy} is obtained simply as the angle $\pm\pi/2$ off the orientation of the transmission axis when the intensity vanishes; this is used as an essential ingredient of ellipsometric *nulling schemes*.

A general elliptic polarization can also be probed by registering the intensity transmitted through a linear polarizer set to suitable orientations. Assume the polarized vibrations of Eqs. (1.87), characterized by the ellipsometric angles (1.90). Let the transmission axis of an ideal linear polarizer form an angle α with the x axis of our coordinate system. Setting the irrelevant absolute phase δ_y to zero, the harmonic vibration along the transmission axis, $E_x(t) \cos\alpha + E_y(t) \sin\alpha$, becomes

$$\begin{aligned} E_\alpha(t) = E_0[&(\sin \psi_{xy} \cos \Delta_{xy} \cos \alpha + \cos \psi_{xy} \sin \alpha) \cos \omega t \\ &- (\sin \psi_{xy} \sin \Delta_{xy} \cos \alpha) \sin \omega t]. \end{aligned} \qquad (1.109)$$

The time average of its square contains two equal terms, $\langle \cos^2 \omega t \rangle = \langle \sin^2 \omega t \rangle = 1/2$, and one vanishing term, $\langle \cos \omega t \cdot \sin \omega t \rangle = 0$. A straightforward calculation gives the relative intensity of the wave $E_\alpha(t)$ as

$$\frac{I(\alpha)}{I_0} = \frac{1}{2}(1 - \cos 2\psi_{xy} \cos 2\alpha + \sin 2\psi_{xy} \cos \Delta_{xy} \sin 2\alpha). \qquad (1.110)$$

Using Eqs. (1.97) and (1.98), the intensity can be written in terms of the azimuth and ellipticity angles as follows

$$\frac{I(\alpha)}{I_0} = \frac{1}{2}[1 + \cos 2\gamma \cos 2(\vartheta_{xy} - \alpha)]. \tag{1.111}$$

Finally, the representation of the elliptic polarization by the complex number of Eqs. (1.100) and (1.101) provides the relative intensity as

$$\frac{I(\alpha)}{I_0} = \frac{1}{2}\left[1 + \frac{1 - |\chi_{xy}|^2}{1 + |\chi_{xy}|^2}\cos 2\alpha + \frac{2\mathrm{Re}(\chi_{xy})}{1 + |\chi_{xy}|^2}\sin 2\alpha\right]. \tag{1.112}$$

The intensity is an even function of the phase shift Δ_{xy}, ellipticity angle γ, and imaginary part of the complex value χ_{xy}. This means identical data observed for polarizations differing solely in the handedness. Since the orientations of the polarizer differing by half of a full rotation are equivalent, the intensity profiles are periodic functions of 2α.

Figure 1.6 shows the dependence of the relative intensity on the orientation of the linear polarizer for elliptic, linear, and circular polarizations. The positions of maximum intensity I_{max} in the first two quadrants are identified as the azimuths ϑ_{xy} of the polarization state; they are marked by dashed arrows. The ratio of minimum to maximum intensity is ellipticity squared,

$$\frac{I_{\min}}{I_{\max}} = \frac{1 - \cos 2\gamma}{1 + \cos 2\gamma} = \tan^2 \gamma. \tag{1.113}$$

The harmonic dependence on the angle $2\alpha_{xy}$ is superposed on the constant contribution of $1/2$; the latter is observed for both circular polarizations, since half of the intensity is rejected by the polarizer at any angular position. Two real parameters of the polarization state in Eqs. (1.109) through (1.111) are encoded in the "cosine" and "sine" Fourier coefficients, a_{xy} and b_{xy}, of the dependence

$$I(\alpha) = \frac{I_0}{2}(1 + a_{xy}\cos 2\alpha + b_{xy}\sin 2\alpha), \tag{1.114}$$

with

$$\begin{aligned} a_{xy} &= -\cos 2\psi_{xy} = \cos 2\gamma \cos 2\vartheta_{xy} = (1 - |\chi_{xy}|^2)/(1 + |\chi_{xy}|^2), \\ b_{xy} &= \sin 2\psi_{xy} \cos \Delta_{xy} = \cos 2\gamma \sin 2\vartheta_{xy} = 2\mathrm{Re}(\chi_{xy})/(1 + |\chi_{xy}|^2). \end{aligned} \tag{1.115}$$

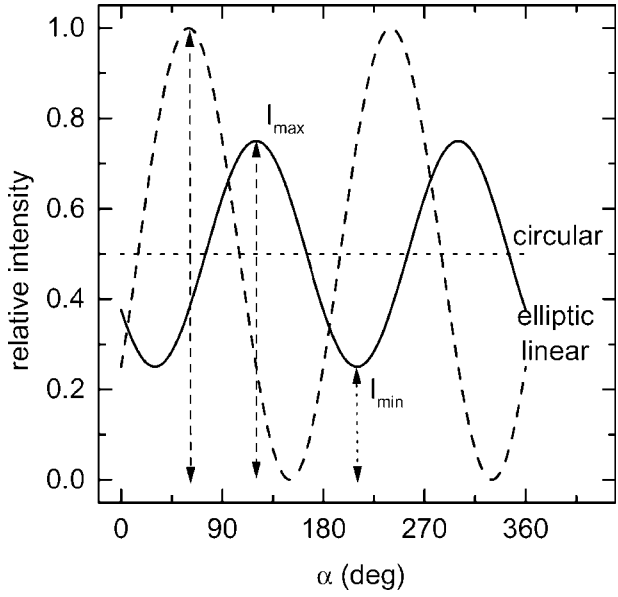


Figure 1.6 Relative intensity detected behind an ideal linear polarizer with the orientation α with respect to the x axis for an elliptic ($\gamma = \pm 30°$, $\theta_{xy} = 120°$; solid line), linear ($\gamma = 0$, $\theta_{xy} = 60°$; dashed line), and circular ($\gamma = \pm 45°$; dotted line) polarization.

In order to determine the pair of real values of a_{xy} and b_{xy}, intensity for at least three positions of the polarizer has to be measured. A larger number of the intensity readings is usually performed in *photometric ellipsometric* systems.

The handedness can be detected using a linear polarizer in combination with a *retarder*, which shifts the phases in a controlled manner. In an ideal linear retarder, the light polarized linearly along its *fast axis* moves faster then for the polarization along its *slow axis*. This results in a phase retardation of the latter polarization by an angle δ_r. If the fast axis of the retarder is oriented along the x axis of the coordinate system, the value of δ_r is subtracted from the relative phase Δ_{xy} of Eq. (1.90). A particularly useful value of the retardation is $\pi/2$, called quarter–wave. Figure 1.7 shows the intensities behind a polarizer for the elliptically polarized light. The solid line results for both right– and left–handed light of $\gamma = \pm 5°$ and $\vartheta_{xy} = 70°$ (or, equivalently, $\psi_{xy} = 20.5°$ and $\Delta_{xy} = \pm 15.3°$). However, an ideal linear quarter–wave retarder inserted in front of the polarizer, having its fast axis parallel to the x axis, changes the state of polarization seen by the

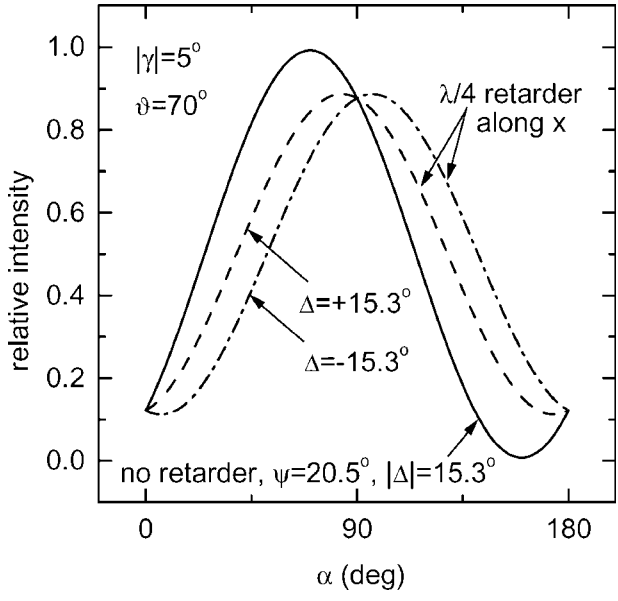


Figure 1.7 Relative intensity detected for an elliptic polarization of $\gamma = \pm 5°$ and $\vartheta_{xy} = 70°$ behind an ideal linear polarizer (solid line). An ideal linear quarter–wave retarder, inserted in front of the polarizer with its fast axis along *x*, distinguishes between the right–handed (dashed line) and left–handed (dash–dotted line) polarizations.

polarizer. Note that the sign of the relative phase Δ_{xy} of the probed polarization has no influence on the ellipticity angle behind the retarder, γ_r. Using the second of Eqs. (1.97) with the retardation of $\pi/2$ subtracted from the phase shift, we get

$$\sin 2\gamma_r = \sin 2\psi_{xy} \sin(\Delta_{xy} - \pi/2) = -\sin 2\psi_{xy} \cos \Delta_{xy},$$

which is an even function of Δ_{xy}. This is the reason for the identical maximum and minimum values of intensity for the polarizations of opposite handedness. On the contrary, the azimuth of the elliptic polarization behind the retarder depends on the handedness of the analyzed light. This is seen as the differing positions of the maximum intensity for positive and negative phase in Fig. 1.7. Thus, a proper use of retarders allows us to retrieve the polarization state including the sign of its ellipticity. Another benefit of using a retarder in photometric ellipsometric setups is the increased accuracy of the parameters, most importantly for the phase shifts Δ_{xy} close to zero or π.

1.4 Propagation of Polarized Light

The description of propagation of polarized light through an ellipsometer becomes cumbersome if not organized properly. A convenient scheme of dealing with pure elliptic polarization is based on *Jones vectors* and linear operations with them, represented by *Jones matrices*. The partially polarized light is best described using quantum mechanical formalism; for practical purposes, Stokes vectors and Mueller matrices are a convenient extension of the matrix methods.

1.4.1 Jones Vectors

The state of polarization described in various ways in Section 1.3 can be encoded into the column vector with complex components,

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} X \exp(i\delta_x) \\ Y \exp(i\delta_y) \end{bmatrix}. \tag{1.116}$$

We assume here the monochromatic wave of Eq. (1.87) traveling in the positive z direction in a right–handed x–y–z coordinate system. The Jones vector, Eq. (1.116), multiplied by the harmonic time–dependence $\exp(-i\omega t)$, represents a superposition of the harmonic vibration with real amplitudes X and Y in the x and y directions, respectively.

As discussed in Section 1.3.5 the light intensity I is proportional to $X^2 + Y^2$, i.e., to the product of the Jones vector (1.116) with its Hermitian adjoint,

$$I \sim X^2 + Y^2 = E_x E_x^* + E_y E_y^* = \begin{bmatrix} E_x \\ E_y \end{bmatrix}^\dagger \begin{bmatrix} E_x \\ E_y \end{bmatrix} = [E_x^* \quad E_y^*] \begin{bmatrix} E_x \\ E_y \end{bmatrix}. \tag{1.117}$$

The dagger denotes the Hermitian adjoint of a matrix, defined as the complex conjugate of the transposed matrix. The asterisk is the symbol for complex conjugation; the transpose of a column vector is a row vector. Since we are not interested in absolute intensities, the Jones vectors are usually normalized, $E_x E_x^* + E_y E_y^* = 1$.

Two Jones vectors differing merely by a multiplicative factor $\exp(i\delta)$, δ real, describe the same polarization state; in fact, this multiplication changes only the non–observable absolute phases of E_x and E_y by the same amount. We can therefore use the angles of Eq. (1.90) in the

following two equivalent normalized Jones vectors of a general elliptic polarization,

$$\begin{bmatrix} \sin\psi_{xy}\exp(i\Delta_{xy}) \\ \cos\psi_{xy} \end{bmatrix}, \begin{bmatrix} \sin\psi_{xy} \\ \cos\psi_{xy}\exp(-i\Delta_{xy}) \end{bmatrix}. \tag{1.118}$$

The pairs of orthogonal polarizations defined in Section 1.3.3 are represented by the Jones vectors satisfying the usual condition of orthogonality,

$$\begin{bmatrix} E_x \\ E_y \end{bmatrix}^{\dagger} \begin{bmatrix} E'_x \\ E'_y \end{bmatrix} = [E_x^* \quad E_y^*] \begin{bmatrix} E'_x \\ E'_y \end{bmatrix} = 0. \tag{1.119}$$

For example, the two orthogonal states of linear polarization along the x and y directions are

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{1.120}$$

More generally, the orthogonal linear polarizations with the first having its direction of oscillations inclined to the x axis by the angle α are

$$\begin{bmatrix} \cos\alpha \\ \sin\alpha \end{bmatrix}, \begin{bmatrix} -\sin\alpha \\ \cos\alpha \end{bmatrix}. \tag{1.121}$$

Another pair of orthogonal polarizations are the right– and left–handed circular polarizations,

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ i \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} i \\ 1 \end{bmatrix}. \tag{1.122}$$

Finally, the following two normalized Jones vectors,

$$\frac{1}{\sqrt{1+|\chi|^2}}\begin{bmatrix} 1 \\ \chi \end{bmatrix}, \frac{1}{\sqrt{1+|\chi|^2}}\begin{bmatrix} -\chi^* \\ 1 \end{bmatrix}, \tag{1.123}$$

represent two orthogonal elliptic polarizations for any complex number χ. The first of these vectors represents the linear polarization along x and y for $\chi = 0$ and $|\chi| \to \infty$, respectively. The state of polarization described by the second Jones vector of Eq. (1.123) can be represented equivalently by the following normalized vector,

$$\sqrt{1+1/|\chi|^2}\begin{bmatrix} 1 \\ -1/\chi^* \end{bmatrix}. \tag{1.124}$$

The vector representation of the orthogonal states is evidently equivalent to that found in Section 1.3.4, see Eq. (1.105).

The formalism of Jones vectors shows very directly how a particular polarization state is formed as a superposition of a chosen pair of orthogonal states. For example, the first of the vectors of Eq. (1.123) is the superposition

$$\frac{1}{\sqrt{1+|\chi|^2}}\begin{bmatrix}1\\ \chi\end{bmatrix} = c_x\begin{bmatrix}1\\ 0\end{bmatrix} + c_y\begin{bmatrix}0\\ 1\end{bmatrix},\ c_x = \frac{1}{\sqrt{1+|\chi|^2}},\ c_y = \frac{\chi}{\sqrt{1+|\chi|^2}} \tag{1.125}$$

of the pair of linear polarizations of Eq. (1.120). Similarly, the same vector can be written as

$$\begin{aligned}\frac{1}{\sqrt{1+|\chi|^2}}\begin{bmatrix}1\\ \chi\end{bmatrix} &= \frac{c_+}{\sqrt{2}}\begin{bmatrix}1\\ i\end{bmatrix} + \frac{c_-}{\sqrt{2}}\begin{bmatrix}i\\ 1\end{bmatrix},\\ c_+ &= \frac{1-i\chi}{\sqrt{2(1+|\chi|^2)}},\ c_- = \frac{\chi - i}{\sqrt{2(1+|\chi|^2)}},\end{aligned} \tag{1.126}$$

i.e., as the superposition of the two circular polarizations of Eq. (1.122) with the weight coefficients c_+ and c_-.

In the above superpositions, there are only two relevant real numbers describing the polarization state; the two can be chosen as the real and imaginary parts of the complex number χ. Equivalently, the pair of (real) ellipsometric angles (ψ_{xy},Δ_{xy}) of Eq. (1.118) describes uniquely the state of polarization. As explained in Section 1.3.1, another convenient pair of real numbers is formed by the azimuth and ellipticity angles of the polarization ellipse. Let us repeat the reasons for the reduction of the four real numbers entering the vector (1.116) to the relevant number of only two real parameters:

- The information on light intensity is skipped by using normalized Jones vectors. The intensity is not related to the state of polarization.
- The absolute phase of the vibrations is irrelevant; both components of the Jones vectors can be multiplied simultaneously by a complex number of unit modulus, leaving observable quantities of polarized vibrations unchanged.

The information on the polarization state is therefore contained solely in the *direction* of the Jones vector in the abstract space of two ordered complex numbers of Eq. (1.116). A Jones vector multiplied by any non–zero complex number describes the same polarization state.

1.4.2 Jones Matrices

When light passes through, or is reflected from, a polarizing optical element, its polarization state changes. We assume only linear changes from initial field amplitudes forming the Jones vector of Eq. (1.116) to the amplitudes resulting from the action of the element,

$$E'_\xi = J_{11}E_x + J_{12}E_y \, , \, E'_\eta = J_{21}E_x + J_{22}E_y, \tag{1.127}$$

or, the same transformation written down in the convenient matrix form,

$$\begin{bmatrix} E'_\xi \\ E'_\eta \end{bmatrix} = \hat{J}\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}\begin{bmatrix} E_x \\ E_y \end{bmatrix}. \tag{1.128}$$

The modified field amplitudes are, in general, related to another coordinate system ξ–η chosen in the plane of vibrations. The 2×2 matrix $\hat{J}$ of Eq. (1.128) is called *Jones matrix* of the element. It should always be remembered that Jones matrices are related to the particular coordinate systems used to specify the orientations of the incident and transformed light beams, and the optical element.

One of basic transformations is just a *rotation* of the coordinate system used to describe a given vibration; the two Jones vectors are related by

$$\begin{aligned} \begin{bmatrix} E_\xi \\ E_\eta \end{bmatrix} &= \hat{R}(\alpha)\begin{bmatrix} E_x \\ E_y \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} E_x \\ E_y \end{bmatrix}, \\ \begin{bmatrix} E_x \\ E_y \end{bmatrix} &= \hat{R}(-\alpha)\begin{bmatrix} E_\xi \\ E_\eta \end{bmatrix} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} E_\xi \\ E_\eta \end{bmatrix}. \end{aligned} \tag{1.129}$$

Here the ξ–η coordinate system results from a counter–clockwise rotation of the x and y axes by the angle α about the common z axis. The 2×2 matrices in Eqs. (1.129) are inverse one to another; they represent inverse operations of the rotations by the angle α and $-\alpha$, respectively.

It is useful to find the transformation of the Jones matrix of Eq. (1.128), resulting from the rotation of both coordinate systems involved in its definition. Let ξ'–η' be rotated in the positive sense by the angle α from the ξ–η axes, and x'–y' rotated the same way from x–y. The Jones matrix $\hat{J}'$ in primed coordinates links the corresponding field amplitudes,

$$\begin{bmatrix} E'_{\xi'} \\ E'_{\eta'} \end{bmatrix} = \hat{J}'\begin{bmatrix} E_{x'} \\ E_{y'} \end{bmatrix}. \tag{1.130}$$

The Jones vectors written down in the initial and primed coordinates are related by the rotation matrix of Eq. (1.129)

$$\begin{bmatrix} E'_{\xi'} \\ E'_{\eta'} \end{bmatrix} = \hat{R}(\alpha)\begin{bmatrix} E'_{\xi} \\ E'_{\eta} \end{bmatrix}, \quad \begin{bmatrix} E_{x'} \\ E_{y'} \end{bmatrix} = \hat{R}(\alpha)\begin{bmatrix} E_{x} \\ E_{y} \end{bmatrix}. \tag{1.131}$$

Inserting these vectors into Eq. (1.130), multiplying both sides by $\hat{R}(-\alpha)$ (which is the inverse of $\hat{R}(\alpha)$), and recalling the definition of Eq. (1.128), we obtain

$$\hat{J}' = \hat{R}(\alpha)\hat{J}\hat{R}(-\alpha)\ , \hat{J} = \hat{R}(-\alpha)\hat{J}'\hat{R}(\alpha). \tag{1.132}$$

A cascade of m polarizing elements transforms an input Jones vector into the output vector

$$\begin{bmatrix} E_{\xi}^{out} \\ E_{\eta}^{out} \end{bmatrix} = \hat{J}_m \ldots \hat{J}_2\hat{J}_1\begin{bmatrix} E_{x}^{in} \\ E_{y}^{in} \end{bmatrix}. \tag{1.133}$$

The form of individual Jones matrices in the matrix product describing the combined action of the cascade depend, of course, on the orientation of the individual polarizing elements with respect to the coordinate systems used.

1.4.2.1 Linear Polarizer

A linear polarizer is an element that modifies any polarization state to the state of linear polarization. The transmitted or reflected wave is polarized linearly in the direction of the *polarization axis* of the polarizer. For the polarization axis oriented parallel to ξ, the Jones matrix of Eq. (1.128) is

$$\hat{J}_{P0} = t_p \exp(i\delta_P)\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \tag{1.134}$$

where a transmission–type polarizer is assumed, with coinciding ξ–η and x–y coordinate systems. As far as only the polarization states are concerned, the attenuation $t_P \leq 1$ and phase shift δ_P of the polarizer are irrelevant. A polarizer with no attenuation ($t_P = 1$) and zero phase shift is called *ideal*. The linear polarization perpendicular to the transmission axis is completely absent at the output of the linear polarizer,

$$\begin{bmatrix} E'_{x} \\ E'_{y} \end{bmatrix} = \hat{J}_{P0}\begin{bmatrix} E_{x} \\ E_{y} \end{bmatrix} = t_p \exp(i\delta_P)E_x\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{1.135}$$

giving $E'_y = 0$.

For the polarization axis rotated counter–clockwise by the angle α about z axis, the Jones matrix of the ideal polarizer, in the coinciding ξ–η and x–y coordinate systems, becomes

$$\hat{J}_{P\alpha} = \hat{R}(-\alpha) \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \hat{R}(\alpha) = \begin{bmatrix} \cos^2\alpha & \cos\alpha \sin\alpha \\ \cos\alpha \sin\alpha & \sin^2\alpha \end{bmatrix}. \tag{1.136}$$

Note that the transmission axis of the polarizer has to be rotated by $-\alpha$ to be aligned parallel to the x axis; this allows us to use the transformation of Eq. (1.132) and the simple Jones matrix (1.134) to obtain the last result.

Assume the elliptically polarized wave is represented by the first vector of Eq. (1.123), i.e., characterized by the complex value of χ in the x–y coordinate system. After passing through the ideal linear polarizer with the transmission axis inclined by α from the x axis, the transformed vector is

$$\frac{1}{\sqrt{1 + |\chi|^2}} \hat{J}_{P\alpha} \begin{bmatrix} 1 \\ \chi \end{bmatrix} = \frac{1}{\sqrt{1 + |\chi|^2}} \begin{bmatrix} \cos^2\alpha + \chi \cos\alpha \sin\alpha \\ \cos\alpha \sin\alpha + \chi \sin^2\alpha \end{bmatrix}. \tag{1.137}$$

This vector need not be normalized, and contains also information on intensity. According Eq. (1.117), the intensity of the transmitted wave is proportional to

$$\begin{aligned} &\frac{|\cos^2\alpha + \chi \cos\alpha \sin\alpha|^2 + |\cos\alpha \sin\alpha + \chi \sin^2\alpha|^2}{1 + |\chi|^2} \\ &= \frac{\cos^2\alpha + 2\mathrm{Re}(\chi) \cos\alpha \sin\alpha + |\chi|^2 \sin^2\alpha}{1 + |\chi|^2}. \end{aligned} \tag{1.138}$$

After a minor rearrangement we arrive quickly at the light intensity

$$I \sim \frac{1}{2} + \frac{1 - |\chi|^2}{2(1 + |\chi|^2)} \cos 2\alpha + \frac{\mathrm{Re}(\chi)}{(1 + |\chi|^2)} \sin 2\alpha. \tag{1.139}$$

This is the result of Eq. (1.112) that has been obtained from the time–averaging procedure of the Poynting vector.

1.4.2.2 Elliptic Polarizer

An elliptic polarizer transforms any input polarization state into a given elliptic polarization. Let us specify the transmitted wave by the normalized Jones vector

$$\frac{1}{\sqrt{1 + |\chi_{EP}|^2}} \begin{bmatrix} 1 \\ \chi_{EP} \end{bmatrix}. \tag{1.140}$$

Here we assume that the polarization state emerging from the polarizer is represented by the complex value χ_{EP} in a selected coordinate system x–y. An ideal elliptic polarizer passes the transmitted wave with no attenuation, and extinguishes completely the orthogonal polarization,

$$\frac{1}{\sqrt{1 + |\chi_{EP}|^2}} \begin{bmatrix} -\chi_{EP}^* \\ 1 \end{bmatrix}. \tag{1.141}$$

This action is accomplished with the Jones matrix

$$\hat{J}_{EP} = \frac{1}{1 + |\chi_{EP}|^2} \begin{bmatrix} 1 & \chi_{EP}^* \\ \chi_{EP} & |\chi_{EP}|^2 \end{bmatrix}, \tag{1.142}$$

defined in the same x–y coordinate system as χ_{EP}. As expected this Hermitian matrix simplifies to

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \text{ or } \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \tag{1.143}$$

for $\chi_{EP} = 0$ or $|\chi_{EP}| \to \infty$, respectively. In fact, the two matrices describe the ideal linear polarizers with the transmission axis parallel to x or y.

Real linear polarizers may actually be slightly elliptic. It is therefore useful to express the Jones matrix of Eq. (1.142) in terms of the azimuth, ϑ, and ellipticity, γ, angles of the transmitted polarization χ_{EP}. Using Eq. (1.101), the sought matrix is

$$\hat{J}_{EP} = \frac{1 + \tan^2\vartheta \tan^2\gamma}{(1 + \tan^2\vartheta)(1 + \tan^2\gamma)} \begin{bmatrix} 1 & \dfrac{\tan\vartheta - i\tan\gamma}{1 + i\tan\vartheta\tan\gamma} \\ \dfrac{\tan\vartheta + i\tan\gamma}{1 - i\tan\vartheta\tan\gamma} & \dfrac{\tan^2\vartheta + \tan^2\gamma}{1 + \tan^2\vartheta\tan^2\gamma} \end{bmatrix}. \tag{1.144}$$

Note that the azimuth ϑ is the angle between the coordinate axis x and the major axis of the transmitted ellipse. For a small ellipticity $\tan\gamma$ and the magnitude of the azimuth $|\vartheta|$ not too close to $\pi/2$, this result can be expressed as

$$\hat{J}_{EP} \approx \begin{bmatrix} \cos^2\vartheta & \cos\vartheta(\sin\vartheta - i\gamma) \\ \cos\vartheta(\sin\vartheta + i\gamma) & \sin^2\vartheta \end{bmatrix} \text{ for } \tan\gamma \approx \gamma \ll 1. \tag{1.145}$$

1.4.2.3 Linear Retarder

The most important action of a linear retarder is a different phase shift of the linearly polarized waves with the orientations along its *fast* and *slow* axes. An ideal retarder does not attenuate any of these linear polarizations; its Jones matrix is

$$\hat{J}_R = \begin{bmatrix} 1 & 0 \\ 0 & \exp(-i\delta_R) \end{bmatrix} \tag{1.146}$$

for the fast axis parallel to x, and the phase of the fast–axis wave advancing the slow–axis wave by δ_R. For the fast axis rotated by the angle α about z axis, the Jones matrix transforms as

$$\begin{aligned} \hat{J}_{R\alpha} &= \hat{R}(-\alpha) \begin{bmatrix} 1 & 0 \\ 0 & \exp(-i\delta_R) \end{bmatrix} \hat{R}(\alpha) \\ &= \exp\left(\frac{-i\delta_R}{2}\right) \begin{bmatrix} \cos\left(\frac{\delta_R}{2}\right) + i\sin\left(\frac{\delta_R}{2}\right)\cos 2\alpha & i\sin\left(\frac{\delta_R}{2}\right)\sin 2\alpha \\ i\sin\left(\frac{\delta_R}{2}\right)\sin 2\alpha & \cos\left(\frac{\delta_R}{2}\right) - i\sin\left(\frac{\delta_R}{2}\right)\cos 2\alpha \end{bmatrix}. \end{aligned} \tag{1.147}$$

An ideal quarter–wave retarder has the phase shift $\delta_R = \pi/2$. Its Jones matrix results from Eq. (1.147) as

$$\hat{J}_{Q\alpha} = \frac{1-i}{2} \begin{bmatrix} 1 + i\cos 2\alpha & i\sin 2\alpha \\ i\sin 2\alpha & 1 - i\cos 2\alpha \end{bmatrix}, \tag{1.148}$$

where α is the azimuth of its fast axis. For the azimuths of $+\pi/4$ and $-\pi/4$, the Jones matrices are

$$\hat{J}_{Q+} = \frac{1-i}{2} \begin{bmatrix} 1 & i \\ i & 1 \end{bmatrix}, \qquad \hat{J}_{Q-} = \frac{1-i}{2} \begin{bmatrix} 1 & -i \\ -i & 1 \end{bmatrix}, \tag{1.149}$$

respectively. One or both of these two symmetric positions of a quarter–wave retarder are often used in ellipsometric setups.

Note that an ideal circular polarizer can be achieved as an ideal linear polarizer followed by an ideal quarter–wave retarder with its fast axis inclined by $+\pi/4$ (right–handed) or by $-\pi/4$ (left–handed) with respect to the transmission axis of the polarizer.

1.4.2.4 Isotropic Reflecting Surface

The polarization state of a light wave is modified by the reflection on a planar surface. The coordinate systems in the plane of polarization of the incident and reflected waves must be different for oblique incidence. Even in the case of normal incidence (i.e., for the wavevectors of the incident and reflected waves perpendicular to the surface), caution is required in the choice of the coordinate systems. The mirror reflection changes the handedness (parity) of the polarization state. For an idealized normal reflection on a perfect conductor, the electric and magnetic intensities vanish on the surface due to the infinite conductivity. Consequently, the amplitudes of incident and reflected waves are of the same magnitude but opposite direction, irrespective of the direction of polarization in the surface (x–y) plane. We can choose the x and y axes for the reflected beam parallel to those for the incident beam; however, one of them must be of opposite sense to provide the right–handed x–y–z system with the positive sense of the z axis in the direction of propagation. Thus, the Jones matrix of the reflection on a perfect conductor is

$$\hat{J}_{PS} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} = -\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{1.150}$$

Since the absolute phase is irrelevant, both $\hat{J}_{PS}$ and $-\hat{J}_{PS}$ describe the same change of the polarization upon reflection. Either of these matrices can be called a reflection matrix.

Assume x and ξ directions in the polarization planes of the incident and reflected wave parallel to the plane of incidence, with y and η directions perpendicular to it. Then, the Jones matrix of an isotropic surface is given by

$$\hat{J}_S = \begin{bmatrix} -r_x \exp(i\delta_x) & 0 \\ 0 & r_y \exp(i\delta_y) \end{bmatrix} = r_y \exp(i\delta_y) \begin{bmatrix} -\dfrac{r_x}{r_y} \exp[i(\delta_x - \delta_y)] & 0 \\ 0 & 1 \end{bmatrix}, \tag{1.151}$$

where r_x and r_y are the magnitudes of amplitude reflectivities, δ_x and δ_y are the corresponding phase shifts.

1.4.3 Quantum Mechanical Description, Partial Polarization

The properties of polarized light are best described using the language of quantum theory. In fact, polarization of photons is often used as *the* illustration of quantum mechanical concepts, see, e.g., ref. 14 and

ref. 8. The discussion of Sections 1.4.1 and 1.4.2 can be reformulated straightforwardly in terms of the state vectors, their superposition, and linear operations on them.

Let us introduce the symbols of *ket vectors* describing quantum mechanical states and their *representation* in the *x–y* coordinate system. Some of the examples given in Section 1.4.1 are

ket vector	x–y representation	description
$\|x\rangle$	$\begin{bmatrix}1\\0\end{bmatrix}$	linear polarization along x
$\|y\rangle$	$\begin{bmatrix}0\\1\end{bmatrix}$	linear polarization along y
$\|R\rangle$	$(1/\sqrt{2})\begin{bmatrix}1\\i\end{bmatrix}$	right-handed circular polarization
$\|L\rangle$	$(1/\sqrt{2})\begin{bmatrix}i\\1\end{bmatrix}$	left-handed circular polarization
$\|\chi\rangle$	$(1/\sqrt{1+\|\chi\|^2})\begin{bmatrix}1\\\chi\end{bmatrix}$	elliptic polarization represented by χ.

(1.152)

The ket vectors are identified with the normalized Jones vectors. We use a suitable symbol inside the delimiters of the ket symbol, $|\cdots\rangle$, to point to different polarization states. The corresponding *bra vectors* are represented by the Hermitian adjoints of the column vectors. For example, the last polarization state of Eq. (1.152), and the orthogonal elliptic polarization, are described by

bra vector	x–y representation	description
$\langle\chi\|$	$(1/\sqrt{1+\|\chi\|^2})\,[1 \quad \chi^*]$	elliptic polarization represented by χ
$\langle -1/\chi^* \|$	$\sqrt{1+1/\|\chi\|^2}\,[1 \quad -1/\chi]$	orthogonal to the above,

(1.153)

as seen from Eqs. (1.123) and (1.124).

Any ket vector is a linear combination of two basis kets. The Hilbert spaces of ket and bra vectors are linked by scalar (inner) products $\langle\chi|\chi\rangle$ formed, in a particular representation, following the usual rules of matrix multiplications. A very convenient way of handling the ket and bra vectors is provided by introducing the outer product of two state vectors, $|\chi_1\rangle\langle\chi_2|$. This is a linear operator projecting any ket vector on $|\chi_1\rangle$ and any bra vector on $\langle\chi_2|$. In particular, the outer product $|\chi\rangle\langle\chi|$ is

a Hermitian operator describing the action of an ideal polarizer with the output state $|\chi\rangle$.

A *mixed* state of polarization, or statistical mixture, is not represented by a state vector. Instead, its representation is the *density operator*

$$\hat{D}_{\chi,w_\chi} = w_\chi|\chi\rangle\langle\chi| + (1 - w_\chi)|-1/\chi^*\rangle\langle-1/\chi^*|, \tag{1.154}$$

where $w_\chi \in \langle 0,1\rangle$ is the probability of finding the pure state $|\chi\rangle$. The density operator is normalized to the unit trace,

$$Tr(\hat{D}_{\chi,w_\chi}) = \langle\chi|\hat{D}_{\chi,w_\chi}|\chi\rangle + \langle-1/\chi^*|\hat{D}_{\chi,w_\chi}|-1/\chi^*\rangle = 1. \tag{1.155}$$

A limiting case is the totally unpolarized state obtained for $w_\chi = 1/2$,

$$\hat{D}_{unpol} = \frac{1}{2}(|\chi\rangle\langle\chi| + |-1/\chi^*\rangle\langle-1/\chi^*|) = \frac{1}{2}\hat{I}, \tag{1.156}$$

where $\hat{I}$ is unit operator. It results from the incoherent addition of any two orthogonal states with the equal weight of $1/2$.

In the context of ellipsometry, an important issue is the completeness of characterization of individual photons taking part in the measurements. First, they are usually not of the exact same energy; in other words, the beam is not strictly monochromatic. Further, they do not interact in exactly the same way with the optical elements and the sample in an ellipsometer; roughly speaking, individual photons could be distinguished according to the direction of propagation, and also a fraction of the sample surface involved in the reflection process. Thus, the complete description requires a more general density matrix

$$\hat{D}_{\chi,w_\chi;\xi,w_\xi} = \sum_\xi w_\xi|\xi\rangle\hat{D}_{\chi,w_\chi}\langle\xi|, \tag{1.157}$$

where ξ denotes the set of values characterizing the state except for the polarization, and w_ξ are the probabilities of their occurrence. If the quantities ξ are not measured, a measurement of the polarization state reports the *reduced* density matrix

$$\hat{D}_{\chi,w_\chi} = \sum_\xi w_\xi\langle\xi|\hat{D}_{\chi,w_\chi;\xi,w_\xi}|\xi\rangle. \tag{1.158}$$

The measured (reduced) state of partial polarization is completely characterized by three real numbers, two of them giving the complex value of χ. The third is the probability w_χ. However, the remaining degrees of freedom were involved in the formation of the reduced density matrix and have to be taken into account in predicting the results of measurements.

1.4.4 Stokes Vectors

As shown in the previous section, all possible states of polarization, including partial polarization, can be represented by a set of three real numbers. They can be conveniently complemented by the light intensity, giving the total of four real, experimentally accessible quantities, called Stokes parameters.

The Stokes parameters are defined using a fixed x–y coordinate system in the plane of polarization. Each of the parameters has the dimension of intensity, and a fairly simple interpretation in ellipsometric measurements. Let us denote by I_0 the total intensity of an analyzed beam. Further, let $I_x, I_y, I_{\pi/4,}$ and $I_{-\pi/4}$ denote the intensities passed through an ideal linear polarizer aligned at the angles of 0, $\pi/2$, $\pi/4$, and $-\pi/4$, respectively. Finally, let I_R and I_L be the intensities transmitted by ideal right– and left–circular polarizers, respectively. The four Stokes parameters, grouped in the column vector $\mathbf{S}$, can be defined as

$$\mathbf{S} = \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} I_0 \\ I_x - I_y \\ I_{\pi/4} - I_{-\pi/4} \\ I_R - I_L \end{bmatrix}. \tag{1.159}$$

Note that $I_0 = I_x + I_y = I_{\pi/4} + I_{-\pi/4} = I_R + I_L$.

The Stokes vectors can be written as rows (i.e., transposed columns) in order to save space. For unpolarized light all three differences of intensities in Eq. (1.159) vanish, and the row Stokes vector is $\mathbf{S}^T_{unpol} = [S_0 \ 0 \ 0 \ 0]$. On the other hand, a totally elliptically polarized beam, characterized by the ellipsometric angles (ψ,Δ), or the azimuth and ellipticity angles (ϑ,γ), or the complex number $\chi = \chi_r + i\chi_i$, has the Stokes vector

$$\mathbf{S}_\chi = S_0 \begin{bmatrix} 1 \\ -\cos 2\psi \\ \sin 2\psi \cos\Delta \\ -\sin 2\psi \sin\Delta \end{bmatrix} = S_0 \begin{bmatrix} 1 \\ \cos 2\gamma \cos 2\vartheta \\ \cos 2\gamma \sin 2\vartheta \\ \sin 2\gamma \end{bmatrix} = S_0 \begin{bmatrix} 1 \\ \dfrac{1 - \chi_r^2 - \chi_i^2}{1 + \chi_r^2 + \chi_i^2} \\ \dfrac{2\chi_r}{1 + \chi_r^2 + \chi_i^2} \\ \dfrac{2\chi_i}{1 + \chi_r^2 + \chi_i^2} \end{bmatrix}. \tag{1.160}$$

This result is obtained by substitutions of the intensities for corresponding azimuths into the definition (1.159) of the Stokes vector. For

the elliptic polarization orthogonal to that of Eq. (1.160), the Stokes vector

$$\mathbf{S}_{-1/\chi^*} = S_0 \begin{bmatrix} 1 \\ -\cos 2\gamma \cos 2\vartheta \\ -\cos 2\gamma \sin 2\vartheta \\ -\sin 2\gamma \end{bmatrix} \tag{1.161}$$

changes sign of the S_1, S_2 and S_3 components; this is easily seen from the orthogonality of the azimuths, and the change of the sign of the ellipticity angle (see Section 1.3.3). Note that the Stokes vectors of the orthogonal polarization states are orthogonal in the usual sense, i.e., their scalar (inner) product vanishes,

$$\mathbf{S}_{\chi}^{T}\mathbf{S}_{-1/\chi^*} = S_0^2[1-\cos^2 2\gamma(\cos^2 2\vartheta + \sin^2 2\vartheta) - \sin^2 2\gamma] = 0. \tag{1.162}$$

The Stokes vectors of linear and circular polarizations are

$$\mathbf{S}_{lin} = S_0 \begin{bmatrix} 1 \\ \cos 2\vartheta \\ \sin 2\vartheta \\ 0 \end{bmatrix}, \qquad \mathbf{S}_{circ} = S_0 \begin{bmatrix} 1 \\ 0 \\ 0 \\ \pm 1 \end{bmatrix}, \tag{1.163}$$

where ϑ is the azimuth of the linear polarization, and the plus and minus signs in $\mathbf{S}_{circ}$ represent the right– and left–circular polarizations, respectively.

Only two real parameters describing the state of polarization are involved in any of the three forms of the Stokes vector (1.160), due to the independence of the intensity S_0. For the totally polarized light,

$$S_0^2 = S_1^2 + S_2^2 + S_3^2 \quad \text{or} \quad (S_3/S_0)^2 = 1 - (S_1/S_0)^2 - (S_2/S_0)^2 \tag{1.164}$$

results from Eq. (1.160), demonstrating the interdependence of the three *relative* intensities. Obviously, we have some redundance in the six intensities used to define the Stokes vector; measuring three suitably chosen intensities must be sufficient to retrieve the polarization state. Another way of treating the case of complete elliptic polarization is the following: we have to find two ideal elliptic polarizers, one that passes and another that stops completely the analyzed beam. They are perpendicular to each other, and either that of maximum (total), or minimum (zero), registered intensity specifies the sought polarization state. Note that the scheme of intensity measurements provides the necessary information.

The general case of a partially polarized beam is more complex. We assume each of the incoming photons to be in the same *mixed state*; we only have the probability w_χ of being passed through the corresponding aligned elliptic polarizer, or the probability $1-w_\chi$ of being stopped by it. The latter occurs whenever the incoming photon happens to be in the orthogonal polarization state $-1/\chi^*$. Thus, the measured signals are composed of two contributions, $\mathbf{S}_\chi$ and $\mathbf{S}_{-1/\chi^*}$, weighted by the probabilities, w_χ and $1-w_\chi$. Using Eqs. (1.160) and (1.161), the resulting Stokes vector can be put in the form

$$\mathbf{S}_{\chi,part} = w_\chi \mathbf{S}_\chi + (1 - w_\chi)\mathbf{S}_{-1/\chi^*}$$

$$= S_0(2w_\chi - 1)\begin{bmatrix} 1 \\ \cos 2\gamma \cos 2\vartheta \\ \cos 2\gamma \sin 2\vartheta \\ \sin 2\gamma \end{bmatrix} + S_0 2(1 - w_\chi)\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{for } w_\chi \in \langle 1/2,\, 1\rangle. \tag{1.165}$$

We have limited the range of probabilities in order to obtain the simple form of the result. The probabilities w_χ from zero to $1/2$ are actually included in Eq. (1.165) when used for the orthogonal polarization, $-1/\chi^*$. The statistical mixture of the two orthogonal polarization states behaves as the incoherent addition of the fully polarized and unpolarized waves. Inserting $w_\chi = 1/2$ into Eq. (1.165), we obtain the Stokes vector of unpolarized light; it should be noted that this result is independent of the basis used to define the unique mixed state of maximum entropy. Observable results are the same for equal mixtures of any pair of orthogonal polarizations.

The state of partial polarization described by the Stokes vector of Eq. (1.165) is defined by three real numbers. The real quantity to be added to the two real characteristics of the limiting elliptic polarization can be the *degree of polarization*. It is defined as the ratio of the intensity of the polarized part to the total intensity,

$$P_\chi = \frac{\sqrt{S_1^2 + S_2^2 + S_3^2}}{S_0} = 2w_\chi - 1. \tag{1.166}$$

The degree of polarization varies from zero for $w_\chi = 1/2$ to unity for the totally polarized light of $w_\chi = 1$. In place of Eq. (1.164) the Stokes parameters satisfy in general the following inequality,

$$S_0^2 \geq S_1^2 + S_2^2 + S_3^2, \tag{1.167}$$

where the equality holds for the totally polarized light, $P_\chi = 1$. Using the degree of polarization, the Stokes vector of Eq. (1.165) can be written as

$$\mathbf{S}_{\chi,part} = S_0[P_\chi \tilde{\mathbf{S}}_\chi + (1 - P_\chi)\tilde{\mathbf{S}}_{unpol}]. \tag{1.168}$$

Here S_0 is the total intensity and the vectors

$$\tilde{\mathbf{S}}_\chi = \begin{bmatrix} 1 \\ \cos 2\gamma \cos 2\vartheta \\ \cos 2\gamma \sin 2\vartheta \\ \sin 2\vartheta \end{bmatrix}, \quad \tilde{\mathbf{S}}_{unpol} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{1.169}$$

represent the directions of the elliptically polarized and unpolarized light, in the four–dimensional space of Stokes vectors.

1.4.5 Mueller Matrices

The Mueller matrix scheme can handle propagation of partially polarized light through optical systems, including possible depolarizing elements. It transforms Stokes vectors by successive multiplications with 4×4 matrices belonging to individual optical elements. In general, the Mueller matrix

$$\begin{bmatrix} S_0' \\ S_1' \\ S_2' \\ S_3' \end{bmatrix} = \hat{M} \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} = \begin{bmatrix} M_{00} & M_{01} & M_{02} & M_{03} \\ M_{10} & M_{11} & M_{12} & M_{13} \\ M_{20} & M_{21} & M_{22} & M_{23} \\ M_{30} & M_{31} & M_{32} & M_{33} \end{bmatrix} \begin{bmatrix} S_0 \\ S_1 \\ S_2 \\ S_3 \end{bmatrix} \tag{1.170}$$

transforms the input Stokes vector to the (primed) output vector. This matrix must preserve the limitation imposed by Eq. (1.167) for the output vector, and for passive devices (not emitting light) also the condition $S_0' \leq S_0$. The scheme is similar to the Jones matrix formalism of Section 1.4.2. In particular, a cascade of polarizing elements transforms an input Stokes vector into the output vector

$$\mathbf{S}_{out} = \hat{M}_m \cdots \hat{M}_2 \hat{M}_1 \mathbf{S}_{inp}. \tag{1.171}$$

The Stokes matrices are built from the corresponding Jones matrices acting on specified pure polarization states, complemented by the probabilities contained in the density matrix of mixed polarization states.

Since the Stokes vectors are defined with respect to a definite coordinate system within the plane of polarization, special Mueller matrices are

needed to change the coordinate systems. A rotation by the angle α described for the Jones vectors by Eqs. (1.129) has its Mueller matrix counterpart given by

$$\hat{R}(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\alpha & \sin 2\alpha & 0 \\ 0 & -\sin 2\alpha & \cos 2\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{1.172}$$

The Mueller matrices in original and rotated (primed) coordinates are

$$\hat{M}' = \hat{R}(\alpha)\hat{M}\hat{R}(-\alpha), \;\; \hat{M} = \hat{R}(-\alpha)\hat{M}'\hat{R}(\alpha). \tag{1.173}$$

Mueller matrices of given optical elements result from the Jones matrices describing the action on pairs of orthogonal polarizations, and from the probabilities of obtaining a pure polarization in a mixed polarization state.

1.4.5.1 Ideal Polarizers

The Mueller matrix describing an ideal elliptic polarizer can be easily found from the following considerations. The matrix should convert any polarization state into the total elliptic polarization of Eq. (1.160), and this state should be passed untouched through the polarizer. This is achieved with the halved outer product, $(1/2)\tilde{\mathbf{S}}_\chi\tilde{\mathbf{S}}_\chi^T$, of the vectors of Eq. (1.169). This is a 4×4 symmetric matrix $\hat{M}_\chi$ with the elements

$$\begin{aligned} &(\hat{M}_\chi)_{kl} = (1/2)S_{k\chi}S_{l\chi}, \quad k,l = 0,\cdots,3, \text{ where} \\ &\quad S_{0\chi} = 1, \; S_{1\chi} = \cos 2\gamma \cos 2\vartheta, \; S_{2\chi} = \cos 2\gamma \sin 2\vartheta, \; S_{3\chi} = \sin 2\gamma. \end{aligned} \tag{1.174}$$

The factor of $1/2$ results from the fact that the squared length of the vector $\mathbf{S}_\chi$ is two, $\mathbf{S}_\chi^T\mathbf{S}_\chi = 2$, and we require $\hat{M}_\chi^2 = \hat{M}_\chi$. The action of an ideal elliptic polarizer can be summarized as follows,

$$\hat{M}_\chi\mathbf{S}_\chi = \mathbf{S}_\chi, \;\; \hat{M}_\chi\mathbf{S}_{-1/\chi^*} = 0, \;\; \hat{M}_\chi\mathbf{S}_{unpol} = (S_0/2)\mathbf{S}_\chi. \tag{1.175}$$

As expected, it passes without attenuation its eigenpolarization, stops completely the orthogonal elliptic polarization, and rejects half of the incoming photons in totally unpolarized beam.

The Mueller matrices of an ideal linear polarizer with its transmission axis parallel to x, parallel to y, and rotated by α from x are

$$\hat{M}_{LPx} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \hat{M}_{LPy} = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \text{ and}$$
$$\hat{M}_{LP\alpha} = \frac{1}{2}\begin{bmatrix} 1 & \cos 2\alpha & \sin 2\alpha & 0 \\ \cos 2\alpha & \cos^2 2\alpha & \sin 2\alpha \cos 2\alpha & 0 \\ \sin 2\alpha & \sin 2\alpha \cos 2\alpha & \sin^2 2\alpha & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{1.176}$$

respectively. The Mueller matrices of an ideal right– and left–circular polarizer are

$$\hat{M}_{RC} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} \text{ and } \hat{M}_{LC} = \frac{1}{2}\begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}. \tag{1.177}$$

1.4.5.2 Partial Polarizers

Let us define a *partial polarizer* as the device transforming any input beam into a fixed state of partial polarization at its output. The desired outgoing partial polarization will be given by the parameters of an elliptic polarization state, χ, and the polarization degree, P_χ. Thus, the product of the corresponding Mueller matrix $\hat{M}_{\chi,part}$ with any input vector must be proportional to the Stokes vector of Eq. (1.168). This is achieved with the outer product,

$$[1/(1 + P_\chi^2)]\,[P_\chi \tilde{\mathbf{S}}_\chi + (1 - P_\chi)\tilde{\mathbf{S}}_{unpol}]^T [P_\chi \tilde{\mathbf{S}}_\chi + (1 - P_\chi)\tilde{\mathbf{S}}_{unpol}]$$

of the vectors of Eq. (1.169). We obtain a symmetric matrix $\hat{M}_{\chi,part}$ with the elements

$$(\hat{M}_{\chi,part})_{kl} = [1/(1 + P_\chi^2)] S_{k\chi} S_{l\chi}\,, \quad k,l = 0,\cdots,3, \quad \text{where}$$
$$S_{0\chi} = 1, \quad S_{1\chi} = P_\chi \cos 2\gamma \cos 2\vartheta, \quad S_{2\chi} = P_\chi \cos 2\gamma \sin 2\vartheta, \quad S_{3\chi} = P_\chi \sin 2\gamma. \tag{1.178}$$

Written explicitly, the Mueller matrix is

$$\hat{M}_{\chi,part} = \left[1/\left(1 + P_\chi^2\right)\right] \times$$
$$\begin{bmatrix} 1 & P_\chi \cos 2\gamma \cos 2\vartheta & P_\chi \cos 2\gamma \sin 2\vartheta & P_\chi \sin 2\gamma \\ P_\chi \cos 2\gamma \cos 2\vartheta & P_\chi^2 \cos^2 2\gamma \cos^2 2\vartheta & P_\chi^2 \cos^2 2\gamma \sin 2\vartheta \cos 2\vartheta & P_\chi^2 \sin 2\gamma \cos 2\gamma \cos 2\vartheta \\ P_\chi \cos 2\gamma \sin 2\vartheta & P_\chi^2 \cos^2 2\gamma \sin 2\vartheta \cos 2\vartheta & P_\chi^2 \cos^2 2\gamma \sin^2 2\vartheta & P_\chi^2 \sin 2\gamma \cos 2\gamma \sin 2\vartheta \\ P_\chi \sin 2\gamma & P_\chi^2 \sin 2\gamma \cos 2\gamma \cos 2\vartheta & P_\chi^2 \sin 2\gamma \cos 2\gamma \sin 2\vartheta & P_\chi^2 \sin^2 2\gamma \end{bmatrix}. \tag{1.179}$$

The expected action of the partial polarizer is confirmed by examining the following simple cases. The result of a partial elliptic polarization of a totally unpolarized input beam is

$$\hat{M}_{\chi,part}\begin{bmatrix} S_0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{S_0}{1+P_\chi^2}\begin{bmatrix} 1 \\ P_\chi \cos 2\gamma \cos 2\vartheta \\ P_\chi \cos 2\gamma \sin 2\vartheta \\ P_\chi \sin 2\gamma \end{bmatrix}$$

$$= \frac{S_0 P_\chi}{1+P_\chi^2}\begin{bmatrix} 1 \\ \cos 2\gamma \cos 2\vartheta \\ \cos 2\gamma \sin 2\vartheta \\ \sin 2\gamma \end{bmatrix} + \frac{S_0(1-P_\chi)}{1+P_\chi^2}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}. \qquad (1.180)$$

We obtain the desired statistical mixture of the fraction $P_\chi/(1+P_\chi^2)$ of the total intensity elliptically polarized; the rest is unpolarized. The degree of polarization is, of course, P_χ. The output intensity, $S_0/(1+P_\chi^2)$, increases from $S_0/2$ for an ideal polarizer ($P_\chi = 1$; half of the incoming photons is rejected) to the total intensity of the incoming beam, S_0, for $P_\chi = 0$. The latter limiting case does not depend on the ellipticity and azimuth angles of the polarizer; this idealized device should be called *ideal depolarizer*. Its Mueller matrix is

$$\hat{M}_{depol} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \qquad (1.181)$$

Another instructive simple case is the action of the partial polarizer on the beam polarized linearly along the x axis,

$$\hat{M}_{\chi,part}\begin{bmatrix} S_0 \\ S_0 \\ 0 \\ 0 \end{bmatrix} = \frac{S_0}{1+P_\chi^2}(1+P_\chi \cos 2\gamma \cos 2\vartheta)\begin{bmatrix} 1 \\ P_\chi \cos 2\gamma \cos 2\vartheta \\ P_\chi \cos 2\gamma \sin 2\vartheta \\ P_\chi \sin 2\gamma \end{bmatrix}. \qquad (1.182)$$

As expected, the resulting state of partial polarization differs from that of Eq. (1.180) in the total intensity only. When rotating the partial polarizer, i.e., changing its azimuth ϑ, the intensity oscillates between

$$I_{\min} = S_0 \frac{1-P_\chi|\cos 2\gamma|}{1+P_\chi^2} \quad \text{and} \quad I_{\max} = S_0 \frac{1+P_\chi|\cos 2\gamma|}{1+P_\chi^2}, \qquad (1.183)$$

unless the two values are the same for a circular polarizer ($\gamma = \pm\pi/4$). The transmitted intensity depends on both polarization degree P_χ and ellipticity angle γ of the polarizer.

The Mueller matrices of a partial linear polarizer with its transmission axis parallel to x and rotated by ϑ from x are

$$\hat{M}_{LPx,part} = \frac{1}{1+P^2}\begin{bmatrix} 1 & P & 0 & 0 \\ P & P^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{1.184}$$

$$\hat{M}_{LP\alpha,part} = \frac{1}{1+P^2}\begin{bmatrix} 1 & P\cos 2\vartheta & P\sin 2\vartheta & 0 \\ P\cos 2\vartheta & P^2\cos^2 2\vartheta & P^2\sin 2\vartheta\cos 2\vartheta & 0 \\ P\sin 2\vartheta & P^2\sin 2\vartheta\cos 2\vartheta & P^2\sin^2 2\vartheta & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The right– and left–circular partial polarizers have the following Mueller matrices,

$$\hat{M}_{RC,part} = \frac{1}{1+P^2}\begin{bmatrix} 1 & 0 & 0 & P \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ P & 0 & 0 & P^2 \end{bmatrix}, \quad \hat{M}_{LC,part} = \frac{1}{1+P^2}\begin{bmatrix} 1 & 0 & 0 & -P \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -P & 0 & 0 & P^2 \end{bmatrix}. \tag{1.185}$$

The definition of the partial polarizer may be completed by specifying a real value of its attenuation, $0 < T \leq 1$, which is the scalar factor multiplying the output Stokes vector $\hat{\mathbf{S}}_{out} = \hat{M}_{\chi,part}\hat{\mathbf{S}}_{inp}$. It should be noted that the definition of partial polarizer, Eqs. (1.178) or (1.179), differs from that used in Ref. 9. The latter assumes different attenuations along principal axes of the polarizer, but no depolarizing action.

1.4.5.3 Linear Retarder

An ideal linear retarder with the phase retardation δ and its fast axis along x, and rotated by α from x, has the Mueller matrices

$$\hat{M}_{Rx} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\delta & \sin\delta \\ 0 & 0 & -\sin\delta & \cos\delta \end{bmatrix} \quad \text{and}$$

$$\hat{M}_{R\alpha} = \tag{1.186}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos^2 2\alpha + \cos\delta\sin^2 2\alpha & \sin 2\alpha\cos 2\alpha(\cos\delta - 1) & \sin\delta\sin 2\alpha \\ 0 & \sin 2\alpha\cos 2\alpha(\cos\delta - 1) & \sin^2 2\alpha + \cos\delta\cos^2 2\alpha & \sin\delta\cos 2\alpha \\ 0 & -\sin\delta\sin 2\alpha & -\sin\delta\cos 2\alpha & \cos\delta \end{bmatrix}.$$

1.4.5.4 Isotropic Reflecting Surface

An ideal isotropic reflecting surface with the ellipsometric angles ψ and Δ defined in the x–y coordinate system of the Stokes vectors has the following Mueller matrix, resulting from the Jones matrix of Eq. (1.151):

$$\hat{M}_S = \frac{|r_x|^2 + |r_y|^2}{2}\begin{bmatrix} 1 & -\cos 2\psi & 0 & 0 \\ -\cos 2\psi & 1 & 0 & 0 \\ 0 & 0 & \sin 2\psi \cos\Delta & \sin 2\psi \sin\Delta \\ 0 & 0 & -\sin 2\psi \sin\Delta & \sin 2\psi \cos\Delta \end{bmatrix}. \tag{1.187}$$

Here $|r_x|^2$ and $|r_y|^2$ are the power (intensity) reflectivities for the p– and s–polarized light, respectively; $R_u = (|r_x|^2 + |r_y|^2)/2$ is the power reflectivity for either unpolarized, circularly polarized, or linearly polarized light with the azimuth of $\pi/4$. Since $|r_x|^2/|r_y|^2 = \tan^2\psi$, the matrix (1.187) can be written as

$$\hat{M}_S = \frac{|r_x|^2}{2\cos^2\psi}\begin{bmatrix} 1 & -N & 0 & 0 \\ -N & 1 & 0 & 0 \\ 0 & 0 & C & S \\ 0 & 0 & -S & C \end{bmatrix} = \frac{|r_y|^2}{2\sin^2\psi}\begin{bmatrix} 1 & -N & 0 & 0 \\ -N & 1 & 0 & 0 \\ 0 & 0 & C & S \\ 0 & 0 & -S & C \end{bmatrix}, \tag{1.188}$$

where $N = \cos 2\psi$, $C = \sin 2\psi \cos\Delta$, $S = \sin 2\psi \sin\Delta$. The reflecting surface of Eq. (1.187) acts as an ideal polarizer aligned along y and x for $\psi = 0$ and $\pi/2$, respectively.

A possible depolarizing action results from the presence of different values of ψ and Δ, related, e.g., to a spread of the optical constants and/or film thicknesses over the illuminated region, or to a spread of the angles of incidence. Let us assume an instructive case of only two pairs (ψ_1,Δ_1) and (ψ_2,Δ_2) involved in the reflection. In addition, we simplify the situation further by assuming equal intensity reflectivities of the linear polarizations along x and y, $|r_x|^2 = |r_y|^2 = R_u$. The latter condition implies the equality $\psi_1 = \psi_2 = \psi$. Let us denote by w_1 the probability that the reflected photon encountered the surface characterized by (ψ_1,Δ_1). Consequently, the second possibility, (ψ_2,Δ_2), occurs with the probability $w_2 = 1 - w_1$. The polarization state of the reflected light is given by the Stokes vector of the incident light and the average Mueller matrix $\hat{M}'_S = w_1\hat{M}^{(1)}_S + w_2\hat{M}^{(2)}_S$; this corresponds to the averaging in the density matrix of Eq. (1.158). The two Mueller matrices $\hat{M}^{(1)}_S$ and $\hat{M}^{(2)}_S$

share the value of $N = \cos 2\psi$ and differ in the remaining two parameters, C and S, due to the difference in the phase shifts Δ. The average Mueller matrix is

$$\hat{M}'_S = R_u \begin{bmatrix} 1 & -N & 0 & 0 \\ -N & 1 & 0 & 0 \\ 0 & 0 & C' & S' \\ 0 & 0 & -S' & C' \end{bmatrix}, \tag{1.189}$$

where

$$\begin{aligned} C' &= \sin 2\psi [w_1 \cos \Delta_1 + (1 - w_1) \cos \Delta_2], \\ S' &= \sin 2\psi [w_1 \sin \Delta_1 + (1 - w_1) \sin \Delta_2]. \end{aligned} \tag{1.190}$$

The light reflected from a surface described by Eq. (1.189) remains totally linearly polarized for the incoming light polarized linearly along x or y. However, for the incoming light beam of unit intensity and linear polarization with the azimuth $\vartheta = \pi/4$, the Stokes vector of reflected light is

$$\hat{M}'_S \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = R_u \begin{bmatrix} 1 \\ -N \\ C' \\ -S' \end{bmatrix}. \tag{1.191}$$

Accordingly, its degree of polarization

$$\begin{aligned} P_{\pi/4} &= \sqrt{N^2 + C'^2 + S'^2} \\ &= \sqrt{\cos^2 2\psi + \sin^2 2\psi [1 + 2w_1(1 - w_1)(\cos(\Delta_2 - \Delta_1) - 1)]} \end{aligned} \tag{1.192}$$

equals one only in the trivial cases of $\Delta_1 = \Delta_2$, or w_1 either vanishing or being equal to one. In other words, the fraction $1 - P_{\pi/4}$ of the incoming intensity becomes unpolarized upon reflection. The depolarization is total for $w_1 = 1/2$ and $\Delta_2 - \Delta_1 = \pi$, which is the expected result: the reflected light is a mixture of equal contributions of any pair of pure orthogonal polarization states.

The above considerations can easily be generalized to the case of a continuous mixture of contributions of different pairs $\psi(\xi)$,$\Delta(\xi)$, and the power reflectivities $R_u(\xi)$, occurring with the probabilities $f(\xi)d\xi$. Here ξ denotes a set of parameters influencing the amplitude reflectivities of the surface; it can cover a multidimensional domain of varying optical constants, film thicknesses, wavelengths, and/or angles of incidence. The

function $f(\xi)$ is the *probability density* of the occurrence of a particular value of ξ. The Mueller matrix can be written as

$$\hat{M}'_S = \langle R_u \rangle \begin{bmatrix} 1 & -\langle N \rangle & 0 & 0 \\ -\langle N \rangle & 1 & 0 & 0 \\ 0 & 0 & \langle C \rangle & \langle S \rangle \\ 0 & 0 & -\langle S \rangle & \langle C \rangle \end{bmatrix}, \tag{1.193}$$

where

$$\begin{aligned} \langle R_u \rangle &= \int R_u(\xi) f(\xi) d\xi, \\ \langle N \rangle &= \frac{\int R_u(\xi) \cos 2\psi(\xi) f(\xi) d\xi}{\langle R_u \rangle}, \\ \langle C \rangle &= \frac{\int R_u(\xi) \sin 2\psi(\xi) \cos \Delta(\xi) f(\xi) d\xi}{\langle R_u \rangle}, \\ \langle S \rangle &= \frac{\int R_u(\xi) \sin 2\psi(\xi) \sin \Delta(\xi) f(\xi) d\xi}{\langle R_u \rangle}. \end{aligned} \tag{1.194}$$

Since any contribution to the resulting Mueller matrix preserves the linear polarizations along x and y, the reflection does not depolarize them. On the other hand, the linearly polarized light with the azimuth $\vartheta = \pm\pi/4$, or a circularly polarized light, becomes partially elliptically polarized with the degree of polarization

$$P_{\pm\pi/4,C} = \sqrt{\langle N \rangle^2 + \langle C \rangle^2 + \langle S \rangle^2}. \tag{1.195}$$

Obviously, the degree of polarization cannot exceed unity; the averaging procedure of Eqs. (1.194) guarantees the fulfillment of this condition for any incident polarization state.

For completely unpolarized incident light, the degree of polarization of the reflected beam is $|\langle N \rangle|$. It reaches unity for a single contributing value of $\psi = 0$ or $\pi/2$ and an arbitrary value of Δ; in this case, the reflected light becomes linearly polarized along x or y.

The action of the ideal linear retarder with its fast and slow axes along x and y, described by Eq. (1.186), is formally the same as that of Eq. (1.187) with $\psi = \pi/4$ and $\Delta = \delta$. Consequently, a possible depolarization by a retarder can be treated along the same lines as that of a reflecting surface. It should be emphasized that (in general) neither the retarder nor the reflecting surface produce a definite pure or partial polarization state independent of the incoming beam.

1.5 Reflection and Transmission of Polarized Light at Planar Interfaces

1.5.1 Matching Plane Waves at a Planar Interface

Let us consider the oblique incidence of a monochromatic optical plane wave on the planar interface between two semi–infinite media as shown schematically in Fig. 1.8. The right–handed coordinate system $x-y-z$ has the y axis pointing to the reader. The scalar complex refractive index changes abruptly from N_1 for $z < 0$ to N_2 for $z > 0$ at the interface, $z = 0$. The light field in both media is assumed to consist of plane waves of the form

$$\vec{E}(\vec{r},t) = \vec{E}_\omega \exp[i(k_x x + k_z z - \omega t)], \tag{1.196}$$

i.e., their wavevectors lie in the x–z plane, which is called *plane of incidence*. This choice is forced by the symmetry with respect to translations in the y direction, with the consequence $k_y = 0$. The plane waves of Eq. (1.196) solve the wave equation for

$$k_x^2 + k_z^2 = \frac{\omega^2}{c^2}\varepsilon(\omega) = \frac{\omega^2}{c^2}N^2(\omega), \tag{1.197}$$

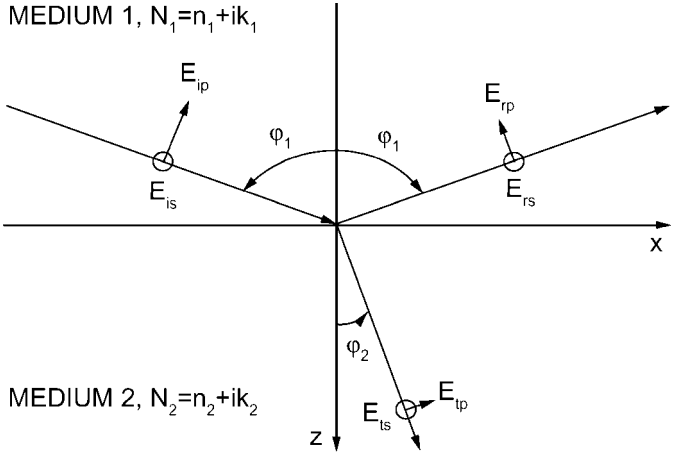


Figure 1.8 Reflection and transmission of a plane wave at the planar interface.

with the complex permitivity $\varepsilon = \varepsilon_1 + i\varepsilon_2 = N^2 = (n + ik)^2$. The total light field in medium 1 is the superposition of the *incident* and *reflected* plane waves

$$\begin{aligned} \vec{E}_i(\vec{r},t) &= \vec{E}_{\omega i} \exp[i(k_{xi}x + k_{zi}z - \omega t)], \\ \vec{E}_r(\vec{r},t) &= \vec{E}_{\omega r} \exp[i(k_{xr}x + k_{zr}z - \omega t)], \quad z < 0. \end{aligned} \tag{1.198}$$

The *transmitted* wave in medium 2 is

$$\vec{E}_t(\vec{r},t) = \vec{E}_{\omega t} \exp[i(k_{xt}x + k_{zt}z - \omega t)], \quad z > 0. \tag{1.199}$$

We define the *angle of incidence* φ_1 by expressing the x– and z–components of the wavevector as

$$k_{xi} = \frac{\omega}{c} N_1 \sin\varphi_1, \quad k_{zi} = \frac{\omega}{c} N_1 \cos\varphi_1. \tag{1.200}$$

Here, $\sin\varphi_1$ and $\cos\varphi_1$ can be complex, although satisfying the condition $\sin^2\varphi_1 + \cos^2\varphi_1 = 1$ which is required by Eq. (1.197). We can still declare $\varphi_1 = \arctan(k_{xi}/k_{zi})$ and use the complex trigonometric functions of complex argument. The latter are defined, for $\xi = \xi_1 + i\xi_2$, ξ_1 and ξ_2 real, as

$$\sin\xi = [\exp(i\xi) - \exp(-i\xi)]/2i\,, \ \cos\xi = [\exp(i\xi) + \exp(-i\xi)]/2. \tag{1.201}$$

Since the tangential components of the plane waves (1.198) and (1.199) match in any point x of the interface $z = 0$, the x–components of the three wavevectors have to coincide,

$$k_{xi} = k_{xr} = k_{xt}\,. \tag{1.202}$$

The first of these equations, together with Eq. (1.197), implies $k_{zr}^2 = k_{zi}^2$. Consequently, we choose $k_{zr} = -k_{zi}$; the reflected wave moves away from the interface, and its wavevector is given by the refractive index and angle of incidence,

$$k_{xr} = \frac{\omega}{c} N_1 \sin\varphi_1, \quad k_{zr} = -\frac{\omega}{c} N_1 \cos\varphi_1. \tag{1.203}$$

In other words, the angle of reflection equals the angle of incidence.

The second of Eqs. (1.202), together with Eq. (1.197), leads to the following relation for the remaining unspecified component of the wavevectors,

$$k_{zt}^2 = \frac{\omega^2}{c^2} N_2^2 \left(1 - \frac{N_1^2}{N_2^2} \sin^2 \varphi_1\right). \tag{1.204}$$

It is convenient to write down the wavevector of the transmitted wave in the same form as that of the incident wave, Eq. (1.200),

$$k_{xt} = \frac{\omega}{c} N_2 \sin \varphi_2, \quad k_{zt} = \frac{\omega}{c} N_2 \cos \varphi_2, \tag{1.205}$$

where $\varphi_2 = \arctan(k_{xt}/k_{zt})$ is the *angle of refraction*. Since $k_{xi} = k_{xt}$, we obtain *Snell's law*

$$N_1 \sin \varphi_1 = N_2 \sin \varphi_2. \tag{1.206}$$

For both media transparent, refractive indices are real, denoted as n_1 and n_2. Further, φ_1 and $\sin \varphi_1$ are also real. The angle of refraction, φ_2, is real for $n_1 \sin \varphi_1 \geq n_2$. This condition is not satisfied for

$$n_1 > n_2\,, \quad \varphi_1 \geq \varphi_c \equiv \arcsin \frac{n_2}{n_1}. \tag{1.207}$$

For fixed values of $n_2 < n_1$ the angle of refraction behaves as follows: with the angle of incidence increasing from zero to the *critical angle* φ_c of Eq. (1.207), it increases from zero to $\pi/2$. With the further increase of the angle of incidence, the real part of φ_2 stays at $\pi/2$ and a non–zero imaginary part appears. Using the Snell's law (1.206) with the real–valued sine function of Eq. (1.201) for $\xi_1 = \pi/2$, we obtain the following result for the angle of refraction,

$$\varphi_2 = \frac{\pi}{2} + i \ln \frac{n_1 \sin \varphi_1 - \sqrt{(n_1 \sin \varphi_1)^2 - n_2^2}}{n_2}, \quad \varphi_1 \geq \varphi_c. \tag{1.208}$$

From Eq. (1.204), the z–component of the transmitted wavevector is imaginary,

$$k_{zt} = i \frac{\omega}{c} \sqrt{(n_1 \sin \varphi_1)^2 - n_2^2}, \quad \varphi_1 \geq \varphi_c. \tag{1.209}$$

The resulting field pattern is an inhomogeneous plane wave with no oscillations in the z direction; it is called an *evanescent* wave. The planes of constant phase in medium 2 are perpendicular to the x axis and move rightwards with the velocity of $c/(n_1 \sin\varphi_1)$. The modulus of the field strength decreases exponentially with z, as obtained from Eqs. (1.199) and (1.209),

$$|\vec{E}_t(\vec{r},t)| = |\vec{E}_{\omega t}| \exp\left(-z\frac{\omega}{c}\sqrt{(n_1 \sin\varphi_1)^2 - n_2^2}\right), \quad \varphi_1 \geq \varphi_c. \tag{1.210}$$

A convenient measure for this decrease is the depth z_e, where the field strength is reduced by the factor of $e \approx 2.71$,

$$z_e = \frac{\lambda}{2\pi\sqrt{(n_1 \sin\varphi_1)^2 - n_2^2}}, \quad \varphi_1 > \varphi_c. \tag{1.211}$$

Here $\lambda = 2\pi c/\omega$ is the wavelength in vacuum. As expected, the penetration depth becomes infinite for $\varphi_1 \to \varphi_c$. It decreases with increasing angle of incidence; the lowest value, obtained for $n_2 \ll n_1$ and $\varphi_1 \to \pi/2$, is

$$z_{e\,\min} = \frac{\lambda}{2\pi n_1}, \tag{1.212}$$

which is the wavelength in medium 1 reduced by the factor of 2π. Since the field vanishes for $z \to \infty$, and medium 2 is non–absorbing, there is no transfer of energy across the interface.

An exponential damping of the field strength of a plane wave occurs in any absorbing medium. For a real refractive index $N_1 = n_1$ and a complex refractive index N_2, the obliquely refracted wave has the following z dependence of the modulus of its amplitude,

$$|\vec{E}_t(\vec{r},t)| = |\vec{E}_{\omega t}| \exp\left(-z\frac{\omega}{c}\,\mathrm{Im}\sqrt{(n_2 + ik_2)^2 - (n_1 \sin\varphi_1)^2}\right). \tag{1.213}$$

At the distance d_e below the interface, the field strength is reduced by the factor of e,

$$d_e = \frac{\lambda}{2\pi}\frac{\sqrt{2}}{\left[\sqrt{(n_2^2 - k_2^2 - n_1^2\sin^2\varphi_1)^2 + 4n_2^2k_2^2} - (n_2^2 - k_2^2 - n_1^2\sin^2\varphi_1)\right]^{1/2}}.$$

. (1.214)

This penetration depth reduces to

$$d_e = \frac{\lambda}{2\pi k_2} \tag{1.215}$$

in the case of normal incidence, $\varphi_1 = 0$.

The significance of the real and imaginary parts of a complex angle of refraction $\varphi_2 = \varphi_{2r} + i\varphi_{2i}$ can be found by examining the angle φ_{2ph} between the normals to the interface and to the planes of constant phase. Namely, the real valued angles are related by

$$\begin{aligned}\tan\varphi_{2ph} &= \frac{n_1 \sin\varphi_1}{n_2 \cos\varphi_{2r} \cosh\varphi_{2i} + k_2 \sin\varphi_{2r} \sinh\varphi_{2i}} \\ &= \frac{n_2^2 + k_2^2}{n_2^2 - k_2^2 \tan^2\varphi_{2r}} \tan\varphi_{2r} = -\frac{n_2^2 + k_2^2}{2n_2k_2} \sinh(2\varphi_{2i}).\end{aligned} \tag{1.216}$$

In the usual case of n_2 larger than n_1, the real and imaginary parts of φ_2 vary fairly strongly with the extinction coefficient k_2. However, the direction φ_{2ph} of the equiphase planes does not deviate dramatically from the value of $\arcsin(n_1 \sin\varphi_1/n_2)$, observed in a non–absorbing medium of the same phase velocity c/n_2.

All results of this paragraph concern only wavevectors. Consequently, they are independent of polarization state of the plane waves. To proceed further, it is necessary to apply the continuity conditions for the tangential (i.e., $x - y$) components of $\vec{E}$ and $\vec{H}$ at the interface. For the plane waves with the wavevector $\vec{k}$ the field intensities are related by $\vec{H} = -(\vec{k} \times \vec{E})/(\omega\mu_0)$, $\vec{E} = (\vec{k} \times \vec{H})/\omega\varepsilon\varepsilon_0$. The cartesian components are

$$H_x = \frac{k_z E_y}{\omega\mu_0}, \quad H_y = \frac{k_x E_z - k_z E_x}{\omega\mu_0}, \quad H_z = \frac{-k_x E_y}{\omega\mu_0}, \tag{1.217}$$

$$E_x = \frac{-k_z H_y}{\omega\varepsilon\varepsilon_0}, \quad E_y = \frac{k_z H_x - k_x H_z}{\omega\varepsilon\varepsilon_0}, \quad E_z = \frac{k_x H_y}{\omega\varepsilon\varepsilon_0}, \tag{1.218}$$

since $k_y = 0$. For the waves linearly polarized perpendicular to the plane of incidence (i.e., $E_x = E_z = 0$), there are only three nonvanishing field components, E_y, H_x, and H_z. This case is called *s*–polarization, from the German word *senkrecht,* meaning perpendicular. When the plane waves are polarized parallel to the plane of incidence (i.e., $E_y = 0$), the only nonvanishing field components are E_x, E_z, and H_y. This case is called *p*–polarization. A general case can be treated as a superposition of the *s*– and *p*–polarizations.

1.5.2 Fresnel Coefficients

1.5.2.1 *s*–Polarization

Let us denote by E_{is}, E_{rs}, and E_{ts} the amplitudes of the incident, reflected, and transmitted waves, respectively. They represent the y components of the vectors $\vec{E}_{\omega i}$, $\vec{E}_{\omega r}$, and $\vec{E}_{\omega t}$ of Eqs. (1.198) and (1.199). Using Eqs. (1.202), the condition of matching the tangential (y) components at the interface $z = 0$ becomes

$$E_{is} + E_{rs} = E_{ts}. \tag{1.219}$$

From Eqs. (1.217), the tangential (x) component of the magnetic intensity vector of the incident wave has the amplitude $k_{zi}E_{is}/(\omega\mu_0)$. Similarly, the amplitudes of the x components of the reflected and transmitted waves are $k_{zr}E_{rs}/(\omega\mu_0)$ and $k_{zt}E_{it}/(\omega\mu_0)$, respectively. With the help of Eqs. (1.203) and (1.205), the condition of matching the tangential components of magnetic field strength at the interface $z = 0$ can be expressed in terms of the electric field strengths as

$$N_1 \cos\varphi_1(E_{is} - E_{rs}) = N_2 \cos\varphi_2 E_{ts}. \tag{1.220}$$

Combined with Eq. (1.219) it leads to the resulting Fresnel reflection and transmission coefficients, which are the ratios of the field amplitudes,

$$r_s \equiv \frac{E_{rs}}{E_{is}} = \frac{N_1 \cos\varphi_1 - N_2 \cos\varphi_2}{N_1 \cos\varphi_1 + N_2 \cos\varphi_2}, \tag{1.221}$$

$$t_s \equiv \frac{E_{ts}}{E_{is}} = \frac{2N_1 \cos\varphi_1}{N_1 \cos\varphi_1 + N_2 \cos\varphi_2}. \tag{1.222}$$

Snell's law, Eq. (1.206), can be used to obtain the following equivalent expressions for the coefficients,

$$r_s = \frac{\sin(\varphi_2 - \varphi_1)}{\sin(\varphi_2 + \varphi_1)} = \frac{N_1 \cos\varphi_1 - \sqrt{N_2^2 - N_1^2 \sin^2\varphi_1}}{N_1 \cos\varphi_1 + \sqrt{N_2^2 - N_1^2 \sin^2\varphi_1}} \equiv \tan\psi_s \exp(i\Delta_s), \tag{1.223}$$

$$t_s = \frac{2 \sin\varphi_2 \cos\varphi_1}{\sin(\varphi_2 + \varphi_1)}. \tag{1.224}$$

The reflection coefficient of Eq. (1.223) is expressed with the help of two angles, ψ_s and Δ_s. This is a convenient way of handling separately the ratio of magnitudes of the reflected and incident field amplitudes, $\tan\psi_s$, and their relative phase, Δ_s.

For *s* polarization the intensity of magnetic field has both tangential and normal components. From Eqs. (1.217) and (1.202) we confirm the continuity of the normal components,

$$H_{zi} + H_{zr} = H_{zt}. \tag{1.225}$$

This is, of course, the general boundary condition for the normal components of magnetic inductance for materials with $\vec{B}_\omega = \mu_0 \vec{H}_\omega$.

1.5.2.2 *p*–Polarization

The Fresnel coefficients are derived similarly to the *s* case. We denote by H_{ip}, H_{rp}, and H_{tp} the amplitudes of the magnetic intensity of the incident, reflected, and transmitted waves, respectively. Using Eqs. (1.202), matching the tangential (y) components of $\vec{H}$ at the interface $z = 0$ requires

$$H_{ip} + H_{rp} = H_{tp}. \tag{1.226}$$

From Eqs. (1.218), the tangential (x) component of the electric intensity vector of the incident wave has the amplitude $E_{xi} = -k_{zi}H_{ip}/(\omega\varepsilon\varepsilon_o)$. Similarly, the amplitudes of the x components of the reflected and transmitted waves are $E_{xr} = -k_{zt}H_{rp}/(\omega\varepsilon\varepsilon_o)$ and $E_{xt} = -k_{zt}H_{tp}/(\omega\varepsilon\varepsilon_o)$, respectively. Using Eqs. (1.203) and (1.205), matching the tangential components of the electric field at the interface $z = 0$ can be expressed in terms of the magnetic field strengths as

$$N_2 \cos\varphi_1 \, (H_{ip} - H_{rp}) = N_1 \cos\varphi_2 H_{tp}. \tag{1.227}$$

Fresnel reflection and transmission coefficients are ratios of the amplitudes of the *electric* field strengths, E_{ip}, E_{rp}, and E_{tp}, of the incident, reflected, and transmitted waves, respectively. These are, however, simply related to the amplitudes of the magnetic field strengths; using Eqs. (1.218) we arrive at

$$r_p \equiv \frac{E_{rp}}{E_{ip}} = \frac{E_{xr}}{E_{xi}} = -\frac{E_{zr}}{E_{zi}} = -\frac{H_{rp}}{H_{ip}}, \quad t_p \equiv \frac{E_{tp}}{E_{ip}} = \frac{N_1}{N_2}\frac{H_{tp}}{H_{ip}}. \tag{1.228}$$

The relative amplitudes of the magnetic field intensities are readily obtained from Eqs. (1.226) and (1.227) leading to the reflection and transmission coefficients

$$r_p = \frac{N_1 \cos\varphi_2 - N_2 \cos\varphi_1}{N_1 \cos\varphi_2 + N_2 \cos\varphi_1} = \frac{N_1/\cos\varphi_1 - N_2/\cos\varphi_2}{N_1/\cos\varphi_1 + N_2/\cos\varphi_2}, \tag{1.229}$$

$$t_p = \frac{2N_1 \cos\varphi_2}{N_1 \cos\varphi_2 + N_2 \cos\varphi_1} = \frac{2N_1/\cos\varphi_1}{N_1/\cos\varphi_1 + N_2/\cos\varphi_2}. \tag{1.230}$$

Equivalent expressions for the Fresnel coefficients are obtained from Snell's law,

$$r_p = \frac{\tan(\varphi_2 - \varphi_1)}{\tan(\varphi_2 + \varphi_1)} = \frac{N_1/\cos\varphi_1 - N_2/\sqrt{1 - (N_1/N_2)^2 \sin^2\varphi_1}}{N_1/\cos\varphi_1 + N_2/\sqrt{1 - (N_1/N_2)^2 \sin^2\varphi_1}} \equiv \tan\psi_p \exp(i\Delta_p), \tag{1.231}$$

$$t_p = \frac{2 \sin\varphi_2 \cos\varphi_1}{\sin(\varphi_2 + \varphi_1) \cos(\varphi_2 - \varphi_1)}. \tag{1.232}$$

The reflection coefficient of Eq. (1.231) is again expressed with the help of two angles, ψ_p and Δ_p, in order to quantify separately the magnitude of the reflected amplitude and its relative phase.

For p polarization the intensity of the electric field has both tangential and normal components. From Eqs. (1.218) and (1.202) we confirm the continuity of the normal components of $\varepsilon\vec{E}$,

$$\varepsilon_1(E_{zi} + E_{zr}) = \varepsilon_2 E_{zt}. \tag{1.233}$$

Here, ε_1 and ε_2 are the complex dielectric functions of the first and second medium, not to be confused with their real and imaginary part. Equation (1.233) represents the general boundary condition for the normal components of electric inductance, $\vec{D}_\omega = \varepsilon\varepsilon_0\vec{E}_\omega$.

1.5.3 Special Values of the Angle of Incidence

Fresnel amplitudes derived in the previous paragraph simplify for several special values of the angle of incidence:

- Normal incidence
  At normal incidence, $\varphi_1 = 0$, the amplitude reflectivities and transmittances of Eqs. (1.221), (1.222), (1.229), and (1.230) reduce to

$$r = r_s = r_p = \frac{N_1 - N_2}{N_1 + N_2}, \quad t = t_s = t_p = \frac{2N_1}{N_1 + N_2}. \tag{1.234}$$

The two orthogonal polarizations become indistinguishable. For an interface between air ($N_1 \approx 1$) and a highly polarizable medium ($|N_2| >> 1$), the transmitted field vanishes ($t \approx 0$). Accordingly, the reflected wave immediately above the interface has the same magnitude as the incident wave, but the opposite sign ($r \approx -1$). The power reflectivity is the ratio of reflected and incident intensities, i.e.,

$$R = |r|^2 = \frac{(n_2 - n_1)^2 + (k_2 - k_1)^2}{(n_2 + n_1)^2 + (k_2 + k_1)^2}. \tag{1.235}$$

- Critical angle
  The modulus of the amplitude reflectivity of Eq. (1.223), rewritten in the form

$$r_s = \frac{1 - iw}{1 + iw}, \quad w = \frac{\sqrt{N_2^2 - N_1^2 \sin^2\varphi_1}}{iN_1 \cos\varphi_1}, \tag{1.236}$$

  is unity if w is real. This case is called *total reflection*, since the power reflectivity $R_s = r_s^* r_s = |r_s|^2$ is unity. When medium 1 is non–absorbing, $N_1 = n_1$, this condition can be satisfied only for a real negative value of the argument of the square root in Eq. (1.236), which coincides with Eq. (1.207). The same holds for the p–polarized reflectivity of Eq. (1.231), as seen from the form

$$r_p = \frac{1 - iw}{1 + iw}, \quad w = \frac{N_2 \cos\varphi_1}{iN_1\sqrt{N_2^2 - N_1^2 \sin^2\varphi_1}}. \tag{1.237}$$

  Note that the existence of evanescent waves in medium 2 means no energy transfer across the interface for both polarizations. However, the phase shifts differ and depend on the angle of incidence.
- Brewster angle
  For both media transparent ($N_1 = n_1$, $N_2 = n_2$), the p–polarized reflectivity vanishes for the angle of incidence equal to the *Brewster angle*

$$\varphi_B = \arctan(n_2/n_1). \tag{1.238}$$

  At this angle the incident p–polarized wave is totally refracted into the second medium. Consequently, a wave with an arbitrary elliptic polarization, except for the linear p–polarization, is reflected as linearly

s–polarized. The same is true for unpolarized or partially polarized light. This is the reason for calling the Brewster angle the *polarizing angle*. Note that the phase shift Δ_p defined in Eq. (1.231) jumps from the value of $-\pi$ for $\varphi_1 < \varphi_B$ to zero above the Brewster angle, since r_p changes sign from negative to positive values. Vanishing of the reflectivity can, of course, be described as vanishing of the angle ψ_p.

The angle ψ_p does not reach zero at the Brewster angle when medium 2 is not transparent, i.e., has a nonvanishing extinction coefficient k_2. However, it exhibits a minimum at *pseudo–Brewster* angle.

- Principal angle
  The *principal angle* φ_P is defined as the angle of incidence at which the difference $\Delta_p - \Delta_s$ is $\pi/2$. A linearly polarized wave, different from *s* or *p* wave, is reflected as elliptically polarized with the major and minor axes parallel or perpendicular to the plane of incidence when the angle of incidence is φ_P. The difference between principal and pseudo–Brewster angles is small for low values of k_2 and tends to zero for vanishing absorption in medium 2.

1.5.4 Ratio of Amplitude Reflectivities

Reflection ellipsometry determines the state of polarization of a reflected wave with suitably prepared polarization state of an incident wave. A general elliptic polarization of the incident and reflected waves can be resolved into the two basic linear polarizations, parallel and perpendicular to the plane of incidence.

Let us find the link of the observed polarization states to the amplitude reflectivities derived in Section 1.5.2. The simplest approach is to assume the incident linearly polarized wave composed of in–phase *s*– and *p*–polarized vibrations of equal amplitudes. A convenient representation of this polarization state requires a suitable choice of the coordinate system in the plane perpendicular to the wavevector. Using Fig. 1.8, the x_i axis will be in the plane of incidence, oriented along the vector $\vec{E}_{ip}$; the y_i axis will be perpendicular to the plane of incidence, pointing to the reader. In the notation of paragraph 1.3.1, the ellipsometric angles of the incident wave are $\psi_{x_i y_i} = \pi/4$, $\Delta_{x_i y_i} = 0$. The plane of polarization is inclined by 45° from the plane of incidence in the positive sense (counterclockwise) when looking into the incident beam. The reflected wave undergoes different changes of its *p*– and *s*–polarized components, with the relative amplitudes given by Eqs. (1.229) and (1.221). We have to specify the polarization state of the reflected wave in a suitably chosen coordinate system in the plane perpendicular to its wavevector. One of the high–symmetry choices is the x_r axis in the plane of incidence, oriented along the vector $\vec{E}_{rp}$ in Fig. 1.8. The y_r axis is perpendicular to the plane of incidence, pointing to the reader; it coincides with y_i.

Since the electric vector of the *s*–polarized component is parallel to both axes y_i and y_r, the amplitude of the reflected wave in the x_r–y_r coordinate system is r_s times the amplitude of the incident wave. The corresponding ratio of the *p*–polarized amplitudes in the two coordinate systems, however, is $-r_p$. This can be seen from the first of Eqs. (1.228) and the inverted sense of the projections of x_i and x_r on the x axis of Fig. 1.8. Consequently, the state of polarization of the reflected wave, represented by the ellipsometric angles in the x_r–y_r coordinate system, is

$$\tan\psi_{x_r y_r} \exp i\Delta_{x_r y_r} = -\frac{r_p}{r_s} = -\frac{\tan\psi_p}{\tan\psi_s} \exp i(\Delta_p - \Delta_s). \qquad (1.239)$$

The inverted sign of r_p above is the consequence of the particular choice of the coordinate systems. If the x_r–y_r axes were rotated by 180° in their plane, the resulting ellipsometric angles of Eq. (1.239) would be the same; however, the minus sign would appear in the ratio of the *s*–polarized amplitudes, because of the antiparallel y_i and y_r axes. The minus sign of the ratio appears due to the mirror reflection which interchanges the left– and right–handed rotations. Our choice here can be described as adopting the *Fresnel convention* for the *p*–polarized amplitude reflectance, which differs by the sign from the *Verdet convention*[6].

The ratio of the amplitude reflectivities is the basis for the interpretation of reflection ellipsometry measurements. The fundamental ellipsometric equation (1.239) is usually written in the form

$$\tan\psi \exp i\Delta = \frac{\tilde{r}_p}{\tilde{r}_s}, \qquad (1.240)$$

with a tacitly assumed reference to a proper coordinate system for the definition of the ellipsometric angles, and a modified *definition* of the r_p Fresnel coefficient, $\tilde{r}_p = -r_p$, while retaining $\tilde{r}_s = r_s$. This definition has the drawback of giving opposite amplitude reflectivities of the *p*– and *s*–polarized wave at normal incidence, where both waves are physically equivalent and cannot be distinguished one from another. On the other hand, the two schemes are very easily identified, since the only difference is the opposite sign of the reflectance ratio, or, equivalently, the shift of relative phase Δ by 180°.

We use in this chapter the Fresnel amplitudes of Eqs. (1.221) and (1.229) and form their ratio,

$$\rho \equiv \frac{r_p}{r_s} = \frac{\tan\psi_p}{\tan\psi_s} \exp i(\Delta_p - \Delta_s) \equiv \tan\psi \exp i\Delta. \qquad (1.241)$$

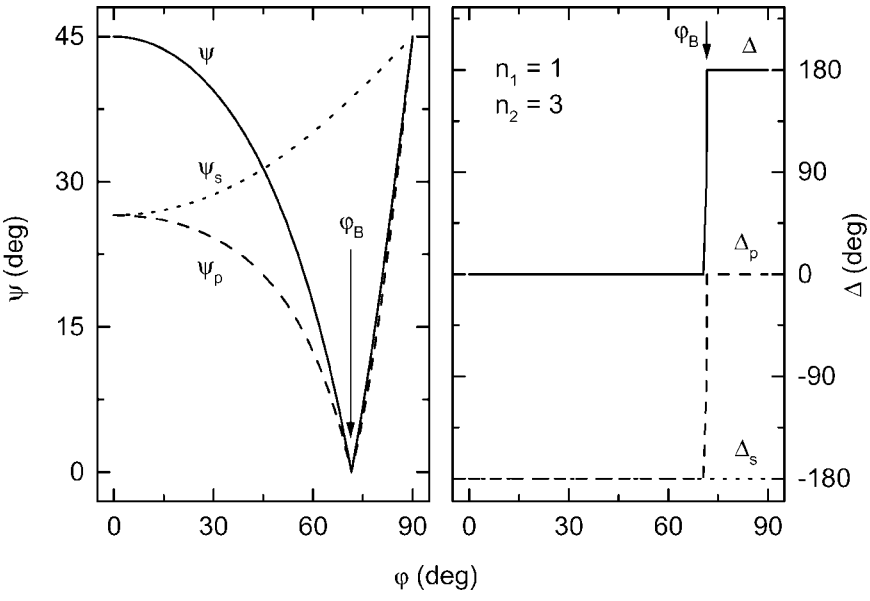


Figure 1.9 The ratios (solid lines) of Fresnel amplitudes for *p*–polarized (dashed lines) and *s*–polarized (dotted lines) waves as functions of the angle of incidence for the interface between two non–absorbing media ($n_1 < n_2$). The arrows point to Brewster angle.

A typical dependence of this complex quantity is shown in Fig. 1.9 for an interface between two non–absorbing media. Note the different behavior of the magnitudes of the *p*– and *s*–components, with vanishing ψ_p at the Brewster angle of Eq. (1.238). The electric field of the incident and reflected *s*–waves is always opposite (i.e., phase–shifted by 180°). The same is true for the *p*–waves below the Brewster angle. When the angle of incidence crosses φ_B, the phase jumps by 180°. When illuminated with a *p*–polarized convergent light beam at the Brewster angle, the interface produces a significant depolarization; it is due to the addition of almost orthogonal polarization states for the angle of incidence below and above φ_B.

The influence of absorption in a semi–infinite sample on the ellipsometric ratio is shown in Fig. 1.10. Note the gradual increase and shift of the minimum in ψ with increasing extinction at (pseudo)–Brewster angle. The position where Δ crosses 90° defines the principal angle. The situation when the incident light enters the sample from an optically denser medium is shown in Fig. 1.11. Since both media are non–absorbing, the ellipsometric ratio vanishes at Brewster angle. Both *p*– and *s*–polarizations are totally reflected for the angle of incidence above the critical angle. Consequently, ψ is constant and equal to 45° for $\varphi > \varphi_c$. While the

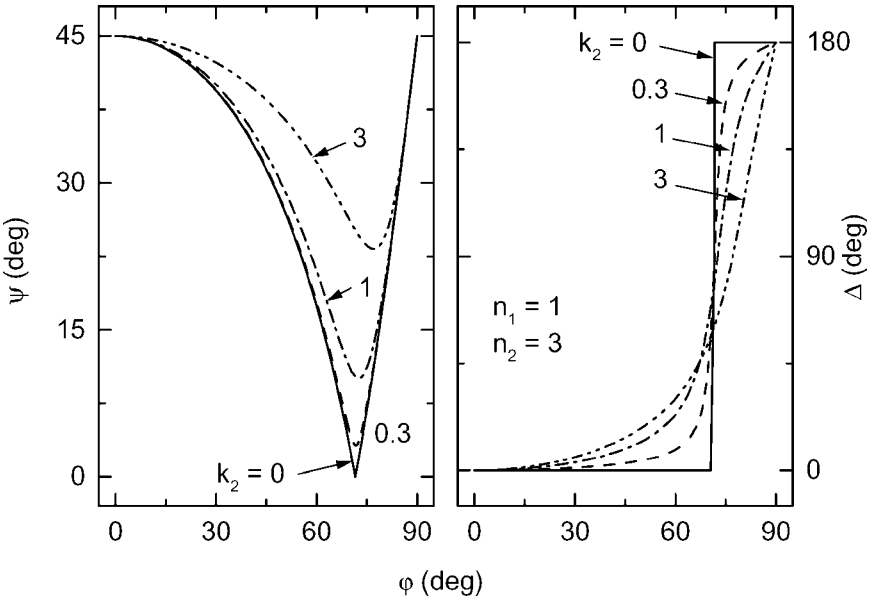


Figure 1.10 The ratios of Fresnel amplitudes as functions of the angle of incidence for the interface between the non–absorbing ambient of $n_1 = 1$ and the sample of $n_2 = 3$ and different values of k_2.

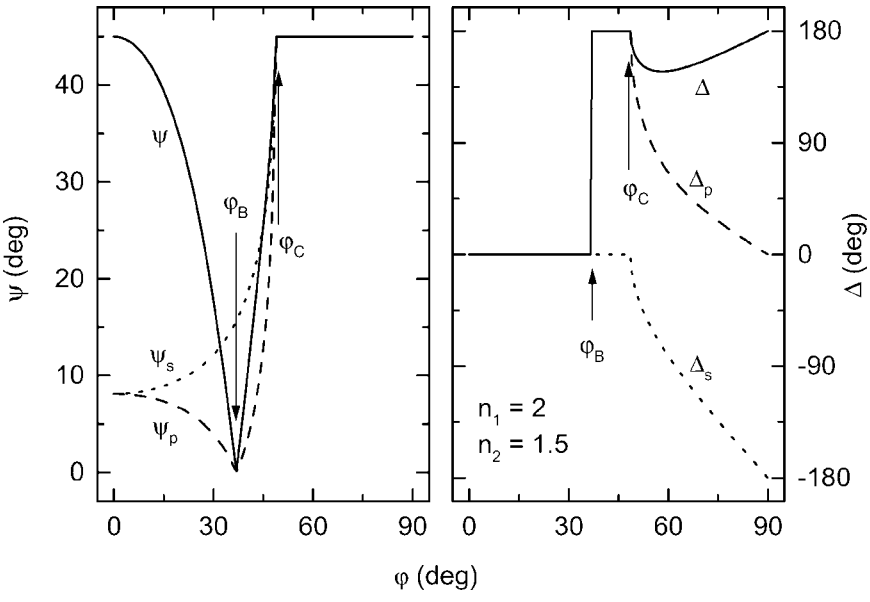


Figure 1.11 The ratios (solid lines) of Fresnel amplitudes for p–polarized (dashed lines) and s–polarized (dotted lines) waves as functions of the angle of incidence for the interface between two non–absorbing media ($n_1 > n_2$). The arrows point to Brewster and critical angles.

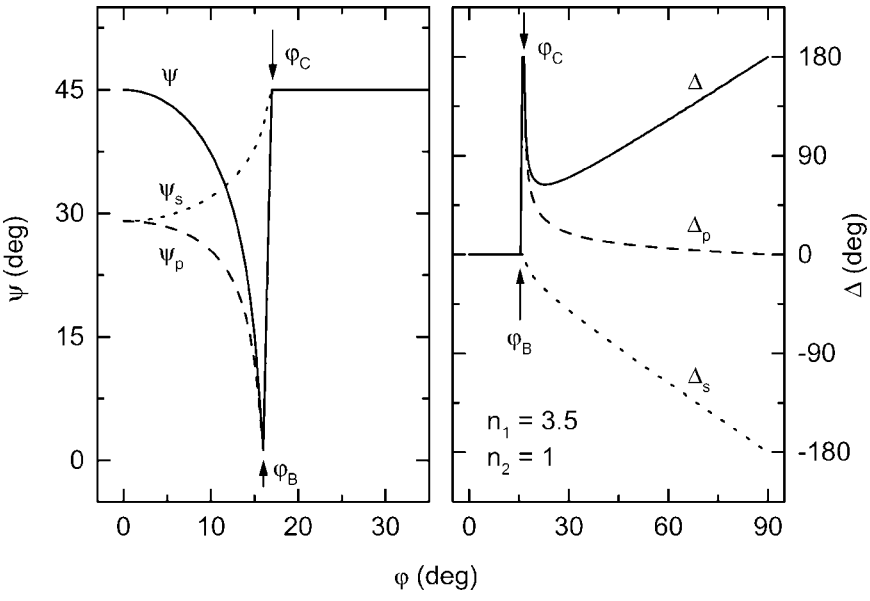


Figure 1.12 Same as Fig. 1.11, but for different refractive indices (when the ratio n_1/n_2 is large).

phase of the *s*–polarization stays at zero up to the critical angle, the *p*–component undergoes a jump by 180° at the Brewster angle. Above φ_c the phases of both components vary rather steeply. When the ratio n_1/n_2 is large, both Brewster and critical angles of Eqs. (1.238) and (1.207) are low and close to one another. Figure 1.12 shows the ellipsometric ratio for refractive indices representative of the internal reflection at the silicon/air interface in infrared. Note that the total phase shift of 90° (i.e., quarter–wave) can be achieved for the angle of incidence close to 45°. Owing to a very high transparency of pure silicon in infrared, this phenomenon can be used as the basis of prism retarders.

1.5.5 Propagation Matrices, Stratified Structures

Propagation of plane waves through multilayer structures with parallel interfaces and homogeneous layers can be described efficiently using *propagation matrices*. This approach uses the fact that the field amplitudes at different depths in the stratified structure are (mutually) linearly dependent, which is a consequence of the linearity of wave equation.

Let us start with noting that the Fresnel amplitudes for *s*–polarization, Eqs. (1.221) and (1.222), and for *p*–polarization, Eqs. (1.229) and (1.230), are essentially of the same form as those for normal incidence, Eqs. (1.234). Namely, the simple relations for amplitude reflectance and transmittance

$$r_{12} = \frac{\tilde{N}_1 - \tilde{N}_2}{\tilde{N}_1 + \tilde{N}_2}, \quad t_{12} = \frac{2\tilde{N}_1}{\tilde{N}_1 + \tilde{N}_2}, \tag{1.242}$$

can be used for both polarizations, when the following *effective refractive indices*

$$\begin{aligned} \tilde{N}_j &= N_j \cos\varphi_j \quad \text{for s-polarization,} \\ \tilde{N}_j &= N_j/\cos\varphi_j \quad \text{for p-polarization,} \quad j = 1,2 \end{aligned} \tag{1.243}$$

are inserted into Eqs. (1.242). Further, the reflectance and transmittance coefficients for the incident wave illuminating the interface from the second medium are obtained by interchanging the indices in Eq. (1.242),

$$r_{21} = \frac{\tilde{N}_2 - \tilde{N}_1}{\tilde{N}_1 + \tilde{N}_2}, \quad t_{21} = \frac{2\tilde{N}_2}{\tilde{N}_1 + \tilde{N}_2}. \tag{1.244}$$

The coefficients of Eqs. (1.242) and (1.244) are related by

$$r_{21} = -r_{12}\,, \quad t_{21}t_{12} + r_{12}^2 = 1. \tag{1.245}$$

Using Eqs. (1.242) and (1.244), we can describe fairly easily the matching of *four distinct plane waves*, traveling upwards and downwards in both media of Fig. 1.8. The simplified framework of normal incidence allows us to treat all field amplitudes as scalars. Both vectors of electric and magnetic intensity are perpendicular to the direction of propagation, which is either z or $-z$. The former waves will be denoted by the upper index $+$, and the latter by the upper index $-$. The intensities of electric field in the first medium are

$$\begin{aligned} E_1^+(z,t) &= E_1^+ \exp[i(k_{z1}z - \omega t)], \\ E_1^-(z,t) &= E_1^- \exp[i(-k_{z1}z - \omega t)], \quad z < 0, \end{aligned} \tag{1.246}$$

with $k_{z1} = (\omega/c)\tilde{N}_1$. Similarly, the intensities in the second medium are

$$\begin{aligned} E_2^+(z,t) &= E_2^+ \exp[i(k_{z2}z - \omega t)], \\ E_2^-(z,t) &= E_2^- \exp[i(-k_{z2}z - \omega t)], \quad z > 0, \end{aligned} \tag{1.247}$$

with $k_{z2} = (\omega/c)\tilde{N}_2$. Matching the fields at the interface $z = 0$ results in a linear dependence of the amplitudes, which can be represented by a 2×2 matrix $\hat{I}_{12}$ in the following relation,

$$\begin{bmatrix} E_1^+ \\ E_1^- \end{bmatrix} = \hat{I}_{12} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix}. \tag{1.248}$$

Identification of amplitude ratios with reflectivities and transmittances,

$$\frac{E_1^-}{E_1^+} = r_{12}, \quad \frac{E_2^+}{E_1^+} = t_{12}, \quad \frac{E_2^+}{E_2^-} = r_{21}, \quad \frac{E_1^-}{E_2^-} = t_{21}, \tag{1.249}$$

and the use of Eqs. (1.245), leads readily to the following result,

$$\hat{I}_{12} = \frac{1}{t_{12}} \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}. \tag{1.250}$$

This matrix "transfers" the electric field amplitudes, i.e., the incoming and outgoing fields at $z = 0$, across the interface from the first to the second medium. In the reverse order, the corresponding matrix $\hat{I}_{21}$ is the inverse of $\hat{I}_{12}$,

$$\hat{I}_{21} = \hat{I}_{12}^{-1} = \frac{1}{t_{21}} \begin{bmatrix} 1 & r_{21} \\ r_{21} & 1 \end{bmatrix} = \frac{t_{12}}{1 - r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix}. \tag{1.251}$$

Besides the transfer across interfaces, we have to relate the field strengths after passing a distance d through a homogeneous layer of refractive index N. From Eqs. (1.246) or (1.247) it is clear that the amplitudes at z and $z+d$ are described by a diagonal matrix $\hat{T}_d$,

$$\begin{gathered} \begin{bmatrix} E_{z+d}^+ \\ E_{z+d}^- \end{bmatrix} = \hat{T}_d \begin{bmatrix} E_z^+ \\ E_z^- \end{bmatrix}, \quad \hat{T}_d = \begin{bmatrix} \exp(i\theta) & 0 \\ 0 & \exp(-i\theta) \end{bmatrix}, \\ \theta = \frac{\omega}{c} dN \cos\varphi = \frac{2\pi}{\lambda} dN \cos\varphi. \end{gathered} \tag{1.252}$$

The phase θ depends on the angle φ related to the z–component of the wavevector in the medium, $k_z = (\omega/c)N\cos\varphi$. Note that we can calculate the phase θ using the effective refractive index $\tilde{N} = N\cos\varphi$ of the s–polarization.

The propagation of polarized light through a stratified structure can be described by a series of matrix multiplications, with the matrices of Eqs. (1.250) and (1.252). This scheme can be modified to involve a single

matrix for the propagation through a homogeneous film. The modification is best introduced by noting that the sum of the amplitudes in Eqs. (1.246) and (1.247) coincides,

$$E_{12} \equiv E_1^+ + E_1^- = E_2^+ + E_2^-. \tag{1.253}$$

The same is true for their difference, multiplied by the refractive index,

$$H_{12} \equiv \tilde{N}_1(E_1^+ - E_1^-) = \tilde{N}_2(E_2^+ - E_2^-); \tag{1.254}$$

namely, this is the continuity condition of the tangential components of the total magnetic field of Eqs. (1.217). The quantity H_{12} differs from the magnetic field intensity merely by the constant factor of $c\mu_0$. The two amplitudes of the total field can be used as components of the following column vector,

$$\begin{bmatrix} E_{12} \\ H_{12} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \tilde{N}_1 & -\tilde{N}_1 \end{bmatrix} \begin{bmatrix} E_1^+ \\ E_1^- \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \tilde{N}_2 & -\tilde{N}_2 \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix}, \tag{1.255}$$

which remains constant upon crossing the interface. It is worth noting that a simple inversion of one of the matrices of Eqs. (1.255) provides the Fresnel amplitudes occurring in Eq. (1.250),

$$\begin{aligned} \begin{bmatrix} E_1^+ \\ E_1^- \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ \tilde{N}_1 & -\tilde{N}_1 \end{bmatrix}^{-1} \begin{bmatrix} 1 & 1 \\ \tilde{N}_2 & -\tilde{N}_2 \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix} \\ &= \frac{1}{2} \begin{bmatrix} 1 & 1/\tilde{N}_1 \\ 1 & -1/N_1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ \tilde{N}_2 & -\tilde{N}_2 \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix} \\ &= \begin{bmatrix} (\tilde{N}_1 + \tilde{N}_2)/(2\tilde{N}_1) & (\tilde{N}_1 - \tilde{N}_2)/(2\tilde{N}_1) \\ (\tilde{N}_1 - \tilde{N}_2)/(2\tilde{N}_1) & (\tilde{N}_1 + \tilde{N}_2)/(2\tilde{N}_1) \end{bmatrix} \begin{bmatrix} E_2^+ \\ E_2^- \end{bmatrix}. \end{aligned} \tag{1.256}$$

This is expected since the meaning of Eqs. (1.255) is just the continuity of the tangential components of the total electric and magnetic fields.

The total field contains two waves moving in opposite directions. We have to apply Eqs. (1.252) in order to find its change after passing a homogeneous slab of material. For the j–th layer in a stack, having refractive index N_j and thickness d_j, the waves reach the interface with the next layer with the amplitudes

$$\begin{bmatrix} E_{j+1} \\ H_{j+1} \end{bmatrix} = \hat{R}_j \begin{bmatrix} E_j \\ H_j \end{bmatrix} = \begin{bmatrix} \cos\theta_j & \sin\theta_j/\tilde{N}_j \\ \tilde{N}_j \sin\theta_j & \cos\theta_j \end{bmatrix} \begin{bmatrix} E_j \\ H_j \end{bmatrix}, \tag{1.257}$$

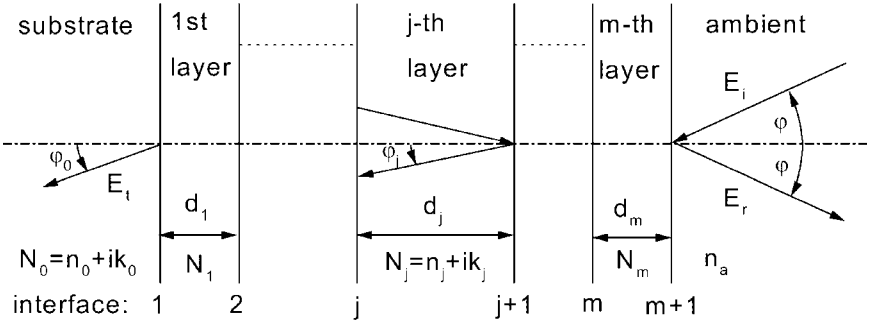


Figure 1.13 Reflection and transmission of a plane wave by a multilayer structure consisting of m films. Transparent ambient with the real refractive index n_a is assumed; φ is the angle of incidence.

$$\cos\theta_j = \frac{\exp(i\theta_j) + \exp(-i\theta_j)}{2}, \quad \sin\theta_j = \frac{\exp(i\theta_j) - \exp(-i\theta_j)}{2}, \qquad \theta_j = \frac{2\pi}{\lambda} d_j N_j \cos\varphi_j. \tag{1.258}$$

Consider a layered structure shown schematically in Fig. 1.13. It consists of m homogeneous layers placed between two semi–infinite media, the ambient (of real refractive index n_a) and the substrate (of refractive index N_0). The optical field in the structure is best computed starting at the first interface, where only the wave propagating into the substrate is present. We are interested in *relative amplitudes* of the incident, reflected, and refracted waves, E_i, E_r, and E_t, respectively. We can therefore choose $E_t = 1$ in the initial column vector

$$\begin{bmatrix} E_1 \\ H_1 \end{bmatrix} = \begin{bmatrix} E_t \\ -N_0 E_t \end{bmatrix} = \begin{bmatrix} 1 \\ -N_0 \end{bmatrix}. \tag{1.259}$$

The total field at the interface with ambient (i.e., at the surface of the structure) is

$$\begin{bmatrix} E_{m+1} \\ H_{m+1} \end{bmatrix} = \hat{R}_m \hat{R}_{m-1} \ldots \hat{R}_1 \begin{bmatrix} 1 \\ -N_0 \end{bmatrix}. \tag{1.260}$$

The final step is to compute the electric field amplitudes at the surface. We can use the inverted matrix of Eq. (1.255) with the effective refractive index of the ambient to obtain

$$\begin{bmatrix} E_r \\ E_i \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1/\tilde{n}_a \\ 1 & -1/\tilde{n}_a \end{bmatrix}\begin{bmatrix} E_{m+1} \\ H_{m+1} \end{bmatrix} = \hat{R}_a \begin{bmatrix} E_{m+1} \\ H_{m+1} \end{bmatrix}, \tag{1.261}$$

$$\begin{aligned} \tilde{n}_a &= n_a \cos\varphi \qquad \text{for } s\text{-polarization,} \\ \tilde{n}_a &= n_a/\cos\varphi \qquad \text{for } p\text{-polarization.} \end{aligned} \tag{1.262}$$

These amplitudes yield the sought reflectance and transmittance coefficients as

$$r = \frac{E_r}{E_i}, \quad t = \frac{1}{E_i}. \tag{1.263}$$

The coefficients differ for s– and p–polarization because of the different effective refractive indices in the propagation matrices $\hat{R}$. Equations (1.204) and (1.205) yield the following convenient expressions containing the real values of φ and n_a,

$$\begin{aligned} \tilde{N}_j &= N_j\sqrt{1 - (n_a \sin\varphi/N_j)^2} \qquad \text{for } s\text{-polarization,} \\ \tilde{N}_j &= N_j/\sqrt{1 - (n_a \sin\varphi/N_j)^2} \qquad \text{for } p\text{-polarization.} \end{aligned} \tag{1.264}$$

It should be remembered that there is no ambiguity in choosing the sign of the square root in Eqs. (1.264). In fact, both n_j and k_j entering the imaginary part of its argument,

$$\frac{2n_a^2 n_j k_j \sin^2\varphi}{(n_j^2 - k_j^2)^2 + 4n_j^2 k_j^2},$$

are non–negative. This ensures that both real and imaginary parts of the square root itself are non–negative. The opposite square root, having both real and imaginary parts negative, is unphysical.

1.5.6 Substrate–Film–Ambient System

The procedures from Section 1.5.5 are well suited for finding explicit expressions for the reflection and transmission coefficients for a single film (thickness d_1, refractive index N_1) placed on the substrate (refractive index N_0). Two matrices of Eq. (1.250) are to be built from the Fresnel coefficients for the 0–1 (substrate–film) and 1–2 (film–ambient) interfaces. The transfer across the film is described by the matrix of Eq. (1.252),

yielding finally the reflection and transmission coefficients of the filmed structure as

$$r = \frac{r_{21} + r_{10}\exp(i2\theta)}{1 + r_{21}r_{10}\exp(i2\theta)}, \quad t = \frac{t_{21}t_{10}\exp(i\theta)}{1 + r_{21}r_{10}\exp(i2\theta)}, \tag{1.265}$$

$$\theta = \frac{2\pi}{\lambda} d_1 N_1 \cos\varphi_1 = \frac{2\pi}{\lambda} d_1 N_1 \sqrt{1 - (n_a \sin\varphi / N_1)^2}. \tag{1.266}$$

Here n_a is the real refractive index of the ambient. The results of matrix multiplications of Eqs. (1.260) and (1.261) inserted into Eq. (1.263) yield the following equivalent form of the reflectivity coefficient,

$$r = \frac{\hat{N}_1(\hat{n}_a - \hat{N}_0) - i(\hat{n}_a\hat{N}_0 - \hat{N}_1{}^2)\tan\theta}{\hat{N}_1(\hat{n}_a + \hat{N}_0) - i(\hat{n}_a\hat{N}_0 + \hat{N}_1{}^2)\tan\theta}. \tag{1.267}$$

It is worth noting that for real refractive index of the film, $N_1 = n_1$, obeying $n_1 > n_a \sin\varphi$, both *p*– and *s*–polarized reflectivities are periodic functions of the real angle θ. The period is π as seen from Eqs. (1.265) or (1.267). In other words, the reflectivity at a given angle of incidence is the same for the film thicknesses differing by

$$D_1 = \frac{\lambda}{2\sqrt{n_1^2 - (n_a \sin\varphi)^2}}. \tag{1.268}$$

The period varies between $\lambda/(2n_1)$ for normal incidence, and $\lambda/(2\sqrt{n_1^2 - n_a^2})$ for grazing incidence.

Two major simplifications of the result for the reflectivity follow from Eq. (1.267) for

- $\theta = 0$, which occurs for zero thickness, or multiples of the period D_1. The reflectivity coefficient is equal to that of the bare substrate,

$$r_{subs} = \frac{\hat{n}_a - \hat{N}_0}{\hat{n}_a + \hat{N}_0}. \tag{1.269}$$

- $\theta = \pi/2$, which occurs for thicknesses equal to $D_1/2$ plus multiples of the period D_1. The reflectivity is

$$r_{D/2} = \frac{\hat{n}_a\hat{N}_0 - \hat{n}_1{}^2}{\hat{n}_a\hat{N}_0 + \hat{n}_1{}^2}. \tag{1.270}$$

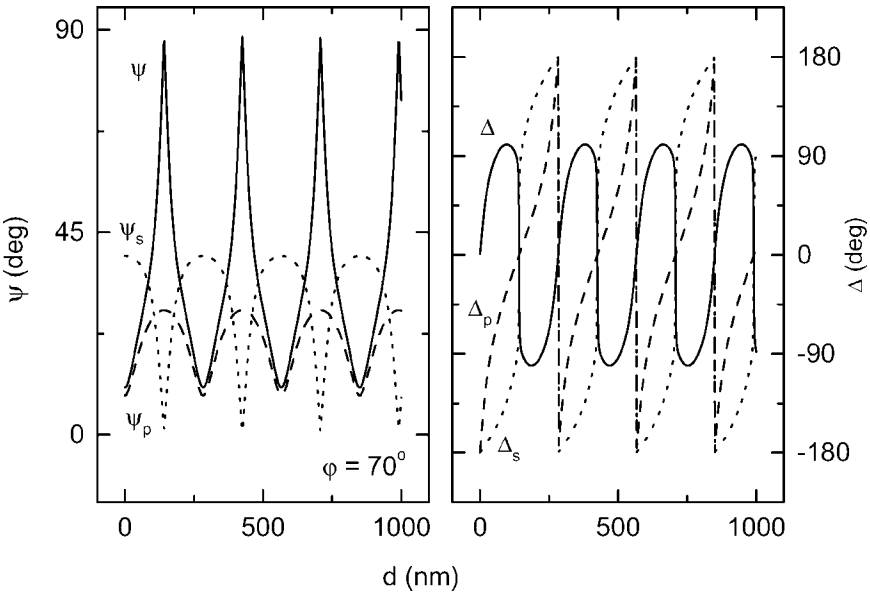


Figure 1.14 The ratios (solid lines) of reflectivity amplitudes for *p*–polarized (dashed lines) and *s*–polarized (dotted lines) waves as functions of the film thickness for SiO_2/Si structure. The optical constants for the wavelength of 633 nm are: $n_1 = 1.46$, $N_0 = 3.87 + 0.02i$.

For a non–absorbing substrate with the real refractive index n_0, the reflectivity vanishes for

$$\hat{n}_1^{(0)} = \sqrt{\hat{n}_a \hat{n}_0}. \tag{1.271}$$

Note that we are using the effective refractive indices, dependent on the angle of incidence, and different for *p*– and *s*–polarization.

The periodic behavior of both reflectivity coefficients and their ratio is shown in Fig. 1.14 for the non–absorbing SiO_2 film on Si substrate. The angle of incidence of 70° is slightly below the pseudo Brewster angle of the substrate for the wavelength of 633 nm. Note the low values of the modulus for *s* polarization at about the half of the period; in this case, the condition of Eq. (1.271) is nearly satisfied. Since the *p*–polarized reflectivity is fairly high for these thicknesses, the ellipsometric angle ψ almost reaches 90°. The phase shift Δ exhibits steep changes at well defined thicknesses; either thickness variations or finite resolution would lead to pronounced depolarization around these values for certain incident polarization states.

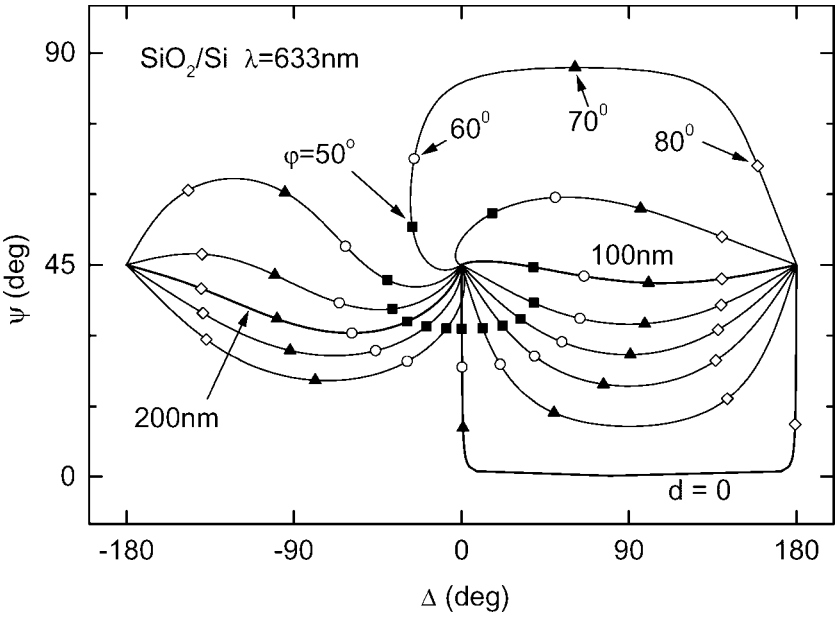


Figure 1.15 The ratios of reflectivity amplitudes for *p*– and *s*–polarized waves for different film thickness in the SiO_2/Si structure. The lines are computed for the angle of incidence varying from zero to 90°, with several values marked by the symbols. The film thickness has been increased from zero in steps of 20 nm. The optical constants for the wavelength of 633 nm are: $n_1 = 1.46$, $N_0 = 3.87 + 0.02i$.

An overview of the ratio $\tan\psi \exp i\Delta = r_p/r_s$ of the reflectivity amplitudes for the SiO_2/Si system at 633 nm is shown in Fig. 1.15. The pairs (ψ,Δ) are repeated with the periods of about 254.6 nm and 293.6 nm for the angles of incidence of 50° and 80°, respectively. This makes multiple–angle–of–incidence measurements suitable also for films thicker than the period of Eq. (1.268). A similar representation of the ellipsometric angles is shown in Fig. 1.16 for an absorbing film. The two contours were calculated for constant angles of incidence and varying film thickness. The points labeled $d = 0$ are the ellipsometric angles for bare silicon at 496 nm. The spirals traveled in the ψ–Δ plane with increasing film thickness are governed by (i) the tendency to the periodicity of Eq. (1.268) which would appear for negligible absorption in the film, and (ii) the gradual loss of the wave reflected at the film–substrate interface, due to the finite penetration depth of Eq. (1.214) in the absorbing film. The spirals converge to points resulting for a semi–infinite sample of the film material. Only a moderate tendency to depolarization

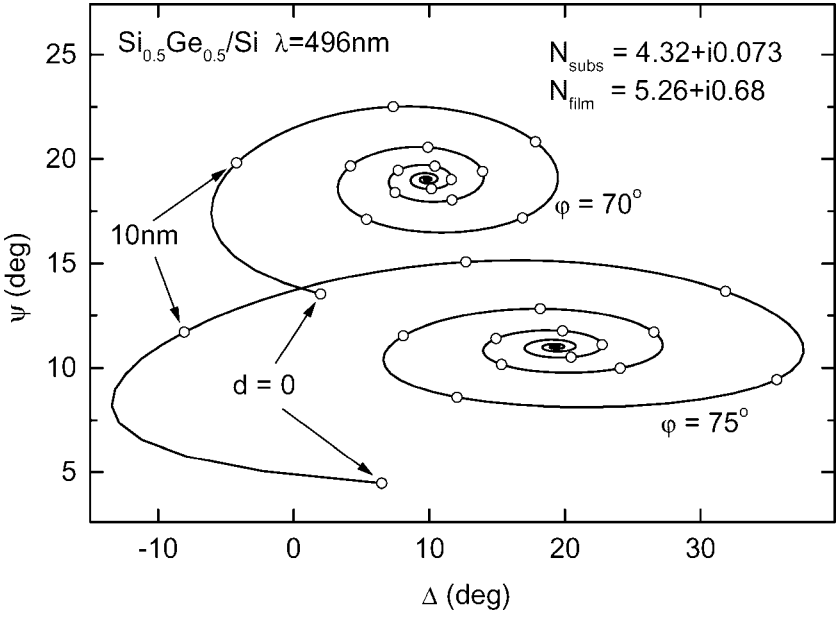


Figure 1.16 The ratios of reflectivity amplitudes for *p*– and *s*–polarized waves for different film thickness in the SiGe/Si structure. The lines are computed for two angles of incidence, 70° (upper spiral) and 75° (lower spiral). The film thickness *d* increases along the lines, with successive steps of 10 nm marked by open circles. The optical constants for the wavelength of 496 nm are: $N_1 = 4.32 + 0.073i$, $N_0 = 5.26 + 0.68i$.

due to the angular spread or variable film thickness is expected in this situation.

The final example in Fig. 1.17 shows the magnitudes of *s*– and *p*–polarized reflectivities for films on highly polarizable substrates. The parameters chosen are representative of infrared response of a dielectric film in the range of strong longitudinal optic vibrations, placed on a metallic substrate. Since the real part of the refractive index of the film material is less than that of the ambient, both *s*– and *p*–polarized waves would be totally reflected ($\psi_s = \psi_p = 45°$) above the critical angle of incidence, Eq. (1.207), if the film was non–absorbing and infinitely thick. Due to the large magnitude of its refractive index ($N = 12 + 50i$ represents silver at the vacuum wavelength of about 15 μm), the bare substrate has the moduli of both reflectivity coefficients fairly close to unity, i.e., $\psi_s \approx \psi_p \approx 45°$, for any angle of incidence. However, the magnitude of reflectance is quite different for both polarizations in the filmed structure.

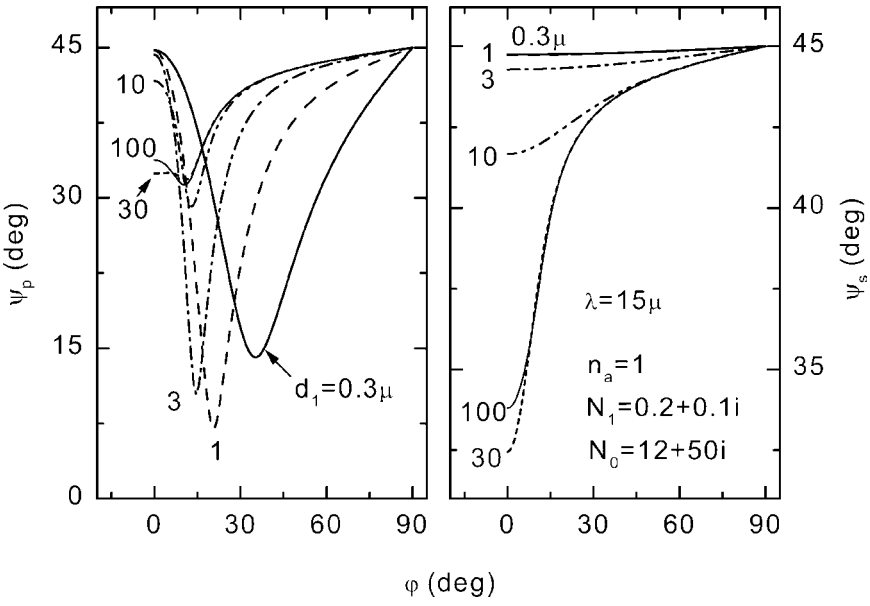


Figure 1.17 Amplitude reflectivity of *p*– and *s*–polarized light shown as the ellipsometric angle ψ; $\tan\psi$ is the modulus of the coefficient.

Since its high polarizability almost extinguishes the field inside the substrate, the difference is related to the different amplitudes of the field in the film.

1.6 References

1. M. Born and E. Wolf, *Principles of Optics*, 6th ed. (Cambridge University Press, 1980), ch. I and II.
2. C. Brosseau, *Fundamentals of Polarized Light: A Statistical Optics Approach* (Wiley, New York, 1998).
3. R. A. Chipman, Polarimetry, in *Handbook of Optics II*, ed. H.B. Crawford and B. Carson (McGraw–Hill, New York, 1995), ch. 22.
4. B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (Wiley Interscience, New York, 1991).
5. J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975).
6. R. T. Holm, Convention Confusions, in *Handbook of Optical Constants of Solids II*, ed. E. D. Palik (Academic Press, New York, 1991), p. 21.
7. H. E. Bennett and J. M. Bennett, Polarization, in *Handbook of Optics*, ed. H.B. Crawford and B. Carson (McGraw–Hill, New York, 1978), ch. 10.

Next Page

8. G. Baym, *Lectures on Quantum Mechanics* (W. A. Benjamin, New York, 1969), ch. 1.
9. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light* (North–Holland, Amsterdam, 1977).
10. D. E. Aspnes, Spectroscopic ellipsometry of solids, in *Optical Properties of Solids: New Developments*, ed. B. O. Seraphin (North–Holland, Amsterdam, 1976, ch. 15).
11. H. G. Tompkins and W. A. McGahan, *Spectroscopic Ellipsometry and Reflectometry: A User's Guide* (Wiley, New York, 1999).
12. D. E. Aspnes and A. A. Studna, Appl. Optics **14**, 220 (1975).
13. S. Li, J. Opsal, H. Chu, and D. E. Aspnes, J. Opt. Soc. Am. A **18**, 426 (2001).
14. P. A. M. Dirac, *The Principles of Quantum Mechanics* (Oxford University Press, London, 1967), ch. I.
15. D. Y. Smith, Dispersion Theory, Sum Rules, and Their Application to the Analysis of Optical Data, in *Handbook of Optical Constants of Solids*, ed. E. D. Palik (Academic Press, New York, 1985), p. 35.
16. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, (Pergamon, Oxford, 1960).
17. G. N. Ramachandran and S. Ramaseshan, *Crystal Optics*, in Handbuch der Physik, S. Fluegge, Ed., (Springer Verlag, Berlin, 1961), Vol. XXV/1.
18. Reference 1, paragraph 14.1.

2 Optical Physics of Materials

Robert W. Collins

Department of Physics and Astronomy
The University of Toledo

Andre S. Ferlauto

Materials Research Laboratory
The Pennsylvania State University

2.1 Introduction

Several different instrument designs for automatic ellipsometry were developed in the late 1960's,[1,2] including those based on self-compensating,[3] rotating analyzer,[4] and phase modulation[5] principles. With the modern versions of such instruments, it is possible to perform ellipsometry experiments on reflecting samples in various multiple-measurement modes with relative ease. Among the experimental variables utilized in these measurement modes include (i) the incident beam wavelength λ (or photon energy E);[6] (ii) the angle of incidence θ_i;[7] (iii) the time t during sample processing (or layer thickness d in the case of thin film growth);[8] (iv) the incident beam polarization ellipse, defined by its tilt and ellipticity angles (Q_i, χ_j);[9] and (v) the real part of the dielectric function ε_{1a} for the ambient medium,[10] as described in the selected reference articles. These different multiple-measurement modes are designed to enhance the capabilities of ellipsometric data analysis. For example, one can deduce additional sample parameters (e.g., those related to surfaces and interfaces) that provide a better understanding of the overall sample structure, or one can deduce sample parameters with greater confidence.

Among the experimental variables listed above, the wavelength is perhaps the most efficient single one for extracting additional sample information. Hence, the technique of spectroscopic ellipsometry has evolved over the last 25 years into a powerful tool for the analysis of specularly reflecting bulk material and thin film samples.[11,12] In a conventional spectroscopic ellipsometry experiment, one measures the ellipsometric angles (ψ, Δ) at 50 to 300 spectral points from near-infrared to ultraviolet wavelengths. The spectral resolution is chosen to be sufficiently high in

order to resolve the sharpest structures in the (ψ, Δ) spectra of the sample without distortion by instrument broadening.

From single-wavelength ellipsometric measurements of specularly reflecting bulk material or thin film samples, it may be possible to extract overlayer/film thicknesses and component material optical properties, i.e., complex dielectric functions, at the one wavelength under certain circumstances (using, e.g., the multiple-angle-of-incidence measurement mode).[1] From the optical properties at the one wavelength, however, it is not often possible to extract further information on the bulk or thin film material with confidence. In contrast, from spectroscopic ellipsometry measurements, one can readily extract not only layer thicknesses, but also the full wavelength dependence of the complex dielectric function for one or more component materials of the sample structure.[6] The resulting complex dielectric function spectra represent an intrinsic characteristic of the material (at least when considered on a linear scale such that the effects of extrinsic defects can be neglected). As a result, the real and imaginary parts of the dielectric function spectra are often known as "optical constants" and have been tabulated extensively for different bulk and thin film materials.[13,14]

The intrinsic nature of the complex dielectric function implies that it provides a fingerprint of the material from which one can extract useful information on alloy composition, void density, crystalline content, and even material temperature, with a very high degree of confidence. These capabilities are particularly useful and have been applied quite extensively for semiconductor analysis. (For a selection of early studies see Refs. 15–18 and for an excellent review see Ref. 19.) In addition, the complex dielectric function spectra provide fundamental information about the nature of electronic states in the material,[20] and on its band structure—in the case of crystalline materials (see, e.g., articles by Cardona and coworkers, Refs. 21 and 22). As a result, one of the first steps in research on the optical properties of a new material involves a determination of its complex dielectric function spectra. Finally, from the complex dielectric function, one can predict the performance of the material in a variety of optical and optoelectronic applications. For example, the spectroscopic reflection, absorption, and transmission properties of a bulk material, calculated from the complex dielectric function spectra and thickness, influence the external quantum efficiency of active solid state devices such as photodetectors and photovoltaic cells.[23]

Here, the several approaches for extracting the complex dielectric function spectra of bulk material and thin film samples from spectroscopic ellipsometry data (ψ, Δ) are reviewed. In this way, a motivation is provided for developing a fundamental understanding of the optical properties in terms of the behavior of electrons and the electronic joint density of states of the material.

Surface cleaning. One approach that has been successful in the analysis of opaque bulk materials requires performing surface chemical and chemo-mechanical treatments on the sample in order to remove overlayers including native oxides and surface roughness.[24] Once these overlayers are fully removed such that the real surface approximates as closely as possible a mathematically abrupt termination of the bulk (as is assumed in the derivations of the Fresnel equations), then the complex dielectric function spectra of the material can be deduced directly from the (ψ, Δ) spectra using an inverted form of the Fresnel equations. In fact, ellipsometry at a single wavelength and at variable times during the surface processing steps can be performed in order to determine the extent to which oxide and roughness removal has been successful. This overall approach can be applied only to oriented single-crystalline bulk materials, since the chemo-mechanical surface treatment methods usually lead to different etching rates for different crystal faces, hence, roughening in the case of polycrystalline bulk materials or thin films, in general.

Exact numerical inversion. A second general approach that can be applied to determine the complex dielectric function spectra of bulk material and thin film samples is exact numerical inversion of spectroscopic ellipsometry data (ψ, Δ) using, for example, a Newton-Raphson algorithm[25] or more modern advanced versions.[26] For such algorithms to be applied, the optical properties of all material components of the sample structure except one must be known; in addition, all thicknesses must be known. The approach begins by assigning first guess values for the real and imaginary parts of the one unknown complex dielectric function (ε_1, ε_2) at the first wavelength point (often the highest value of λ in the spectrum). Then (ψ, Δ) at this wavelength point are computed and compared to the experimental results. In the simplest case of the Newton-Raphson method, the algorithm is then used to determine corrections to the first guess (ε_1, ε_2) values, based on linear extrapolations, that are required to bring the calculated and experimental values into closer agreement. The procedure is iterated and successive corrections are made until the calculated and experimental (ε_1, ε_2) agree to within limits smaller than the precision of the experiment, e.g., within 1×10^{-4}. This process is repeated for each spectral point in turn, applying the inverted (ε_1, ε_2) result from the previous spectral point as a first guess for the next spectral point. The success of this overall approach depends sensitively on the sample structure and the unknown complex dielectric function. The iterative process may diverge if the first guess for (ε_1, ε_2) is not sufficiently close to the correct result or if the assumptions about the sample structure are incorrect; furthermore, multiple solutions may also occur. The validity of the resulting complex dielectric function spectra can be verified first by ensuring that they contain no artifacts (see next paragraph) and second by

comparing them to Kramers-Kronig consistent forms predicted on the basis of the physical mechanisms described later in this chapter.

Artifact minimization. In exact numerical inversion, all experimental errors—including incorrect assumptions about the sample structure, e.g., incorrect thicknesses—are transferred directly to the dielectric function and can lead to unphysical artifacts. A third approach turns this problem to an advantage, allowing one to deduce not only the one unknown complex dielectric function, as described in the previous paragraph, but also one or more unknown thicknesses. These thickness values are chosen by trial-and-error to ensure that no artifacts appear in the complex dielectric function deduced by exact numerical inversion. Such artifacts may arise due to interference fringes that disappear from the real and imaginary parts of the dielectric function spectra only when the film thicknesses in the structure are fixed at their correct values.[27] Artifacts may also arise due to features in the known complex dielectric functions of the components of the sample structure (e.g., the substrate complex dielectric function when the unknown component is an overlying thin film).[28] Again these artifacts can only be eliminated if the one or more film thicknesses are chosen correctly.

Semiconductor or dielectric criterion. A fourth approach to extract the complex dielectric function is similar to artifact minimization to the extent that one can deduce one unknown complex dielectric function, and one or more unknown thicknesses in the sample structure. This approach is only relevant when the unknown dielectric function is associated with a semiconductor or dielectric component material, for which it is known that $\varepsilon_2 = 0$ over some spectral range.[6] In this case, the thickness values are chosen by trial and error so as to ensure that ε_2, again obtained along with ε_1 by exact numerical inversion, vanishes over the required spectral range.

Least-squares regression. One of the most popular approaches for extracting the complex dielectric function of a bulk or thin film material is based on the least-squares regression method as exemplified in Refs. 15 through 19. (See also the excellent reviews by Jellison[29,30] in Refs. 11 and 12.) In this method, the unknown complex dielectric function spectra are expressed in terms of a number of wavelength-independent parameters that are determined along with the thicknesses by least-squares regression. In this analysis procedure, first guesses are made for all the wavelength-independent parameters, including the thicknesses and the optical property parameters. Next, the (ψ, Δ) spectra corresponding to these initial guesses are computed for comparison with the experimental results. The least-squares regression algorithm is then used to compute corrections to the first-guess parameter values that are required in order to bring the calculated and experimental values into closer agreement (as in the exact inversion routine). The iteration is complete when a mean-square error

function between the calculated and experimental ellipsometric spectra is minimized. It should be noted that this quantity cannot be reduced to zero since the number of free parameters in the analysis is generally much less than the number of experimental data values.

The earliest approaches in least-squares regression analysis used an unweighted mean square deviation for the error function.[15] More recently, the advantages of the weighted deviations have been demonstrated, whereby the squares of the deviation between the calculated and experimental data are divided by the experimental errors for each wavelength, before taking the average versus wavelength.[31] With this error function, the structural model for the sample and the optical model for the dielectric function spectra can be considered as valid if the weighted version of the error function is near unity. Thus, in least-squares regression analysis, the optical properties of a given component material can be forced to follow a physically correct form—a distinct advantage over the exact inversion methods. The disadvantage occurs when the complex dielectric function is too complicated to be expressed in terms of several free parameters. In this latter case, the least-squares regression analysis will fail (e.g., the weighted version of the error function will exceed unity), and the exact inversion approach is then advantageous. Additional discussion of the disadvantage of least squares regression analysis will appear later in this section.

The goal of this chapter is to provide a discussion of the physical principles of the optical properties of solid state materials that will be useful to the ellipsometry experimentalist. As the general theme of the chapter, mathematical expressions for the complex dielectric function will be developed in terms of wavelength-independent parameters that describe the nature of the electronic transitions in the solid, the joint densities of electronic states, and the interactions of the electrons with the network of atoms. Although expressions based on simple classical mechanical principles can be derived, the most useful expressions are based on rigorous quantum mechanical principles.[20] Such expressions can be employed in comparisons with the complex dielectric functions obtained experimentally as described above, either by the chemical cleaning methods or by exact inversion of spectroscopic ellipsometry data. The comparisons can provide qualitative information on the nature of the material, whereas rigorous fitting procedures can provide values of specific material parameters such as the energy of the fundamental bandgap, the energies of the critical points in the joint density of states, the electron relaxation times, and the relative oscillator strengths and free electron densities. Alternatively, for crystalline solids, the complex dielectric function obtained experimentally can be twice differentiated with respect to photon energy and then fitted to specific theoretical lineshape functions that provide insights into the band structure of the material.[21,22,32] This procedure

is called "critical point analysis" and provides the energies at which the van Hove singularities occur in the joint density of electronic states. Finally, expressions for the complex dielectric function that are derived from physical principles and include only wavelength-independent parameters can be applied directly in least-squares regression analysis of spectroscopic ellipsometry data. In this sense, the development of physically correct descriptions of the optical properties of solids and the analysis of spectroscopic ellipsometry data are inextricably connected.[29,30]

Another important feature of theoretical expressions for the complex dielectric functions of solids is the ability to express the wavelength-independent parameters in these expressions as continuous (or possibly step-wise continuous) functions of one or more of the material properties, such as the alloy composition, doping level, crystalline grain size, or temperature. As a result, once these parameter-property functions are established, the complex dielectric function can be predicted theoretically, based on the specification of one or more of the material properties. In order to establish such a capability, e.g., for a series of semiconductor alloys, one or more parameterized expressions for the complex dielectric function must be developed that are valid for the entire series of alloys. Then, a database having broad utility can be obtained in one of two ways. First, the complex dielectric functions for materials of several different alloy compositions—deduced from spectroscopic ellipsometry data either by surface cleaning methods or by exact inversion—can be fit to the parameterized expressions in order to extract the optical property parameters that can be plotted versus alloy composition. Alternatively, the optical property parameters in such plots can be deduced directly from spectroscopic ellipsometry data for the series of alloys using least-squares regression analysis. In either case, the resulting optical property parameters can then be fit to polynomial functions of alloy composition. Once this database of $p(n + 1)$ polynomial coefficients is established (where p is the number of optical property parameters and n is the order of the polynomial), it can be incorporated into the least-squares regression analysis algorithm. With this enhancement, the algorithm can be used to extract directly the alloy composition of one or more layers in the sample structure along with the layer thicknesses. In the more general situation, one or more material properties including composition, doping level, crystalline grain size, and temperature may be extracted. As a result, such databases when combined with the least-squares regression algorithm can provide an extremely powerful tool for routine material analysis by spectroscopic ellipsometry. Structures of the III-V ternary alloy system gallium-aluminum arsenide ($Ga_{1-x}Al_xAs$) were the first to be studied in detail in this manner.[17,33–35]

Considering the widespread popularity of the least-squares regression algorithm for determination of a variety of materials characteristics—not

only the optical functions, but also band structure parameters, material composition, temperature, etc.—it is important to emphasize potential pitfalls of the overall approach. For most crystalline materials, the mathematical expressions required to describe adequately their complex dielectric functions include a very large number of wavelength-independent free parameters (typically >20). As a result, the best fit parameters deduced in the least-squares regression algorithm can exhibit very large cross-correlation coefficients and consequently large confidence limits. Such parameter uncertainties can generate two problems in the subsequent analyses. First, if one proceeds to plot these optical property parameters versus a material parameter, such as composition or temperature, and fits the resulting plots to polynomial expressions, then large uncertainties in the resulting polynomial coefficients may be generated. This in turn can give rise to erroneous optical functions if the polynomials are used to interpolate or extrapolate to compositions and temperatures for which samples or measurements are unavailable. Second, if one proceeds to interpret the optical property parameters further and deduces band structure information, then the associated confidence limits must be taken into consideration. Otherwise erroneous interpretations concerning the electronic structure and physical processes may result. Further comments on these pitfalls will be given when specific models and examples are presented.

In the following four paragraphs, the overall framework of this chapter will be previewed.[36] In many sections of the chapter, the treatment follows widely available textbook material where more detailed discussions can be found. In these instances, representative texts are provided as references throughout the sections and their titles are also collected at the end of the chapter in the Bibliography.

In Section 2.2, the electromagnetic theory that forms the basis for the propagation of light in bulk materials is presented.[20,37–39] First, the approximate linear relationships are established that express the complex dielectric function, a macroscopic quantity, as a function of the microscopic electric charges/currents induced by electromagnetic fields in a solid.[20] With these relationships, the exact microscopic form of Maxwell's equations can be converted to an approximate macroscopic form. Then plane wave solutions to the macroscopic form of Maxwell's equations in the solid reveal the effects of the complex dielectric function on the propagation of monochromatic electromagnetic waves in the optical frequency domain. Alternative expressions for the optical constants are also defined in Section 2.2. These include the complex index of refraction, the absorption coefficient, and the complex optical conductivity. In fact, outside the condensed matter and applied physics communities, the complex index of refraction is used more commonly than the complex

dielectric function to express the optical properties. The complex optical conductivity is sometimes used in theoretical descriptions of the optical properties of bulk metals. Because most solids are not optically isotropic, a part of Section 2.2 is devoted to electromagnetic wave propagation in these materials, and the concepts of the complex dielectric tensor, the complex index tensor, the optical activity tensor, and the tensor of circular dichroism are introduced.[38,40,41] Characteristics of the complex dielectric function and tensor dictated by causality are treated at the conclusion of Section 2.2.

In Section 2.3, simple classical mechanical models are presented to describe the interaction of a monochromatic electromagnetic wave with a collection of identical atoms that comprise a hypothetical solid. These models are based on the early work of Drude and Lorentz.[20,42,43] This classical approach yields the induced microscopic atomic charge redistribution, and the macroscopic polarization (and hence the dielectric function), from the Newtonian forces of the optical electric field on the electrons of each atom. The model provides a simple qualitative understanding of the origin of absorption and dispersion, or the wavelength-dependence of the complex dielectric function. It can also account for the differences in optical properties among insulators, semiconductors, and metals, in simple terms based on the strength of the restoring forces exerted on the electrons by the atomic cores. Although the complex dielectric function spectra of all solid state materials exhibit the general features of these models, there are few such spectra that can be closely fit and, thus, quantified in detail using the models.

In general, one must incorporate the quantum mechanical features associated with electronic states in the solid in order to closely fit complex dielectric function spectra of real materials.[20,23,32,44,45] Such features are necessary to account for the differences in the complex dielectric functions observed among different materials within a given group (e.g., among metallic aluminum, silver, and gold, or among semiconducting silicon, germanium, and gallium-arsenide). Such a quantum mechanical description is provided in Section 2.4. In a quantum mechanical model of the optical properties of a solid state material, the one-electron approximation is typically used. With this approximation, the effect of the optical electromagnetic field on the many-electron wavefunction is obtained by determining the response of each separate electron wavefunction within a self-consistent potential. Then, one can consider the optical electric field (in a crystalline solid, for example) as inducing transitions of single electrons from their ground Bloch states to higher-lying excited states within the framework of time-dependent perturbation theory. The complex dielectric function is calculated from the expectation value of the conduction current density averaged over a volume that is large compared to

atomic dimensions. Three situations will be discussed in detail: (i) direct interband transitions in metals, semiconductors, and insulators, i.e., electronic transitions that conserve crystal momentum, from filled bands to partially filled or empty bands; (ii) intraband transitions in metals, i.e., transitions of electrons within partially filled bands; and (iii) indirect, interband transitions in semiconductors or insulators, i.e., transitions of electrons from filled bands to empty bands, along with simultaneous absorption or emission of a quantum of lattice vibrational energy. The ability of the complex dielectric function to provide direct information on the joint density of electronic states in an amorphous or crystalline solid is also discussed.

In Section 2.5, examples are presented that demonstrate the scope of derivations described in Sections 2.3 and 2.4.[36] In the first part of this section, typical shapes of the classical Lorentz oscillator and Drude expressions for the real and imaginary parts of the complex dielectric functions are presented. The corresponding results are also presented in terms of the real and imaginary parts of the complex index of refraction, as well as in terms of the absorption coefficient and reflectance (the latter assuming a single vacuum/material interface). These results provide a better intuitive understanding of the manifestations of the dispersion (or wavelength-dependence) of the optical properties in general. Next, the Lorentz oscillator and Drude expressions, in spite of their limitations, are compared qualitatively and quantitatively (i.e., by fitting) to different materials, including amorphous silicon (a-Si:H) in thin film form,[46] aluminum (Al) and silver (Ag) in particulate form,[47,48] single-crystal gallium arsenide (GaAs),[49,50] and thin film polycrystalline Ag.[36] In the second part of this section, more detailed expressions developed on the basis of quantum mechanical principles are employed specifically to fit a similar group of materials more precisely, including thin film amorphous silicon-germanium (a-$Si_{1-x}Ge_x$:H) and amorphous silicon-carbon (a-$Si_{1-x}C_x$:H) alloys,[51,52] single-crystal Si and GaAs,[53,54] mercury cadmium telluride ($Hg_{1-x}Cd_xTe$) ternary alloys,[55] and thin film Al.[47]

Before proceeding, it should be noted that several topics related to the optical physics of solids are treated only briefly or are omitted from this chapter as they are covered in specialized chapters elsewhere in this Handbook. In fact, the underlying theme of this chapter from a *microscopic* viewpoint is the intrinsic optical processes involving valence and conduction electrons. This limitation means that the discussion of the complex dielectric function itself will be confined to the range of wavelengths where electronic transitions occur between (or among) valence and conduction band states (i.e., states lying within ~±5 eV of the Fermi level). This range is spanned by conventional spectroscopic ellipsometers operating from near-infrared to ultraviolet wavelengths. Measurements of

the complex dielectric function in the infrared range yield information on atomic/molecular vibrations in semiconductors and insulators and on free electron transitions in metals. Only the latter optical processes will be treated here; atomic and molecular vibrations will be treated in Chapter 11: Infrared Ellipsometry. The dielectric function from the deep ultraviolet to the x-ray range provides information on the electronic core levels in the solid and will also not be discussed in this chapter. The optical responses of materials at shorter wavelengths than those covered here ($\lambda < 200$ nm) will be treated in Chapter 10: Ultraviolet Ellipsometry. Finally, with regard to this introduction as well as to the applications presented in Section 2.5, the emphasis rests squarely on isotropic materials. In general, anisotropic materials pose special problems in measurement and data analysis (see Chapter 3: Analysis of Ellipsometric Data; Chapter 9: Anisotropic Materials). Once the complex dielectric tensor has been obtained in diagonal form, however, each of the principal components of this tensor can be modeled using the same physical principles as described in Sections 2.2 through 2.4. Models of the complex optical activity tensor are beyond the scope of this Handbook.

2.2 Propagation of Light in Solids

2.2.1 Optically Isotropic Solids and the Complex Dielectric Function

In describing the propagation of light waves in solid media, an appropriate starting point is the macroscopic version of Maxwell's Equations. These equations can be found in a number of textbooks on electrodynamics and optics,[37–39] and are given by:

$$\vec{\nabla} \cdot \vec{E} = \frac{\rho_{tot}}{\varepsilon_0}, \tag{2.1a}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{2.1b}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{2.1c}$$

$$\vec{\nabla} \times \vec{B} = \mu_0 \left(\varepsilon_0 \frac{\partial \vec{E}}{\partial t} + \vec{J}_{tot} \right). \tag{2.1d}$$

In Eqs. (2.1a) and (2.1d), ε_0 and μ_0 are real constants, the permittivity and permeability of free space, respectively, given by $\varepsilon_0 = 8.8542 \times 10^{-12}$ C^2/Nm^2 and $\mu_0 = (4\pi) \times 10^{-7}$ Ns^2/C^2. These constants are related to the speed of light in vacuum c by $\mu_0\varepsilon_0 = c^{-2}$. In addition, $\vec{E}$, $\vec{B}$, ρ_{tot}, and $\vec{J}_{tot}$ in Eqs. (2.1) are the macroscopic electric field and magnetic induction associated with the wave and the macroscopic total charge and current densities in the solid, respectively. *In this chapter, System International (SI) units are employed throughout;* for useful conversion tables see Ref. 39.

The macroscopic fields and charge or current densities are obtained by averaging the corresponding microscopic quantities over a volume ΔV having linear dimensions of ~10 nm. Such dimensions are small compared to the wavelength range of interest for the light waves, ~200 to 1000 nm, but large compared to atomic dimensions, 0.2 nm. For example, the macroscopic electric field and total charge density are obtained by:

$$\vec{E}(\vec{r}) = \frac{1}{\Delta V}\int_{\Delta V} \vec{E}_{loc}(\vec{r} + \vec{r}')d^3\vec{r}', \tag{2.2a}$$

$$\rho_{tot}(\vec{r}) = \frac{1}{\Delta V}\int_{\Delta V} \rho_{micro}(\vec{r} + \vec{r}')d^3\vec{r}', \tag{2.2b}$$

where $\vec{E}_{loc}$ and ρ_{micro} are the microscopic electric field and charge density that vary on the atomic scale. Similar equations also relate $\vec{B}_{loc}$ to $\vec{B}$ and $\vec{J}_{micro}$ to $\vec{J}_{tot}$. In a classical model, ρ_{micro} and $\vec{J}_{micro}$ are calculated by summing contributions from the individual electrons and atomic nuclei. In a quantum mechanical model, the electron contribution must be calculated from the electronic wavefunction $\psi(\vec{r})$ of the system, i.e., $\rho_{micro}(\text{electrons}) = -e\psi^*(\vec{r})\psi(\vec{r})$, where $-e$ is the charge of the electron (see, e.g., Refs. 56 and 57).

Equations (2.1) ultimately need to be reformulated by introducing the macroscopic optical properties of the solid.[20] As the first step in this procedure, the macroscopic total charge and current densities can be separated into different components, depending on their origin. The charge density can be separated into polarizable charges (ρ_{pol}), i.e., the balanced charges due to the electrons and the background of atomic nuclei, and external charges (ρ_{ext}) introduced into the electrically neutral solid. The current density can be separated into bound currents ($\vec{J}_{bound}$), due to the localized motion of electrons bound to nuclei, and free currents ($\vec{J}_{free}$) due to the extended motion of electrons through the solid.

The first equalities from the left in the following two equations describe these separate components:

$$\rho_{tot} = \rho_{pol} + \rho_{ext} = (-\vec{\nabla} \cdot \vec{P}) + \rho_{ext}, \tag{2.3a}$$

$$\vec{J}_{tot} = \vec{J}_{bound} + \vec{J}_{free} = \frac{\partial \vec{P}}{\partial t} + (\vec{\nabla} \times \vec{M}) + (\vec{J}_{cond} + \vec{J}_{ext}). \tag{2.3b}$$

The second equality of Eq. (2.3a) describes the macroscopic polarizable charge density ρ_{pol} in terms of a spatially nonuniform electric dipole moment per unit volume, or polarization $\vec{P}$. Such polarization can be visualized as a displacement of the electronic clouds with respect to the nuclei. The second equality in Eq. (2.3b) describes the macroscopic bound current density $\vec{J}_{bound}$ in terms of contributions from a time-dependent polarization $\vec{P}$ and from a spatially nonuniform magnetic dipole moment per unit volume, or magnetization $\vec{M}$. In nonmagnetic solids, weak magnetization can result from the interaction of the local $\vec{B}$-field with the electron spins and/or the orbital motion of electrons (e.g., Landau diamagnetism in the case of the orbital motion of free electrons). Such effects on the optical properties of the solid are small and can be neglected under most circumstances. The effect of the much stronger $\vec{M}$-field in a saturated ferromagnetic solid on its optical properties can be significant; however, nonmagnetic solids are assumed in this article. The free current density $\vec{J}_{free}$ in Eq. (2.3b) includes the currents due to conduction electrons ($\vec{J}_{cond}$) in metals and semiconductors as well as external currents ($\vec{J}_{ext}$). Note that the conduction electrons do not generate a net macroscopic negative charge in Eq. (2.3a) due to the balance of the positive charge background of the nuclei.

By defining two new vector fields, the displacement, $\vec{D}$, and the magnetic field intensity, $\vec{H}$, given by:

$$\vec{D} = \varepsilon_0 \vec{E} + \vec{P}, \tag{2.4a}$$

$$\vec{H} = \frac{1}{\mu_0} \vec{B} - \vec{M}, \tag{2.4b}$$

the macroscopic version of Maxwell's Equations can be rewritten as:

$$\vec{\nabla} \cdot \vec{D} = \rho_{ext}, \tag{2.5a}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{2.5b}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{2.5c}$$

$$\vec{\nabla} \times \vec{H} = \frac{\partial \vec{D}}{\partial t} + (\vec{J}_{cond} + \vec{J}_{ext}). \tag{2.5d}$$

Next, the macroscopic optical properties of the solid can be introduced through the constitutive relationships:

$$\vec{P} = \varepsilon_0 \chi_{e,r} \vec{E}, \tag{2.6a}$$

$$\vec{M} = \chi_{m,r} \vec{H}, \tag{2.6b}$$

$$\vec{J}_{cond} = 4\pi\varepsilon_0 \sigma_r \vec{E}. \tag{2.6c}$$

Here $\chi_{e,r}$, $\chi_{m,r}$ and σ_r are, respectively, the real (r) electric (e) and magnetic (m) susceptibilities and the real *optical* conductivity (the latter in units of s^{-1}), which all depend on the frequency of the light wave. In fact, the prefactor of $\vec{E}$ in Eq. (2.6c), $4\pi\varepsilon_0\sigma_r$ is the conductivity expressed in SI units $[(\Omega\ m)^{-1}]$. In Eqs. (2.6), it is assumed that the response to the fields is linear and the solid is optically isotropic. A discussion of wave propagation in optically anisotropic solids is provided in Section 2.2.2 . Two frequency-dependent parameters alternative to the susceptibilities can also be used to describe the medium. These are defined by:

$$\vec{D} = \varepsilon_0 \varepsilon_r \vec{E}, \tag{2.7a}$$

$$\vec{B} = \mu_0 \mu_r \vec{H}, \tag{2.7b}$$

where ε_r is the real dielectric function and μ_r is the real magnetic permeability. These dimensionless parameters are related to the real susceptibilities by: $\varepsilon_r = 1 + \chi_{e,r}$ and $\mu_r = 1 + \chi_{m,r}$. The focus of this article will be on nonmagnetic solids for which $\chi_{m,r} = 0$ and $\mu_r = 1$. In ferromagnetic solids one can still set $\mu_r \approx 1$ for the high frequency field associated with the light wave; however, one must include the effect of the permanent magnetization on the optical properties.

By using Eqs. (2.6c) and (2.7a–b) to eliminate $\vec{D}$, $\vec{H}$, and $\vec{J}_{cond}$ in Eqs. (2.5), one arrives at a re-expression of the macroscopic Maxwell's

Equations that explicitly includes the macroscopic properties of the material (i.e., σ_r, ε_r, μ_r).[20,37–39] However, the new equations are no longer exact since Eqs. (2.6–2.7) describe only an approximate response of the solid to the fields. Assuming no external charges or currents, and an isotropic, homogeneous (i.e., no spatial variations in σ_r, ε_r, and μ_r), and nonmagnetic (i.e., $\vec{M} = 0$, $\mu_r = 1$) material, then these new expressions are given by:

$$\vec{\nabla} \cdot \vec{E} = 0, \tag{2.8a}$$

$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{2.8b}$$

$$\vec{\nabla} \cdot \vec{B} = 0, \tag{2.8c}$$

$$\vec{\nabla} \times \vec{B} = \frac{\varepsilon_r}{c^2}\frac{\partial \vec{E}}{\partial t} + \frac{4\pi\sigma_r}{c^2}\vec{E}. \tag{2.8d}$$

Applying Eq. (2.8d) to eliminate $\vec{B}$ from Eq. (2.8b), the following wave equation is obtained for the $\vec{E}$-field:

$$\vec{\nabla}^2\vec{E} = \frac{\varepsilon_r}{c^2}\frac{\partial^2 \vec{E}}{\partial t^2} + \frac{4\pi\sigma_r}{c^2}\frac{\partial \vec{E}}{\partial t}. \tag{2.9}$$

From Eqs. (2.8a) and (2.9), one finds that the solutions for $\vec{E}$ are transverse plane waves that describe the propagation of the light through the solid in terms of the (real) optical property parameters ε_r and σ_r. These solutions can be expressed in the following complex harmonic form:

$$\vec{E}(\vec{r},t) = \vec{E}_0 \exp[i(\vec{q} \cdot \vec{r} - \omega t)]. \tag{2.10a}$$

Here, a single frequency component ω appears, based on the assumption that the light wave in the medium is monochromatic. In addition, the complex wavevector $\vec{q}$ is perpendicular to the field amplitude vector $\vec{E}_0$ and obeys:

$$q^2 = \left(\frac{\omega}{c}\right)^2\left[\varepsilon_r + i\left(\frac{4\pi\sigma_r}{\omega}\right)\right] = \left(\frac{\omega}{c}\right)^2 N^2. \tag{2.10b}$$

A complex wavevector is required to account for dissipation of energy via the optical conductivity term at the far right in Eqs. (2.8d) and (2.9). In

Eq. (2.10b), ε_r and σ_r denote the real dielectric function and real optical conductivity at the light wave frequency ω. N is the complex index of refraction at the frequency ω, given by:

$$N = n + i\kappa = \sqrt{\varepsilon_r + i\left(\frac{4\pi\sigma_r}{\omega}\right)}. \tag{2.10c}$$

In the first equality, n is the (real) index of refraction, and κ is the extinction coefficient.

With this new description of the optical properties of the solid, the $\vec{E}$-field solution of Eqs. (2.10) can now be written

$$\vec{E}(\vec{r},t) = \vec{E}_0 \exp\left(-\frac{\omega\kappa z}{c}\right) \exp\left\{i\left(\frac{\omega n z}{c} - \omega t\right)\right\}, \tag{2.11}$$

where it is assumed that the wavevector $\vec{q}$, defining the direction of propagation, is parallel to the z-axis. Thus, the electric field vector of Eq. (2.11), confined to the x-y plane, can be expressed as a wave propagating within the medium having a phase velocity $v_p = c/n$, a wavelength $\lambda_m = 2\pi c/\omega n$, and an exponentially decaying amplitude with decay length (at $1/e$) of $\omega\kappa/c$. The irradiance I, or the time average of the energy crossing a unit area normal to $\vec{q}$ per unit time for an isotropic solid, is given by

$$I(z) = \frac{\varepsilon_0 \varepsilon_r v_p}{2} E_0^2 \exp\left(-\frac{2\omega\kappa z}{c}\right).$$

The absorption coefficient α describes the decay of the irradiance due to energy dissipation. Thus,

$$I(z) = I_0 \exp(-\alpha z), \tag{2.12a}$$

$$\alpha = \frac{2\omega\kappa}{c} = \frac{4\pi\kappa}{\lambda}, \tag{2.12b}$$

$$I_0 = \frac{\varepsilon_0 \varepsilon_r v_p}{2} E_0^2, \tag{2.12c}$$

where $I_0 = I(z = 0)$. In Eq. (2.12b), $\lambda = 2\pi c/\omega = n\lambda_m$ is the wavelength for a light wave of the same frequency sustained in vacuum. Equations (2.12) do not apply when $n = 0$ and $\kappa \neq 0$, e.g., for a perfect metal. In this special case, $I(z) = 0$, and since no net energy can be transmitted into the material from vacuum, no energy is dissipated.

In addition to the complex index of refraction, the complex dielectric function ε is commonly used to define the macroscopic optical properties of solids, where

$$\varepsilon = N^2 = \varepsilon_r + i\left(\frac{4\pi\sigma_r}{\omega}\right). \tag{2.13a}$$

The real and imaginary parts of ε, written as ε_1 and ε_2 with $\varepsilon = \varepsilon_1 + i\varepsilon_2$, are given by:

$$\varepsilon_1 = \varepsilon_r = n^2 - \kappa^2, \tag{2.13b}$$

$$\varepsilon_2 = \frac{4\pi\sigma_r}{\omega} = 2n\kappa. \tag{2.13c}$$

Inverting these equations yields:

$$n = \sqrt{\frac{\sqrt{\varepsilon_1^2 + \varepsilon_2^2} + \varepsilon_1}{2}}, \tag{2.13d}$$

$$\kappa = \sqrt{\frac{\sqrt{\varepsilon_1^2 + \varepsilon_2^2} - \varepsilon_1}{2}}. \tag{2.13e}$$

The complex dielectric function can enter directly into the wave equation solution by starting with Maxwell's Equations [Eqs. (2.5)] and redefining a displacement field $\vec{D}$ that is explicitly complex. The complex $\vec{D}$ incorporates both the real displacement field of Eq. (2.7a), now denoted $\vec{D}_r$, and the real current density of Eq. (2.6c) due to conduction electrons, now denoted $\vec{J}_{cond,r}$:

$$\vec{D} = \varepsilon_0\varepsilon\vec{E} = \vec{D}_r + \frac{i}{\omega}\vec{J}_{cond,r}, \tag{2.14}$$

as described in Ref. 20. This also leads to an explicitly complex polarization vector and complex electric susceptibility χ_e in accordance with Eq. (2.6a):

$$\vec{P} = \varepsilon_0\chi_e\vec{E} = \varepsilon_0(\varepsilon - 1)\vec{E}, \tag{2.15}$$

where $\chi_e = \varepsilon - 1 = (\varepsilon_r - 1) + i(4\pi\sigma_r/\omega)$.

A third definition of the optical properties often used in the study of metals is the complex optical conductivity σ, given in terms of the

complex dielectric function ε and the complex electric susceptibility χ_e by:

$$\sigma = \frac{i\omega}{4\pi}(1 - \varepsilon) = -\frac{i\omega}{4\pi}\chi_e. \tag{2.16a}$$

The real and imaginary parts of σ, written as σ_1 and σ_2 with $\sigma = \sigma_1 + i\sigma_2$, are given by:

$$\sigma_1 = \frac{\omega\varepsilon_2}{4\pi} = \sigma_r, \tag{2.16b}$$

$$\sigma_2 = \frac{\omega(1 - \varepsilon_1)}{4\pi} = \frac{\omega(1 - \varepsilon_r)}{4\pi} = -\frac{\omega\chi_{e,r}}{4\pi}. \tag{2.16c}$$

The complex optical conductivity develops naturally in the solution to the wave equation by defining a new conduction electron current in Maxwell's Equations [Eqs. (2.5)] that is explicitly complex and incorporates the real polarization of Eq. (2.6a), now denoted $\vec{P}_r$, along with the real conduction electron current of Eqs. (2.6c) and (2.14):

$$\vec{J}_{cond} = 4\pi\varepsilon_0\sigma\vec{E} = \vec{J}_{cond,r} - i\omega\vec{P}_r. \tag{2.17}$$

In theory, experiments can be devised to measure N for bulk solids based on Eqs. (2.11) and (2.12). For example, one can conceive of illuminating a disc of material having two optically polished plane-parallel surfaces with a monochromatic plane light wave of known frequency ω or vacuum wavelength $\lambda = 2\pi c/\omega$. The optical properties n and κ can then be obtained by measuring the phase shift (relative to a reference wave; for n) and the irradiance change (for κ) that occur when the wave traverses the disc. As a practical matter, however, the measurements must be made external to the solid; thus, one must account for the effects of the interfaces between the solid and ambient in the measurements.[58,59] With this method, the maximum measurable value of α is 10^2 cm^{-1} for polished bulk materials with $d \sim 0.1$ cm, owing to the near opacity of the sample for higher α values. As a result, a reflection measurement such as ellipsometry, rather than a transmission measurement, must be used to obtain the much larger α values ($>10^3$ cm^{-1}) associated with intraband and direct interband transitions (see Sections 2.4 and 2.5). In addition, measurements of the large phase shifts relative to vacuum that occur in such a sample geometry are impractical for solids with typical values of $n = 1.4$ to 4.0. In this case, refraction or reflection measurements are preferred for determining n, depending on whether the material is transmitting or is

opaque, respectively. Further information regarding the physics of reflection, the ellipsometry measurement, and the analysis for determination of (n, κ) [or $(\varepsilon_1, \varepsilon_2)$] is provided in other chapters of the Handbook, as well as in Section 2.1 of this chapter.

2.2.2 Optically Anisotropic Solids and the Dielectric Tensor

For the discussion of Section 2.2.1, the medium was assumed to be optically homogeneous and isotropic. In other words, its dielectric function was assumed to be independent of position in the solid (on a macroscopic scale) and also independent of the orientation of the light wave's orthogonal vectors $(\vec{D}, \vec{H}, \vec{q})$. The assumption of optical isotropy is only strictly valid for crystalline solids that are cubic and for amorphous and nanocrystalline materials that are microstructurally isotropic. Otherwise the solids are optically anisotropic, in general, and may exhibit optical phenomena including linear and circular birefringence and dichroism. For linear birefringence, two light waves traveling with the same propagation vector $\vec{q}$ in the solid, but with orthogonal linear polarization modes (i.e., polarization states that travel unchanged within the solid), denoted by the displacement vectors $\vec{D}_1$ and $\vec{D}_2$, experience different indices of refraction n_1 and n_2. As a result, the two waves travel at different phase speeds v_{p1} and v_{p2} in the solid. Anisotropic solids possessing a helical arrangement of atoms or a helical nanostructure may exhibit circular birefringence. For this case, left- and right-elliptical polarization states are the orthogonal modes, in general, and these experience different indices of refraction and travel at different phase speeds. Linear and circular dichroism may accompany linear and circular birefringence, respectively, in absorbing solids and are characterized by different extinction coefficients for the two orthogonal polarization modes. The indices of refraction and extinction coefficients that describe wave propagation in linearly or circularly birefringent and dichroic materials can be combined to form a complex dielectric tensor of second rank, as described below.

In this subsection, the physics of wave propagation in anisotropic solids will be described. First, such a description is informative to the ellipsometry user for the physical insights it provides into the optical phenomena exhibited by anisotropic samples. More detailed information regarding the physics of reflection from a multilayer sample that includes one or more anisotropic media, the ellipsometric measurement of such samples, and the analysis of data for determination of the thicknesses and dielectric tensors is provided in other chapters of the Handbook. Most often the latter analysis involves a matrix method of computation in a

"cookbook" style that provides no physical insights into the associated optical phenomena. Second, a description of wave propagation in non-absorbing anisotropic solids is informative because it provides insights into the operation of the basic components of the ellipsometer, including polarizers, compensators, and phase modulators. For the modulators, optical anisotropy and its effects such as linear and circular birefringence and dichroism can be induced by external perturbations such as electric fields (for Pockels cell modulators), magnetic fields (for Faraday cell modulators), or mechanical stress (for piezo-optic modulators).

2.2.2.1 Non-Absorbing Solids that are not Optically Active

The scalar dielectric function for an optically isotropic solid, defined by Eq. (2.7a), becomes a second-rank tensor in an anisotropic solid.[38,40,41] For a homogeneous solid that is non-absorbing, nonmagnetic, and not optically active, one can write:

$$D_x = \varepsilon_0(\varepsilon_{r11}E_x + \varepsilon_{r12}E_y + \varepsilon_{r13}E_z), \tag{2.18a}$$

$$D_y = \varepsilon_0(\varepsilon_{r21}E_x + \varepsilon_{r22}E_y + \varepsilon_{r23}E_z), \tag{2.18b}$$

$$D_z = \varepsilon_0(\varepsilon_{r31}E_x + \varepsilon_{r32}E_y + \varepsilon_{r33}E_z), \tag{2.18c}$$

or for simplicity in the different tensor notations:

$$D_i = \varepsilon_0\varepsilon_{rij}E_j, \tag{2.19a}$$

$$\vec{D} = \varepsilon_0[\varepsilon_r]\vec{E}. \tag{2.19b}$$

In Eqs. (2.18) and (2.19), the orthogonal (x, y, z) axes chosen for the reference frame [also designated by indices $i = (1,2,3)$ or $j = (1,2,3)$, respectively] are fixed in the medium. In Eq. (2.19a), the Einstein convention of summing over repeated indices is employed. In Eq. (2.19b), the brackets enclosing the single variable ε_r denote a tensor, in this case, of second rank defined by the elements of a 3×3 matrix.

In restricting the discussion of this part to non-absorbing solids, the tensor elements ε_{rij} are assumed to be real, as indicated by the subscript 'r' in Eqs. (2.18) and (2.19). As a result, this discussion is most useful for understanding the optical characteristics of ellipsometer components such as polarizers, compensators, and phase modulators. For non-absorbing solids, energy conservation requires that $\varepsilon_{rij} = \varepsilon_{rji}$, i.e., that $[\varepsilon_r]$ is symmetric. As a result, it is always possible to select a specific (x, y, z)

coordinate system such that $[\varepsilon_r]$ is diagonal. The (x, y, z) directions for this selection define the principal dielectric axes. For single crystals having orthorhombic and higher symmetries, the principal dielectric axes lie along the symmetry axes of the crystal, and hence their orientation is completely specified in advance, as long as the orientation of the crystal is known. In this case, *at most* three independent quantities, the principal dielectric functions, ε_{rjj} $(j = 1,2,3)$, must be specified in order to define the dielectric tensor and principal axes. In contrast, the absence of symmetry for the triclinic crystal system means that the directions of all three orthogonal principal axes are unspecified even when the crystal orientation is known in advance (assuming that the relation between the crystal axes and the principal dielectric axes had not been previously determined for the crystal). Thus, six independent quantities $[\{\varepsilon_{rjj}, (j = 1,2,3)\}$; plus the coordinate system orientation] must be specified. Table 2-1 provides the number of independent elements needed to define the diagonalized dielectric tensor and the orientation of the principal axes for different crystal systems.[40,41]

In non-absorbing, nonmagnetic solids [with $\vec{H} = \vec{B}$], Maxwell's Equations exhibit linearly polarized, plane-wave solutions requiring that the vectors $(\vec{D}, \vec{H}, \vec{q})$ [or $(\vec{D}, \vec{B}, \vec{q})$] form a right-handed coordinate system. As a result, $\vec{D}$ and $\vec{H}$ always lie in the plane perpendicular to $\vec{q}$. In optically isotropic solids, Eq. (2.7a) shows that $\vec{D}$ and $\vec{E}$ are parallel; so in this case $\vec{E}$ also lies in the same plane perpendicular to $\vec{q}$. In optically anisotropic solids, however, Eq. (2.18) shows that $\vec{D}$ and $\vec{E}$ are *not* necessarily parallel. Because Maxwell's Equations also require that the vector $\vec{E}$ remains perpendicular to $\vec{H}$ [or $\vec{B}$], this implies that $\vec{E}$, $\vec{D}$, and $\vec{q}$ are all coplanar (lying in the plane perpendicular to $\vec{H}$). Hence $\vec{E}$ cannot remain perpendicular to $\vec{q}$, in general, and so the direction of *wave* propagation defined by the propagation vector $\vec{q}$, called the wave normal, is not necessarily parallel to the direction of *energy* propagation defined by the Poynting vector $\vec{S} = \vec{E} \times \vec{H}$, called the ray direction. In fact, the angle ξ between $\vec{q}$ and $\vec{S}$, i.e., between the wave normal and the ray direction, is the same as that between $\vec{D}$ and $\vec{E}$, and so is given by

$$\cos \xi = \frac{\vec{E} \cdot \vec{D}}{|\vec{E}||\vec{D}|}.$$

Since the ray and wavefront exhibit the same frequency and must remain in phase during propagation through the crystal, the speed of the ray, v_u, must exceed that of the wavefront, v_p, by the factor $(1/\cos\xi)$. Thus, $v_u = v_p/\cos\xi$, meaning that the phase speed is the projection of the ray speed on the direction of the wave normal. The wavefront or phase speed

Table 2-1. Optical behavior of transparent solids of different crystal systems. In this table, (*a*, *b*, *c*) designate the principal crystallographic axes which are aligned with the principal dielectric axes (*x*, *y*, *z*) for crystals of orthorhombic and higher symmetry. The semiaxes of the index ellipsoid (see Fig. 2.1) define the principal indices of refraction (n_x, n_y, n_z). For biaxial crystals, (*x*, *y*, *z*) are labelled such that $n_x < n_y < n_z$. For uniaxial crystals, the *z*-axis of the index ellipsoid is the optic axis and is parallel to the *c* crystallographic axis. For these crystals, $n_o = n_x = n_y$ is the ordinary index of refraction, and $n_e = n_z$ is the extraordinary index of refraction.

Crystal systems and point groups	**No. of parameters**	**Nature and orientation of index ellipsoid**	**Optical behavior**	**Variation with λ, temperature, or hydrostatic pressure**	**Point groups exhibiting optical activity**
Triclinic (1, $\bar{1}$)	6	Ellipsoid, principal axes in unspecified directions	Biaxial, optic axes in unspecified directions	Unspecified	1
Monoclinic (*b* axis unique) (2, m, 2/m)	4	Ellipsoid, one principal axis $\parallel$ to *b*, other two $\perp$ to *b* but in unspecified directions	Biaxial, optic axial plane either $\parallel$ or $\perp$ to *b*	Orientation of one principal axis always along *b*	2, m
Orthorhombic (2mm, 222, mmm)	3	Ellipsoid, all three principal axes along *a*, *b*, and *c*	Biaxial, optic axial plane $\parallel$ to *ab*, *bc*, or *ca;* acute bisectrix $\parallel$ to one crystal axis	No change in orientation, but only change in length of principal axes, i.e., n_x, n_y, and n_z	2mm, 222
Trigonal (3, $\bar{3}$, 3m, 32, $\bar{3}$m)	2	Spheroid, with unique axis $\parallel$ to *c*	Uniaxial, optic axis $\parallel$ to *c*	Optic axis always along *c*, but n_o and n_e will vary	3, 32
Tetragonal (4, $\bar{4}$, 4/m, 4mm, $\bar{4}$2m, 422, 4/mmm)	2	Spheroid, with unique axis $\parallel$ to *c*	Uniaxial, optic axis $\parallel$ to *c*	Optic axis always along *c*, but n_o and n_e will vary	$\bar{4}$, 422, $\bar{4}$, 42m
Hexagonal ($\bar{6}$, $\bar{6}$, 6/m, $\bar{6}$m2, 622, 6mm, 6/mmm)	2	Spheroid, with unique axis $\parallel$ to *c*	Uniaxial, optic axis $\parallel$ to *c*	Optic axis always along *c*, but n_o and n_e will vary	6, 622
Cubic (23, m3, $\bar{4}$3m, 432, m3m)	1	Sphere	Isotropic	Always isotropic, but *n* will change	23, 432

is defined as in optically isotropic solids by $v_p = c/n$, where n for anisotropic solids is the index of refraction corresponding to the specific orientations of $\vec{D}$ and $\vec{q}$. In anisotropic solids,

$$v_p = \frac{c\sqrt{\varepsilon_0 \vec{E} \cdot \vec{D}}}{|\vec{D}|},$$

an expression derived starting from the equality $|\vec{D}| = \varepsilon_0 n^2 |\vec{E}| \cos \xi$. In the next paragraphs, further consequences of such wave propagation behavior will be discussed.

The electrical energy density associated with an electromagnetic wave propagating within an optically anisotropic crystal is a constant for a non-absorbing solid.[38,40] This constant is given by:

$$\begin{aligned} u_E &= \frac{\vec{D} \cdot \vec{E}}{2} = \frac{\vec{D} \cdot [a_r]\vec{D}}{2\varepsilon_0} \\ &= \frac{a_{r11}D_x^2 + a_{r22}D_y^2 + a_{r33}D_z^2 + 2a_{r12}D_xD_y + 2a_{r13}D_xD_z + 2a_{r23}D_yD_z}{2\varepsilon_0}, \end{aligned} \tag{2.20a}$$

where $[a_r]$ is the real index (or impermeability) tensor given by $\vec{E} = \frac{1}{\varepsilon_0}[a_r]\vec{D}$ (see, e.g., Ref. 40). Thus, $[a_r]$ is obtained by taking the inverse of the matrix that represents the dielectric tensor, i.e., $[a_r] \equiv [\varepsilon_r]^{-1}$. In the ($x$, y, z) coordinate system of the principal dielectric axes, both $[a_r]$ and $[\varepsilon_r]$ are diagonal with $a_{rjj} = 1/\varepsilon_{rjj}$. In this coordinate system,

$$u_E = \frac{1}{2\varepsilon_0}\left(\frac{D_x^2}{\varepsilon_{r11}} + \frac{D_y^2}{\varepsilon_{r22}} + \frac{D_z^2}{\varepsilon_{r33}}\right). \tag{2.20b}$$

Finally, applying the unitless scaling

$$\vec{R} \rightarrow \frac{\vec{D}}{\sqrt{2\varepsilon_0 u_E}},$$

yields the so-called index ellipsoid (or optical indicatrix):

$$1 = \frac{X^2}{n_x^2} + \frac{Y^2}{n_y^2} + \frac{Z^2}{n_z^2}. \tag{2.20c}$$

Thus, the index ellipsoid is the surface consisting of the locus of $\vec{D}$ vector endpoints, corresponding to an energy density of $(1/2\varepsilon_0)$, and has semi-axes given by the principal indices of refraction, defined by

$$n_x \equiv \sqrt{\varepsilon_{r11}} = \frac{1}{\sqrt{a_{r11}}},$$

$$n_y \equiv \sqrt{\varepsilon_{r22}} = \frac{1}{\sqrt{a_{r22}}},$$

and

$$n_z \equiv \sqrt{\varepsilon_{r33}} = \frac{1}{\sqrt{a_{r33}}}.$$

It should be kept in mind that any coordinate transformations of this ellipsoid are performed via the index tensor $[a_r] \equiv [\varepsilon_r]^{-1}$, as in Eq. (2.20a), rather than via the dielectric tensor $[\varepsilon_r]$.[40]

In fact, the geometric construction of Eq. (2.20c) can be applied to determine the indices of refraction (n_1, n_2) and directions of the displacement vector $(\vec{D}_1, \vec{D}_2)$ associated with the two independent, linearly polarized modes that can propagate along a given wavevector direction $\vec{q}$ (see Fig. 2.1). First, the plane passing through the origin, normal to $\vec{q}$ is defined. This plane intersects the index ellipsoid to form an ellipse, or central section, that provides the index of refraction $n(\vec{D})$ for all possible directions of $\vec{D}$ normal to $\vec{q}$. The minor and major axes of this ellipse

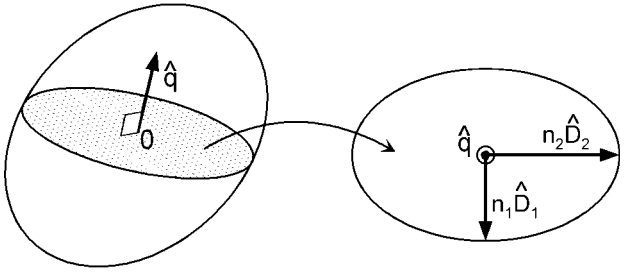


Figure 2.1 Index ellipsoid (left) and its central section (right) for an anisotropic solid demonstrating the procedure used to determine the two possible values of the index of refraction (n_1 and n_2), as well as the directions of the associated orthogonal linear polarization modes ($\vec{D}_1$ and $\vec{D}_2$), given the direction of the wavevector, $\vec{q}$.

define the maximum and minimum phase speeds $v_{p1} = c/n_1$ and $v_{p2} = c/n_2$, respectively, corresponding to the two specific orthogonal directions of $\vec{D}_1$ and $\vec{D}_2$ (or linear polarization directions). Thus, for *any* direction of $\vec{D}$ normal to $\vec{q}$, the displacement can be resolved along these two directions, yielding two orthogonal linearly polarized waves having different phase speeds. When a wave in an optically isotropic medium crosses an interface into an anisotropic medium, the two polarization components will be refracted through different angles, and this is the physical explanation of the unique visual effect of birefringence. This effect is exploited, for example, in MgF_2 Rochon prism polarizers that spatially separate the orthogonal linearly polarized waves.

An approach alternative to that of Eq. (2.20c) and Fig. 2.1 for understanding wave propagation in optically anisotropic crystals provides a direct connection between the direction of $\vec{q}$ and the indices of refraction for the two orthogonally polarized waves.[38,40] Considering the linearly polarized plane-wave solutions to Maxwell's Equations in an anisotropic, non-absorbing solid that is not optically active, the index of refraction n (defining the phase speed, c/n) and the wave propagation direction $\vec{q}/q$ are related by Fresnel's equation:

$$\frac{1}{n^2} = \sum_{j=1}^{3} \left(\frac{q_j}{q}\right)^2 \frac{1}{n^2 - n_{jj}^2}. \tag{2.21}$$

As usual, the principal-axes coordinate system must be used, and the summation j is over the (x, y, z) components of $\vec{q}/q$, designated in Eq. (2.21) as q_j/q (j = 1,2,3), and over the principal indices of refraction (n_x, n_y, n_z), designated as n_{jj} (j = 1,2,3). This equation has two positive solutions for the index of refraction, n, corresponding to the two orthogonal linearly-polarized waves of different phase speeds having wave normals in the direction of $\vec{q}/q = (q_x/q, q_y/q, q_z/q)$. Each of the two solutions for $n(\vec{q})$ can be visualized as a surface in propagation-vector space such that the vector from the origin to a point on the surface in the direction of $\vec{q}/q$ has length n. These surfaces, along with the corresponding surfaces given by $v_p(\vec{q}) = c/n(\vec{q})$, are collectively referred to as normal surfaces. Given the direction of the Poynting vector $\vec{S}$, similar surfaces called ray surfaces can be constructed. These surfaces can be obtained from Eq. (2.21) by making the replacements $(n, n_{jj}, q_j, q) \rightarrow (n_u^{-1}, n_{jj}^{-1}, S_j, S)$, and they provide two possible solutions for the ray or energy index, $n_u(\vec{S})$, and the ray speed, related by $v_u(\vec{S}) = c/n_u(\vec{S})$. For a given wave defined by $\vec{q}$ and $\vec{S}$, the normal and ray surface solutions for the index are related by $n_u = n\cos\xi$.

For crystal systems of orthorhombic and lower symmetry having three unequal principal indices of refraction, the normal surfaces given by

Eq. (2.21) intersect at four coplanar points.[38,40] In this case, two optic axes can be defined, each passing through the origin and one pair of intersecting points. For propagation vectors along these two axes, the two orthogonal linearly polarized waves experience the same index of refraction and therefore exhibit the same phase speed. Thus, the crystals exhibiting this behavior are called "optically biaxial," as shown in Table 2-1. In an optically biaxial crystal, the principal coordinate axes are labeled such that the principal indices of refraction obey $n_x < n_y < n_z$. With this labeling convention, the optic axes lie in the x-z plane. A cross-section of the intersecting normal surfaces, defined in terms of $n(\vec{q})$ whereby $\vec{q}$ lies within the x-z plane, is shown in Fig. 2.2(a) for a biaxial crystal. One cross section is circular, given by $n = n_y$, corresponding to linearly polarized waves with $\vec{D}$ parallel to the y-axis. The other is elliptical, given by:

$$\frac{1}{n^2} = \frac{\cos^2\phi}{n_x^2} + \frac{\sin^2\phi}{n_z^2}, \tag{2.22a}$$

corresponding to linearly polarized waves with $\vec{D}$ perpendicular to the y-axis.[41] In Eq. (2.22a), ϕ is the angle between the z-axis and the propagation vector direction in the x-z plane. It should be noted that the cross sections of the normal surfaces in the x-y and y-z planes do not intersect, meaning that there are two unequal n values for the orthogonal linearly polarized modes when the propagation vector $\vec{q}$ lies in these two planes.

For the higher symmetry optically anisotropic crystals (see Table 2-1) there are two distinct principal indices of refraction, the ordinary index, $n_o \equiv n_x = n_y$, and the extraordinary index, $n_e \equiv n_z$. In this case, one normal surface is a sphere that defines the directionally independent ordinary index of refraction, given by $n = n_o$. The other is a spheroid having an axis of symmetry along the z-direction, given by:

$$\frac{1}{n^2} = \frac{\sin^2\phi}{n_e^2} + \frac{\cos^2\phi}{n_o^2}, \tag{2.22b}$$

where ϕ is the angle of the propagation vector in the x-z plane defined relative to the z-axis.[40,41] Such crystals are called "optically uniaxial", because the normal surfaces now intersect at only two points at $\phi = 0°$ in the x-z plane. The line passing through the origin and joining the two intersecting points defines the optic axis, i.e., the z-axis in this case. If $n_o < n_e$, then the crystal is said to be positive uniaxial, an example being α-quartz (with $n_o = 1.5443$ and $n_e = 1.5534$ at $\lambda = 589.3$ nm), often used to fabricate Rochon polarizers and Babinet-Soleil compensators. If $n_o > n_e$, it is said to be negative uniaxial, an example being calcite ($CaCO_3$) (with

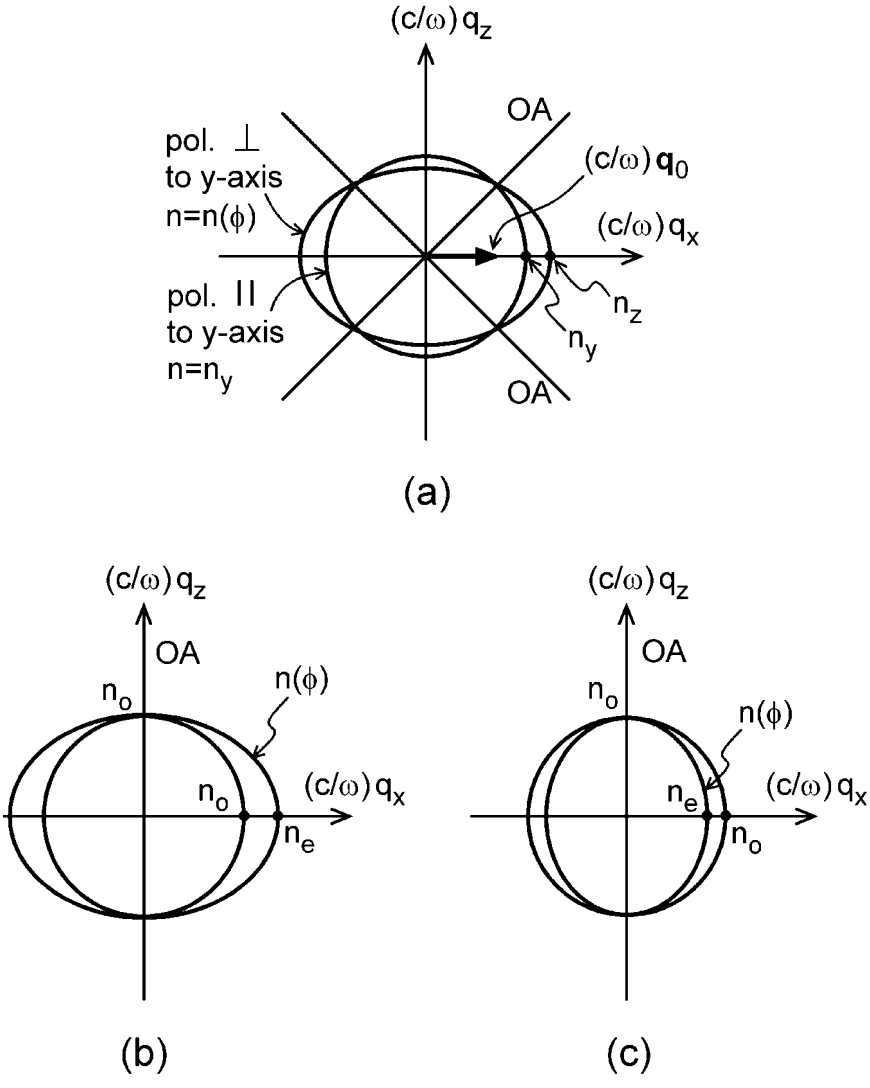


Figure 2.2 Intersections of the two normal surfaces of Eq. (2.21) with the *x*-*z* plane for (a) a biaxial crystal ($n_x < n_y < n_z$); (b) a positive uniaxial crystal ($n_o = n_x = n_y < n_e = n_z$); and (c) a negative uniaxial crystal ($n_e < n_o$). The optic axes, defined by the intersection points of the normal surfaces, are also indicated in each case (by OA). For the biaxial crystal in (a), the principal indices of refraction are indicated for $\vec{q} = \vec{q}_0$ along $\vec{x}$.

$n_o = 1.6584$ and $n_e = 1.4864$ at $\lambda = 589.3$ nm), used to fabricate Glan-Thompson or Glan-Foucault polarizers. The intersections of the *x*-*z* plane with the normal surfaces of positive and negative uniaxial crystals are shown in Figs. 2.2(b) and 2.2(c), respectively.

Finally, the phenomenon of plane wave refraction at an interface between an isotropic ambient (assumed to be air with $n_a = 1$) and an optically uniaxial anisotropic solid will be considered briefly. The boundary conditions specify two angles of refraction, θ_{t1} and θ_{t2}, given by $\sin\theta_{t1} = \sin\theta_i/n_j$ ($j = 1, 2$), where θ_i is the angle of incidence and n_j are the two indices of refraction associated with the orthogonal linearly polarized components $\vec{D}_j$ in the solid. In addition, the boundary conditions specify that the propagation vectors of the two transmitted waves lie within the plane of incidence. For the linearly polarized component associated with the ordinary wave, with $\vec{D}_1$ perpendicular to the optic axis in the solid, the phase speed is independent of the propagation direction (i.e., the normal surface is spherical). Thus, the behavior for θ_{t1} as a function of θ_i is in accordance with Snell's Law for an optically isotropic solid with $n_1 = n_o$. For the orthogonal linearly polarized component $\vec{D}_2$ associated with the extraordinary wave, the phase speed depends on the propagation direction (i.e., the normal surface is elliptical). Thus, the behavior for θ_{t2} as a function of θ_i is in accordance with a much more complicated expression for $\sin\theta_{t2}$ that must include $n(\phi)$ from Eq. (2.22b).

2.2.2.2 Absorbing Solids that are not Optically Active

In optically isotropic solids, energy dissipation due to conduction electron currents can be accounted for by introducing a complex dielectric function (see Section 2.2.1). Following the procedure used in the development of Eq. (2.13a), Eq. (2.19b) for non-absorbing anisotropic solids that are not optically active can also be extended to the case of absorbing solids by introducing a complex dielectric tensor $[\varepsilon]$ with components given by:

$$\varepsilon_{ij} = \varepsilon_{rij} + i\eta_{rij}, \tag{2.23}$$

where the real tensors $[\varepsilon_r]$ and $[\eta_r] \equiv (4\pi/\omega)[\sigma_r]$ are both symmetric in the absence of optical activity.[40] Thus, both tensors can be diagonalized, but the principal axes only coincide when required by the symmetry of the crystal (i.e., for orthorhombic and higher symmetries).

As noted in Section 2.2.2.1, the index ellipsoid given by Eq. (2.20c) for non-absorbing solids relates the polarization mode direction and the phase speed of the wave, and its semiaxes can be defined in terms of the

components of the diagonalized (real) index tensor $[a_r]$. The index tensor is complex in absorbing solids, however, and is written as $[a] = [a_r] - i[b_r] = [\varepsilon]^{-1}$ with components given by $a_{ij} = a_{rij} - ib_{rij}$, where $[b_r]$ denotes the absorption tensor. The ellipsoids representing $[a_r]$ and $[b_r]$ are called the index (as in the case of non-absorbing solids) and the absorption ellipsoids, respectively. Again, the principal axes of these two ellipsoids only coincide when required by the symmetry of the crystal.

In general, the nature of the waves propagating along any direction $\vec{q}$ of the crystal depend on the central sections of the index and absorption ellipsoids, normal to $\vec{q}$. The directions of the major and minor axes of the central sections are called the principal directions of linear birefringence (for the index ellipsoid) and linear dichroism (for the absorption ellipsoid). In many cases, absorption is weak and the tensor elements $[\eta_r]$ are small compared to unity and to the corresponding elements of $[\varepsilon_r]$. In these cases, $[a_r] \approx [\varepsilon_r]^{-1}$ and $[b_r] \approx [\varepsilon_r]^{-1}[\eta_r][\varepsilon_r]^{-1}$ so that the principal axes of $[\varepsilon_r]$ and $[a_r]$ coincide, whereas those of $[\eta_r]$ and $[b_r]$ do not, in general. As a result, the wave propagation behavior in most directions (i.e., sufficiently far from the optic axes in a biaxial crystal) is similar to that for non-absorbing crystals. In particular, the polarization modes that travel unchanged through the crystal are linear and orthogonal, and their directions, phase speeds, and the associated indices of refraction are all determined from the index ellipsoid as in Fig. 2.1. Furthermore, the extinction coefficient that describes the decay of the field vectors is determined from the expression $\kappa = b_r n^3/2$, where ${b_r}^{-1/2}$ and n define the magnitudes of the real (unitless) radius vectors for the absorption and index ellipsoids in the direction of the corresponding polarization mode. In the general case of arbitrary absorption magnitude and propagation direction, the situation is far more complex. In this case, non-diagonal index and absorption tensors are used to compute the two polarization modes that travel unchanged through the crystal for a given propagation direction, as well as the indices of refraction and extinction coefficients for these two polarizations. For the general case, the two polarization states are elliptical with their major axes crossed, but of the same handedness. Thus, the modes are not orthogonal in contrast to the situation for a non-absorbing crystal.

2.2.2.3 Non-Absorbing, Optically Active Solids

Optical activity, or optical rotatory power, describes the phenomenon in which the polarization direction of a linearly polarized light wave rotates as the wave travels within the medium[42,60] Typical magnitudes of this effect range from 20° to 2000° per cm of path length in optically active solids, and the effect can occur in both isotropic media and crystalline

solids. The last column of Table 2-1 lists the crystal point groups that exhibit optical activity. Depending on the crystal symmetry, solids from any of the crystal systems can be optically active, including cubic (e.g., $NaClO_3$, $Bi_{12}GeO_{20}$), optically uniaxial (e.g., Se, Te, α-quartz, HgS), and optically biaxial (e.g., cane sugar and Rochelle salt) crystals. For the anisotropic crystals, the simple polarization rotation effect is observed along the unique optic-axial directions.

Optical activity is generally treated as arising from circular birefringence, i.e., a difference between the indices of refraction n_L and n_R for left and right circularly polarized light waves, respectively, with a typical magnitude of $|n_L - n_R| \sim 10^{-5}$. Optical activity arises from an intrinsic helical arrangement of the atoms in the crystal. For a medium having the property of a screw axis, the displacement vector $\vec{D}$ at a given location must depend not only on the electric field vector $\vec{E}$ but also on its spatial derivatives. Hence, for plane-wave propagation in an optically active medium, Eq. (2.19b) is replaced by:

$$\vec{D} = \varepsilon_0\{[\varepsilon_r]\vec{E} + i\vec{G}_{or} \times \vec{E}\} \tag{2.24a}$$

$$= \varepsilon_0\{[\varepsilon_r] + i[G_{or}]\}\vec{E}. \tag{2.24b}$$

In Eq. (2.24a), $[\varepsilon_r]$ is the real dielectric tensor that would exist in the absence of the optical activity and $\vec{G}_{or}$ is the real (r) gyration vector.[40,41] Equation (2.24b) follows since the vector product $\vec{G}_{or} \times \vec{E}$ can always be represented as the product of a real antisymmetric tensor $[G_{or}]$ and the electric field $\vec{E}$. The following relationships must hold between the tensor elements of $[G_{or}]$ and the vector components of $\vec{G}_{or}$: $G_{or23} = -G_{or32} = -G_{orx}$; $G_{or31} = -G_{or13} = -G_{ory}$; $G_{or12} = -G_{or21} = -G_{orz}$; and $G_{orjj} = 0$ $(j = 1,2,3)$. Equation (2.24b) leads to a new complex dielectric tensor for non-absorbing, optically active (*OA*) solids, given by $[\varepsilon^{OA}] = [\varepsilon_r] + i[G_{or}]$ and represented by a Hermitian matrix, i.e., $\varepsilon_{ij}^{OA} = (\varepsilon_{ji}^{OA})^*$, where the asterisk denotes the complex conjugate operation. From this result, it can be shown that the divergence of the Poynting vector, $\vec{\nabla} \cdot \vec{S} = \vec{\nabla} \cdot (\vec{E} \times \vec{H})$, vanishes so that there is no energy dissipation. The gyration vector $\vec{G}_{or}$ can be expressed further in terms of the wave propagation vector according to $\vec{G}_{or} = [g_{or}]\vec{q}$, where $[g_{or}]$ is a second rank gyration tensor described in general by a real 3×3 matrix.

Equations analogous to Eqs. (2.24), but employing the index tensor, can be written as:

$$\vec{E} = \varepsilon_0^{-1}\{[a_r]\vec{D} - i\vec{\Gamma}_{or} \times \vec{D}\} \tag{2.25a}$$

$$= \varepsilon_0^{-1}\{[a_r] - i[\Gamma_{or}]\}\vec{D}. \tag{2.25b}$$

In Eq. (2.25a), $[a_r]$ is the real index tensor that would exist in the absence of the optical activity and $\vec{\Gamma}_{or}$ is the real optical activity vector.[40] Equation (2.25b) derives from the following relationships between the tensor elements of $[\Gamma_{or}]$ and the vector components of $\vec{\Gamma}_{or}$: $\Gamma_{or23} = -\Gamma_{or32} = -\Gamma_{orx}$; $\Gamma_{or31} = -\Gamma_{or13} = -\Gamma_{ory}$; $\Gamma_{or12} = -\Gamma_{or21} = -\Gamma_{orz}$; and $\Gamma_{orjj} = 0$ ($j = 1,2,3$). Equation (2.25b) leads to a new complex index tensor for non-absorbing, optically active solids, given by $[a^{OA}] = [a_r] - i[\Gamma_{or}]$, again represented by a Hermitian matrix. Like the gyration vector, $\vec{\Gamma}_{or}$ is a function of the wave propagation vector, thus yielding the relation $\vec{\Gamma}_{or} = [\gamma_{or}]\vec{q}$, where $[\gamma_{or}]$ is a second-rank optical activity tensor also described by a 3×3 matrix.

Because $[\Gamma_{or}]_{ij}/[a_r]_{ij} \ll 1$, the normal surfaces of an optically active solid are nearly identical to those without optical activity having $\vec{\Gamma}_{or} = 0$. However, the interpretation of the resulting wave propagation must be modified from the behavior for solids that are not optically active. As an example, consider the case of a uniaxial crystal such as α-quartz. For $\vec{q}$ along the optic axis (i.e., the z-axis), the relevant diagonal index tensor elements are $a_{r11} = a_{r22} \equiv a_r$, and the polarization modes that travel unchanged within the solid are no longer linear and orthogonal, but rather left- and right-circular. In addition, the two normal surfaces no longer make contact, and the resulting indices of refraction for the two modes are given by $(1/n^2) = a_r \pm \Gamma_{orz}$. The rotatory power due to the difference in the two indices of refraction is given by:

$$\rho_{or} = \frac{\pi(n_L - n_R)}{\lambda} = \frac{\pi n^3 \Gamma_{orz}}{\lambda}, \tag{2.26}$$

where n is the average index of refraction, i.e., $n = (n_L + n_R)/2$. As a result, for a linearly polarized wave propagating along the optic axis, the direction of polarization rotates by ρ_{or} [in units of radians per meter in Eq. (2.26)] in the medium. The sign convention most often employed for ρ_{or} in this application has been adopted from the chemical literature and is opposite that typically chosen by physicists. In this convention, optical rotation is clockwise-positive (looking toward the source) for a dextro-rotatory crystal with $n_L > n_R$, and counterclockwise-negative for a levo-rotatory crystal with $n_R > n_L$. The α-quartz crystal exists in both levo- and dextro-rotatory forms with a specific rotatory power of 188°/cm for light with $\lambda = 632.8$ nm, which corresponds to $n^3\Gamma_{orz} = |n_L - n_R| = 6.6 \times 10^{-5}$. Thus, one can determine values of the circular birefringence with very high sensitivity by measuring the rotation that the polarization direction of an incident linearly polarized, monochromatic light wave undergoes when it travels through a known thickness of the solid.

For an optically active biaxial crystal, index splitting similar to that described by Eq. (2.26) is observed for wave propagation along the two optic axes, and again the corresponding polarization modes are left-circular and right-circular. In cubic crystals the same behavior is observed, irrespective of the direction of wave propagation. Wave propagation in optically active, non-absorbing, anisotropic solids in directions other than the optic axes is complicated by the presence of both linear birefringence and optical activity. Under these circumstances the polarization modes that travel unchanged through the solid are left and right elliptical having major axis directions that are parallel to the linearly polarized modes predicted from the index ellipsoid in the absence of optical activity. The ellipticities of the two modes and the associated indices of refraction can be evaluated utilizing coupled-mode theory in which the optical activity is treated as a small perturbation of the dielectric tensor given by $[\Delta\varepsilon_i] = i[G_{or}]$, so that $[\varepsilon^{OA}] = [\varepsilon_r] + [\Delta\varepsilon_i]$[41] An alternative approach is to solve the problem geometrically by the method of superposition using the Poincare sphere.[40]

2.2.2.4 Absorbing, Optically Active Solids

If both absorption and optical activity are present in an anisotropic solid, then the relationships between $\vec{D}$ and $\vec{E}$ take the same forms as Eqs. (2.24) and (2.25), but the real second-rank tensors $[\varepsilon_r]$, $[G_{or}]$, $[g_{or}]$, $[a_r]$, $[\Gamma_{or}]$, and $[\gamma_{or}]$ defined there, as well as the real vectors $\vec{G}_{or}$ and $\vec{\Gamma}_{or}$, become complex. It is convenient to express the wave propagation characteristics of the solid by the four real tensors $[a_r]$, $[b_r]$, $[\gamma_{or}]$, and $[\beta_{or}]$, that describe the complex index tensor $[a]$ and the complex optical activity tensor $[\gamma_o]$, according to $[a] = [a_r] - i[b_r]$ and $[\gamma_o] = [\gamma_{or}] + i[\beta_{or}]$. As in the case of absorbing solids that are not optically active, $[a_r]$ and $[b_r]$ are the index and absorption tensors that characterize the index and absorption ellipsoids. In addition, $[\gamma_{or}]$ is the optical activity tensor, as in Section 2.2.2.3, and $[\beta_{or}]$ is called the tensor of circular dichroism. For a given wave propagation vector, these latter tensors define the complex optical activity vector $\vec{\Gamma}_o$ according to:

$$\vec{\Gamma}_o = \vec{\Gamma}_{or} + i\vec{B}_{or} = ([\gamma_{or}] + i[\beta_{or}])\vec{q}, \tag{2.27a}$$

where the real part is the (real) optical activity vector, $\vec{\Gamma}_{or} = [\gamma_{or}]\vec{q}$ [as in Eq. (2.25a)], and the imaginary part is the circular dichroism vector, $\vec{B}_{or} = [\beta_{or}]\vec{q}$.[40]

Linear birefringence and dichroism can be neglected for absorbing, optically active crystals of orthorhombic and higher symmetries in which

wave propagation occurs along the optic axes. Thus, only circular birefringence and dichroism need to be considered in this case, and the polarization modes continue to be left- and right-circular. As an example, when propagation occurs along the optic axis (or z-axis) in a uniaxial crystal, $a_{11} = a_{22} \equiv a$ (with all other elements zero), and one finds complex indices of refraction for the orthogonal circular polarizations given by $(1/N^2) = a \pm \Gamma_{oz}$. In the case of extinction coefficients that are small compared with the index of refraction, one arrives at the complex expression:

$$\rho_o = \frac{\pi(N_L - N_R)}{\lambda} = \frac{\pi n^3 \Gamma_{oz}}{\lambda}, \tag{2.27b}$$

analogous to the real expression of Eq. (2.26). The decomposition of this equation into real and imaginary parts demonstrates clearly the effects of circular birefringence and dichroism, respectively, and their direct relationship to the tensors $[\gamma_{or}]$ and $[\beta_{or}]$.

2.2.3 Dispersion Relationships

The real and imaginary parts of the optical properties of solids, combined to form any of the complex quantities such as χ_e, N, ε, or σ, are not independent, but are interrelated through dispersion relationships. These relationships can be derived considering the fact that the complex electric susceptibility χ_e, as one example, is a response function that describes the polarization existing at position $\vec{r}$ and time t in terms of the electric field acting at all other positions and times.[20] In solids that are not optically active, spatial dispersion can be neglected for the purposes of this discussion because the wavelength is long compared to the atomic scale of the optical processes. As a result, for a given time t one can write:

$$\vec{P}(t) = \varepsilon_0 \int_{-\infty}^{\infty} \chi_e(t - t')\vec{E}(t')dt', \tag{2.28}$$

which holds for optically isotropic solids. Analogous expressions involving the susceptibility tensor components hold for anisotropic solids that are not optically active. (Optically active solids, in which spatial dispersion is important, have been described briefly in Section 2.2.2.)

Causality requires that there can be no response $\vec{P}(t)$ to the field $\vec{E}(t')$ before it is applied (i.e., for $t < t'$); as a result, $\chi_e(t - t') = 0$ in Eq. (2.28) for $t < t'$. When Eq. (2.28) is Fourier transformed to complex angular frequency space, yielding $\vec{P}(\omega) = \varepsilon_0 \chi_e(\omega)\vec{E}(\omega)$, then the condition of causality

leads to the dispersion relationships. Reverting now to the complex dielectric function, these relationships can be expressed by:

$$\varepsilon_1(\omega) - 1 = \frac{2}{\pi} P \int_0^\infty \frac{\omega' \varepsilon_2(\omega')}{\omega'^2 - \omega^2} d\omega', \tag{2.29a}$$

$$\varepsilon_2(\omega) = -\frac{2\omega}{\pi} P \int_0^\infty \frac{\varepsilon_1(\omega') - 1}{\omega'^2 - \omega^2} d\omega', \tag{2.29b}$$

where ω and ω' denote real angular frequencies and P denotes the principal value of the integral. In deriving these relationships, it has been assumed that one is interested in response functions that yield real outputs. Equations (2.29), called the Kramers-Kronig relations, have important applications for modeling the optical properties of solids, as described in Section 2.5.

2.3 Classical Theories of the Optical Properties of Solids

Next, simple classical theories will be introduced that predict the photon energy dependence of the complex dielectric function $\varepsilon(\omega)$ of metals, semiconductors, and insulators from the standpoint of classical mechanics and electrodynamics. In addition, such theories provide insights into the optical manifestations of the fundamental microscopic behavior of electrons in these solids. The classical theories originate from the work of Lorentz and Drude and serve as useful starting points from which one can model the observed ellipsometric spectra of solids. As will be seen in Section 2.4, the classical models have quantum mechanical analogs that in fact lead to the same mathematical expressions for the dielectric functions, but with a reinterpretation of the parameters. The Lorentz oscillator model is applicable to semiconductors and insulators; the Drude model is applicable to free-electron metals. The quantum mechanical analogs of these two models describe the direct interband transitions of Section 2.4.2 and the intraband transitions of Section 2.4.5, respectively.

2.3.1 Semiconductors and Insulators: the Lorentz Oscillator Model

In deriving the optical properties of an optically isotropic (nonmagnetic) semiconductor or insulator, the Lorentz model starts with the equation

of motion for a spherical charge distribution, having a single electron of charge $-e$ and mass m, bound to a nucleus of charge $+e$ fixed in position in the solid. This system constitutes one "atom". According to Newton's Second Law:

$$m\frac{d^2\vec{r}}{dt^2} = -e\langle\vec{E}_{loc}\rangle - m\omega_0^2\vec{r} - m\gamma\frac{d\vec{r}}{dt}, \tag{2.30}$$

where $\vec{r}$ is the displacement of the center of charge of the electron cloud from the nucleus located at $\vec{r} = 0$. [20] The first term on the right is a driving force generated by the microscopic electric field, spatially averaged over the electron cloud: $\langle\vec{E}_{loc}\rangle = \langle\vec{E}_{0,loc}\rangle\exp(-i\omega t)$. Thus, $\langle\vec{E}_{0,loc}\rangle$ is the spatial average of the field at $t = 0$. The second term is a Hooke's law restoring force with a force constant of $m{\omega_0}^2$, where ω_0 is the resonance frequency for the atomic oscillator. The third term represents viscous damping of the electron cloud's motion with a frictional constant γ. This provides an energy dissipation pathway that occurs in solids through interactions with neighboring atoms. In this classical model, two assumptions are made. First, the mass of the nucleus is assumed to be infinite since it forms a rigid lattice with neighboring nuclei. Second, the force $\vec{F}_M$ due to the magnetic induction $\vec{B}_{loc}$, $\vec{F}_M = -e\langle\vec{v}\times\vec{B}_{loc}\rangle$ is assumed to be negligible since the electron speed v is small compared with the speed of light.

Upon solving Eq. (2.30), one arrives at a complex charge displacement $\vec{r}$, and assuming N_a identical atoms per unit volume, one obtains a complex macroscopic polarization $\vec{P}$:

$$\vec{P} = N_a\vec{p} = -N_a e\vec{r}(t) = N_a\alpha_a(\omega)\langle\vec{E}_{loc}\rangle \tag{2.31a}$$

$$= \frac{N_a e^2\langle\vec{E}_{loc}\rangle}{m[(\omega_0^2 - \omega^2) - i\gamma\omega]}. \tag{2.31b}$$

Here, $\vec{p}$ is the complex induced dipole moment for a single atom, and α_a is the complex microscopic atomic polarizability. Through this model, the frequency dependence of the optical properties can be understood intuitively. The physical mechanism underlying the dissipation via the conduction electrons in Eqs. (2.13a), (2.14), and (2.15) and the frictional term in Eq. (2.31b) must be the same, since they both lead to a complex $\chi_e(\omega)$. Considering $\omega < \omega_0$ for the moment, if $\gamma \ll [({\omega_0}^2 - \omega^2)/\omega]$, then the interaction with neighboring atoms is very weak and the electrons remain "bound". However, if γ is of the order of $[({\omega_0}^2 - \omega^2)/\omega]$, the electrons in the Lorentz model need to be considered as "conduction electrons" that interact strongly with the neighboring atoms. This viewpoint identifies a

problem in the separation of Eqs. (2.3b) and (2.5d): the same electrons that are bound at one frequency become conduction electrons at another frequency. For this reason, the use of a single complex quantity, N, ε, or σ, through the application of Eqs. (2.14) or (2.17) is preferred to account for this behavior when describing the frequency-dependent optical properties.

It should be pointed out that, in general, the macroscopic electric field in the solid is not equivalent to the spatial average of the local electric field over the electron cloud, i.e., $\langle \vec{E}_{loc} \rangle \neq \vec{E}$. As a result, $N_a \alpha_a(\omega) \neq \varepsilon_0 \chi_e(\omega)$, in general. The reason is that the averaging to determine $\langle \vec{E}_{loc} \rangle$ does not include the contribution from the atom at the site of interest (which is "removed" to perform the averaging), nor does it include regions between sites. For conduction electrons, the equality $\langle \vec{E}_{loc} \rangle = \vec{E}$ is expected to be reasonably correct. (In fact, in this case the term "local field" is a misnomer.) At the opposite extreme where the electrons are strongly localized at the atomic sites so as to be considered point dipoles, then Lorentz calculated a field of $\vec{E}_{loc}(site) = \langle \vec{E}_{loc} \rangle = \vec{E} + (\vec{P}/3\varepsilon_0)$. It is anticipated that the correct local field lies between these two extremes, closer to $\vec{E}$ than to $\vec{E} + (\vec{P}/3\varepsilon_0)$.[20]

The choice $\langle \vec{E}_{loc} \rangle = \vec{E}$ [so that $N_a \alpha_a(\omega) = \varepsilon_0 \chi_e(\omega)$] in Eqs. (2.31) gives rise to the complex frequency-dependent dielectric function given by:

$$\varepsilon(\omega) = \varepsilon_{0s} + \frac{e^2}{\varepsilon_0 m} \sum_n \frac{N_{eb,n}}{(\omega_n^2 - \omega^2) - i\gamma_n \omega}, \tag{2.32a}$$

with the real and imaginary parts given by:

$$\varepsilon_1(\omega) = \varepsilon_{0s} + \frac{e^2}{\varepsilon_0 m} \sum_n \frac{N_{eb,n}(\omega_n^2 - \omega^2)}{(\omega_n^2 - \omega^2)^2 + \gamma_n^2 \omega^2}, \tag{2.32b}$$

$$\varepsilon_2(\omega) = \frac{e^2}{\varepsilon_0 m} \sum_n \frac{N_{eb,n} \gamma_n \omega}{(\omega_n^2 - \omega^2)^2 + \gamma_n^2 \omega^2}. \tag{2.32c}$$

These equations have been generalized beyond Eqs. (2.31) in two respects. First, Eqs. (2.32) include the possibility of more than one electron per atom. $N_{eb,n}$ is the concentration of electrons ('*eb*' denoting electrons bound to their atomic cores) having resonance frequency ω_n and frictional constant γ_n. Thus,

$$N_{eb} = ZN_a = \sum_n N_{eb,n}$$

is the total bound electron concentration and Z is the number of electrons per atom. Second, the value of ε_{0s} at the right in Eqs. (2.32a) and (2.32b) represents the contributions to ε_1 from electronic resonances with ω_n much greater than the measurable frequency range. In the absence of such contributions $\varepsilon_{0s} = 1$ as would be predicted from Eqs. (2.31) upon applying the relation $\varepsilon(\omega) = 1 + \chi_e(\omega)$.

The alternative choice for the average local field, $\langle \vec{E}_{loc} \rangle = \vec{E} + (\vec{P}/3\varepsilon_0)$, leads to a dielectric function given by[61]

$$\frac{\varepsilon(\omega) - 1}{\varepsilon(\omega) + 2} = \varepsilon_{0l} + \frac{e^2}{3\varepsilon_0 m} \sum_n \frac{N_{eb,n}}{(\omega_n^2 - \omega^2) - i\gamma_n\omega}. \tag{2.33}$$

In this case, $\varepsilon_{0l} = 0$ in the absence of resonances at frequencies much greater than the measurable frequency range. The simpler expressions of Eqs. (2.32), however, have been used most widely to fit the measured optical properties of insulators and semiconductors.

The expressions of Eqs. (2.32) can also be applied in modeling optical properties of materials in which the dissipation or optical absorption is negligible, i.e., when

$$\gamma_n << \frac{(\omega_n^2 - \omega^2)}{\omega},$$

for all n. This can occur under two circumstances: (i) when all electron resonances lie well above the accessible optical frequency range, as in the measurement of insulating (or very wide band gap semiconducting materials), and (ii) when low optical frequencies are used, corresponding to photon energies below the band gap, in the measurement of any nonmetallic material ($\omega_n > 0$ for all n). In these circumstances, setting $\gamma_n = 0$ in Eq. (2.32a), for example, yields:

$$\varepsilon_1(\omega) = \varepsilon_{0s} + \frac{e^2}{\varepsilon_0 m} \sum_n \frac{N_{eb,n}}{(\omega_n^2 - \omega^2)}, \tag{2.34}$$

with $\varepsilon_2(\omega) = 0$. This relationship, called the Sellmeier expression, is used widely in the analysis of ellipsometric spectra in order to model the optical properties of non-absorbing materials.

Finally, it should be noted that the Sellmeier expression of Eq. (2.34) and the Lorentz oscillator expression of Eqs. (2.32) can also be extended to anisotropic solids in order to model the frequency-dependence of linear birefringence and dichroism. For example, if the direction of linear

polarization (defined by the direction of $\vec{D}$) is parallel to the x, y, or z principal crystallographic axes of an orthorhombic crystal that is not optically active, then the approach outlined above applies and leads to one of the three principal complex dielectric functions. The resonance frequencies, frictional constants, and resonance electron concentrations associated with the three different polarization directions will be different, in general. The mechanical analog in this case is in terms of directionally dependent (or anisotropic) force constants and frictional constants for the electron oscillator. This may arise due to different strengths of the interactions with neighboring atoms in the different directions. In fact, a mechanical analog of an optically active crystal can also be developed. In this case the spatial dispersion is simulated by describing the interaction of the electromagnetic wave with two coupled harmonic oscillators that lie in separate planes of the crystal.[61]

The Lorentz oscillator describes other physical situations as well. For example, the optical properties of thin films consisting of nanometer-sized metallic particles, as well as composite bulk materials consisting of metal particles embedded within an insulator matrix exhibit Lorentz oscillator behavior as discussed in Section 2.3.2. Furthermore, the vibration of ions in ionic crystals and the orientation of permanent dipoles generate optical property characteristics similar to those of Eqs. (2.32) in the infrared and microwave ranges of the electromagnetic spectrum.

2.3.2 Metals: The Drude Free Electron Model

The Drude model for optically isotropic, nonmagnetic metals can be derived directly from the Lorentz oscillator model. This is accomplished by setting the electron restoring forces to zero, i.e., $\omega_n = 0$ for all n, meaning that all electrons are free to move throughout the collection of atoms.[20] In addition, because the free electrons are uniformly distributed, then $\langle \vec{E}_{loc} \rangle \approx \vec{E}$. By setting $\omega_n = 0$ in Eqs. (2.32) and then assuming all electrons are subject to the same frictional force constant $\gamma_f = \gamma_n$ one obtains:

$$\varepsilon_1(\omega) = 1 - \frac{\omega_p^2}{(\omega^2 + \gamma_f^2)}, \tag{2.35a}$$

$$\varepsilon_2(\omega) = \frac{\omega_p^2 \gamma_f}{\omega(\omega^2 + \gamma_f^2)}, \tag{2.35b}$$

where the plasma frequency ω_p now describes the free electron concentration and is given by $\omega_p = (N_{ef} e^2/\varepsilon_0 m)^{1/2}$. The constant term ε_{0s} in

Eqs. (2.32a) and (2.32b) is now unity since all electrons in the solid are assumed to be free, and thus are included in the model. With the approach of Eqs. (2.35), however, the physical origin of the damping term in Eq. (2.30) is unclear.

In fact, in the Drude theory, these results are derived from a somewhat different perspective, by considering the metal as a free electron gas, electrically compensated by a background of positively charged nuclei.[44,45] In this theory, one assumes that the electrons move according to Newton's Laws between instantaneous scattering events that occur with a probability per unit time of $(1/\tau)$ and randomize the velocity direction. The scattering events were attributed by Drude to the same phenomenon giving rise to Joule heating when current passes through the metal. Although Drude ascribed the scattering to charged particle collisions (i.e., electrons with the positively charged nuclei), they are in fact due to deviations from perfect periodicity in the solid, e.g., lattice vibrations at room temperature. With such a model, the equation of motion for the average velocity $\vec{v}$ of an electron is

$$\frac{d\vec{v}}{dt} = -\frac{\vec{v}}{\tau} - \frac{e\vec{E}}{m}, \tag{2.36}$$

assuming no distinction between local and macroscopic electric fields. However, for this equation to be valid, the macroscopic field $\vec{E}$ must not vary significantly over the mean free path L of the electron, i.e., $L = v\tau \ll \lambda$ and $L \ll 2/\alpha$ where λ is the wavelength and $2/\alpha$ is the exponential decay length of the field. Assuming $\vec{E} = \vec{E}_0 \exp(-i\omega t)$, then a steady state solution of the form $\vec{v} = \vec{v}_0 \exp(-i\omega t)$ yields the complex conduction current $\vec{J}_{cond} = -N_{ef} e\vec{v} = 4\pi\varepsilon_0\sigma\vec{E}$, where the complex optical conductivity σ is given by:

$$\sigma = \sigma_1 + i\sigma_2 = \frac{(N_{ef}e^2\tau/4\pi\varepsilon_0 m)}{1 - i\omega\tau} = \frac{(\sigma_0/4\pi\varepsilon_0)}{1 - i\omega\tau}. \tag{2.37a}$$

The real and imaginary parts of the complex optical conductivity are:

$$\sigma_1 = \frac{(\sigma_0/4\pi\varepsilon_0)}{1 + \omega^2\tau^2}, \tag{2.37b}$$

$$\sigma_2 = \frac{(\sigma_0\omega\tau/4\pi\varepsilon_0)}{1 + \omega^2\tau^2}, \tag{2.37c}$$

where $\sigma_0 = 4\pi\varepsilon_0\sigma(\omega = 0) = (N_{ef}e^2\tau/m)$ denotes the dc electrical conductivity in SI units $[(\Omega\ \mathrm{m})^{-1}]$. Equations (2.37) can be seen to be equivalent

to Eqs. (2.35) with $\gamma_f = 1/\tau$, using the relationship between the optical conductivity and dielectric function in Eqs. (2.16).

As will be discussed in greater detail in section 2.5, most metals exhibit the optical characteristics of bound and free electrons within the same frequency range. Furthermore, very heavily doped (i.e., degenerate) semiconductors may also exhibit a combination of the bound and free electron behaviors. In these cases, one can separate the two types of optical characteristics using the following expressions

$$\varepsilon_1(\omega) = \varepsilon_{0s} - \frac{N_{ef}e^2}{\varepsilon_0 m(\omega^2 + \gamma_f^2)} + \frac{e^2}{\varepsilon_0 m}\sum_n \frac{N_{eb,n}(\omega_n^2 - \omega^2)}{(\omega_n^2 - \omega^2)^2 + \gamma_n^2\omega^2}, \quad (2.38a)$$

$$\varepsilon_2(\omega) = \frac{N_{ef}e^2\gamma_f}{\varepsilon_0 m\omega(\omega^2 + \gamma_f^2)} + \frac{e^2}{\varepsilon_0 m}\sum_n \frac{N_{eb,n}\gamma_n\omega}{(\omega_n^2 - \omega^2)^2 + \gamma_n^2\omega^2}. \quad (2.38b)$$

In fact, these represent the most general form of the dispersion equations presented so far, as they can be applied to metallic, semiconducting, and insulating materials.

The optical properties of a thin film consisting of an array of isolated nanometer-sized metallic particles (i.e., particles small compared to the wavelength) are of interest because the electrons that would normally be free in the bulk metal now exhibit a non-zero resonance frequency in the particulate metal. The resonance in this case is understood in terms of the collective oscillation of free electrons within each metal particle. These oscillations are driven by the optical field, and the restoring force is provided by the rigid array of positive nuclei. Thus, true oscillations in charge density that build up and collapse periodically at the particle surfaces can in fact be excited by transverse electromagnetic waves in particulate metals. This situation is in contrast to that in the conventional optical measurement of bulk metals (see Section 2.3.3).

For a single spherical particle, the equation of motion for the displacement $\vec{r}$ of the free electron distribution relative to the fixed, positively charged distribution of nuclei can be written as:

$$m\frac{d^2\vec{r}}{dt^2} = -e\left(\vec{E} - \frac{\vec{P}}{3\varepsilon_0}\right) - \frac{m}{\tau}\frac{d\vec{r}}{dt} = -e\vec{E} - \frac{m}{\tau}\frac{d\vec{r}}{dt} - \frac{N_{ef}e^2\vec{r}}{3\varepsilon_0}. \quad (2.39)$$

Here, m and $-e$ are the electron mass and charge, N_{ef} is the concentration of free electrons within the particle, and $\vec{E} - (\vec{P}/3\varepsilon_0)$ is the field within

the particle, the latter including the applied optical field and the (complex) depolarization field. Because this equation has the same form as Eq. (2.30), solving for the dielectric function yields a Lorentz oscillator expression with a resonance frequency of

$$\omega_{pp} = \sqrt{\frac{N_{ef}e^2}{3\varepsilon_0 m}} = \frac{\omega_p}{\sqrt{3}},$$

which can be interpreted as the plasma frequency for the particle. Such free electron, or plasma oscillations will be discussed further in Section 2.3.3.

2.3.3 Plasmons

To supplement this section, it is of interest to continue the discussion on the characteristics of free electron oscillations in metallic solids in somewhat greater detail. Generally, the collective oscillations of free electrons are called plasmons, and these can be excited in bulk metals, at metal surfaces, and in particulate metal films (as described in Section 2.3.2). Only the surface and particle plasmons can be excited by light waves in an ellipsometry experiment. Surface plasmon excitation can be observed either by illumination of rough surfaces or in specialized sample geometries, whereas the particle plasmon excitation is observed directly through the dielectric function of the sample, without specialized measurement conditions. A complete understanding of such oscillations requires a detailed quantum mechanical treatment of the free electrons in the metal. Here, a brief overview of the phenomena of volume and surface plasmons will be provided, but an in-depth quantum mechanical approach will be avoided.

Because the volume and surface plasmons are self-sustaining and are free to propagate throughout the bulk and surface of the solid, respectively, one can associate propagation vectors with these excitations. These vectors are denoted here as $\vec{q}_p$ for the volume plasmon and $\vec{q}_s$ for the surface plasmon, where $\lambda_p = 2\pi/q_p$ and $\lambda_s = 2\pi/q_s$ are the wavelengths of the associated oscillations in the charge density fluctuations. The expressions connecting the plasmon frequency (or plasmon energy, denoted $\hbar\omega_{qp}$ and $\hbar\omega_{qs}$) and the magnitude of the propagation vector are called dispersion relationships, and these provide insights into the ability of light waves to excite the plasmons. (In the following paragraphs, when 'q' is inserted into the subscript for ω_p or ω_s, the intrinsic propagation vector dependence of the plasmon frequency is being emphasized. Without the

subscript, ω_p and ω_s refer to the bulk and surface plasma frequencies in the limit of short q_p and long q_s, respectively; see Figs. 2.3.)

Volume plasmons are collective oscillations of the electrons in a bulk solid, i.e., oscillations of charge density fluctuations that are self-sustaining, longitudinal, and quantized with energy $\hbar\omega_p$.[20,44,45,62] Because volume plasmons can be present in the absence of an external field, the oscillations can occur when the real and imaginary parts of the dielectric function in the longitudinal direction (i.e., the plasmon propagation direction) satisfy $\varepsilon_1 = 0$ and $\varepsilon_2 \ll 1$. This occurs in Eqs. (2.35) when $\omega = \omega_p$ and $\gamma_f \ll \omega_p$. Because volume plasmons are longitudinal, they cannot be excited by light waves in an ellipsometry experiment, (not even near ω_p). The simplest way to excite volume plasmons is to generate fluctuations in charge density, e.g., by passing monoenergetic electrons through the sample. Peaks in the electron energy loss spectrum of the transmitted electrons occur at energies corresponding to integral multiples of $\hbar\omega_p$. In fact, plasmons generated by electrons exhibit a range of propagation vector magnitudes from $q_p = 0$ to $q_p = q_{pc}$, a characteristic cutoff value, which for a free electron gas is given by $q_{pc} \approx \omega_p/v_F$, where v_F is the maximum velocity of electrons in the distribution (i.e., the Fermi velocity). With typical values of ω_p and v_F for metals, q_p approaches $q_{pc} \sim 1\ \text{nm}^{-1}$ from below, implying that near the cutoff, the wavelength λ_p of the plasmon is approaching atomic dimensions from above. As a result, single electron excitations occur for q_p above q_{pc}, and these screen out the long range Coulomb interactions required to sustain the plasmon. Weak dispersion (or propagation vector dependence) in the plasmon frequency denoted by ω_{qp} occurs due to the thermal motions of electrons. In optically isotropic materials in which the longitudinal and transverse dielectric functions are equal, the plasmon energy ω_p identified from the optical properties, e.g., in an ellipsometric measurement, identifies ω_{qp} in the limit $q_p \to 0$ (even though plasmons are not excited in this measurement). The dispersion characteristics of volume plasmons and single electron excitations in a free electron gas are summarized in Fig. 2.3(a).

Surface plasmons are collective oscillations of electrons in the plane of a solid surface, that are self-sustaining, quantized with energy $\hbar\omega_s$, and can be localized within an atomic distance of the surface (within ~0.1 nm).[63–65] Assuming self-sustaining fields at the surface, Maxwell's Equations yield a dispersion relationship of

$$q_{sx} = \frac{\omega_{qs}}{c}\sqrt{\frac{\varepsilon_a\varepsilon}{\varepsilon_a + \varepsilon}}.$$

Here, q_{sx} and ω_{qs} are the wavevector magnitude and frequency of the surface plasmon, assumed to be propagating along the x-direction at the

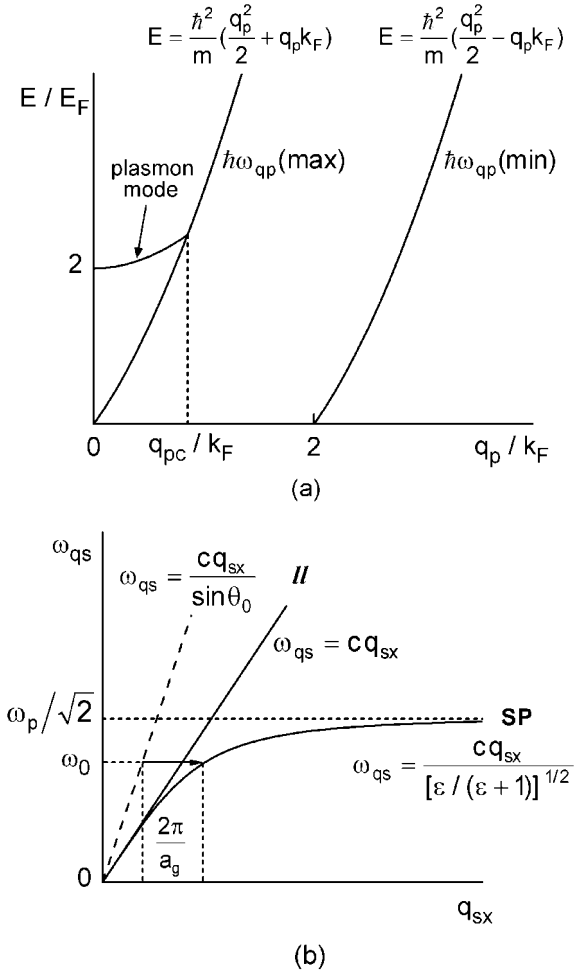


Figure 2.3 (a) Dispersion relationships associated with volume excitations of a free-electron gas. The electron energy E and wavevector q_p are normalized to the Fermi energy E_F and Fermi wavevector k_F, respectively. The region between the lines designated by $\hbar\omega_{qp}$(max) and $\hbar\omega_{qp}$(min) represents the possible one-electron excitations. Below a critical electron wavevector designated q_{pc}, a collective electron excitation or volume plasmon mode is observed. In this plot, the plasma energy $\hbar\omega_p$ (which is the electron energy in the limit $q_p \to 0$) is chosen to be twice the Fermi energy (after Ref. 20). (b) Schematic dispersion relationship for surface plasmons (SP) propagating in the x-direction in a free-electron gas. The photon line ($//$) is also shown. The horizontal arrow denotes the excitation of surface plasmons by light of frequency ω_0 using a grating coupler [see Eq. (2.40a)]. Here ε is the dielectric function of the grating (real; $\varepsilon < -1$), θ_0 is the angle of incidence of the light on the grating plane, and a_g is the grating constant.

surface of a medium. In addition, $\varepsilon = \varepsilon_1 + i\varepsilon_2$ is the dielectric function of the metallic medium, and ε_a is the dielectric function of the ambient (assumed to be real). For a propagating wave at the metal surface with real q_{sx}, $\varepsilon_1 < -\varepsilon_a$ and $\varepsilon_2 \ll 1$ which can occur in a metal for

$$\omega < \frac{\omega_p}{\sqrt{1 + \varepsilon_a}},$$

as long as $\gamma_f \ll \omega$ [see Eq. (2.35)]. For small q_{sx}, which is the situation when $\varepsilon_1 \ll -\varepsilon_a$,

$$\omega_{qs} = \frac{q_{sx}c}{\sqrt{\varepsilon_a}},$$

and the dispersion relation approaches that of the light wave. At large q_{sx}, which is the situation when $\varepsilon_1 \to -\varepsilon_a$, the frequency asymptotes according to

$$\omega_{qs} \to \frac{\omega_p}{\sqrt{1 + \varepsilon_a}},$$

for a free electron metal of volume plasma frequency ω_p.

Figure 2.3(b) shows a schematic of the resulting dispersion relationship for surface plasmons, along with the straight line associated with light in vacuum having the same wavevector and frequency $\omega = \omega_{qs} = cq_{sx}$. From this plot, one can conclude that under usual circumstances light cannot simultaneously satisfy the momentum and energy conservation requirements to excite surface plasmons. However, one can design various geometrical configurations in which photons can be coupled to surface plasmons. Here two approaches are described. In both approaches, for a given optical wave frequency, the propagation vector component of the light wave $q_x = \omega_0/c$ must be increased by a value of Δq_x such that $q_x + \Delta q_x = q_{sx}$.

In a grating coupler, light with frequency ω_0 strikes a metallic grating with grating constant, a_g, and dielectric function ε_1 ($\varepsilon_2 \ll 1$) at an angle of incidence θ_0. With the grating in air ($\varepsilon_a = 1$), coupling can be achieved under the following conditions:

$$\frac{\omega_0 \sin\theta_0}{c} \pm \frac{2\pi\nu}{a_g} = \frac{\omega_0}{c}\sqrt{\frac{\varepsilon_1}{\varepsilon_1 + 1}} = q_{sx}, \tag{2.40a}$$

where ν is an integer. The coupling is shown in Fig. 2.3(b) by the horizontal arrow (assuming $\nu = 1$). From analogous reasoning, it is clear that surface plasmons can also be excited on rough surfaces. In fact, weak bands due to surface plasmon absorption are sometimes detected in the reflectance or ellipsometric spectra of rough metal films.[20]

In an attenuated total reflection (ATR) coupler, an insulating medium (e.g., fused silica) with $\varepsilon = \varepsilon_i > 1$ (real) is formed into the shape of a semi-cylinder with a metal film deposited on the flat surface. A light wave is directed radially inward, normal to the cylinder axis so as to reflect internally from the metal surface. Under these conditions, the resonance condition is met at the outer surface of the metal film (in air), and is given by

$$\frac{\omega_0 \sin\theta_0}{c} \sqrt{\varepsilon_i} = \frac{\omega_0}{c} \sqrt{\frac{\varepsilon_1}{\varepsilon_1 + 1}} = q_{sx}, \tag{2.40b}$$

where θ_0 is the angle of incidence of the wave at the insulator/metal interface measured from the normal to the interface. In this situation, the evanescent wave in the metal is used to excite the surface plasmon. To ensure proper excitation, the thickness of the metal film must be optimized; typical values are on the order of 50 nm. Reflectance and ellipsometric measurements can be performed in this configuration, with the internally reflected beam leaving the semi-cylinder along its radius. The wavelength at which the surface plasmon resonance feature occurs in the reflectance or ellipsometric spectra is extremely sensitive to the presence of thin films on the surface of the metal. Thus, this method can be applied as a probe of thin film adsorption and other phenomena at the metal surface.

2.3.4 Optical Sum Rules

Sum rules are based on the observation that for an electromagnetic wave of sufficiently high frequency such that $\omega \gg \omega_n$, $\omega \gg \gamma_n$, and $\omega \gg \gamma_f$, all electrons in a solid respond to the wave as free electrons with $\varepsilon_1(\omega) = 1 - (\omega_p^2/\omega^2)$ and $\varepsilon_2(\omega) = 0$[66] In the context of any type of solid (insulator, metal, semiconductor) as described by Eqs. (2.38), ω_p is now given by $\omega_p = (N_e e^2/\varepsilon_0 m)^{1/2}$, where N_e is the total concentration of electrons, given from Eqs. (2.38) by $N_e = N_{es} + N_{ef} + \sum_n N_{eb,n}$. Here N_{es} represents the contribution from electrons that generate the ε_{0s} term in Eq. (2.38a); if $\varepsilon_{0s} = 1$, then $N_{es} = 0$. This contribution can be computed as

$$N_{es} = \frac{\varepsilon_0 m \omega_{0s}^2 (\varepsilon_{0s} - 1)}{e^2},$$

where it is assumed that ε_{0s} arises from a collection of electrons having a single resonance frequency ω_{0s}. Applying these considerations to the Kramers-Kronig relations, one can relate the optical properties of an optically isotropic solid to the concentration of all the electrons that interact with the electric field of a (transverse) electromagnetic wave:

$$\int_0^\infty \omega\varepsilon_2(\omega)d\omega = \frac{\pi}{2}\omega_p^2 = \frac{\pi N_e e^2}{2\varepsilon_0 m}. \tag{2.41}$$

It is more practical to employ a finite frequency range, and this leads to:

$$\int_0^{\omega_c} \omega\varepsilon_2(\omega)d\omega = \frac{\pi N_a e^2}{2\varepsilon_0 m} Z_{eff}(\omega_c), \tag{2.42}$$

where m is the free electron mass, N_a is the concentrations of atoms, and $Z_{eff}(\omega_c)$ is the number of electrons per atom contributing to the optical properties up to frequency ω_c. Similar sum rule approaches have been calculated in which $\mathrm{Im}[-1/\varepsilon(\omega)]$ replaces $\varepsilon_2(\omega)$ in Eqs. (2.41) and (2.42). These latter rules are most useful when the electronic excitation occurs by the field of a longitudinal electron wave.[20] For optically isotropic materials in the long wavelength limit, appropriate for plasmons (see Section 2.3.3), as well as for the optical range of electromagnetic waves in Eq. (2.42), the two approaches are interchangeable. Other sum rules, derived from quantum mechanical principles and applicable to the oscillator strengths of transitions in atomic and solid state systems, will be mentioned briefly in Section 2.4.

2.4 Quantum Mechanical Theories of the Optical Properties of Solids

In this section, quantum mechanical descriptions of the dielectric functions of solids will be presented. Although the focus will be on optically isotropic solids, the extensions to anisotropic solids are straightforward; additional comments on anisotropy are provided in the following paragraph. First, in Section 2.4.1, the quantum mechanical analog of the classical Lorentz oscillator expression is derived. In this part, the influence of the optical electric field on the localized electronic wavefunction associated with each atom of the solid is determined using first-order time-dependent perturbation theory. Then, the dielectric function is calculated from the wavefunction characteristics using quantum mechanical expressions for the complex polarization or the complex conduction current. In Section 2.4.2, the expression for the dielectric function derived in Section 2.4.1 is extended to

describe the more useful situation of delocalized one-electron wavefunctions, or Bloch wavefunctions, associated with electrons in a crystalline solid. The dielectric function in this case provides information on the density of electronic states in the crystal as a function of electron energy, as described in Section 2.4.3. In Sections 2.4.2 and 2.4.3, only direct transitions are considered (i.e., transitions that conserve the crystal momentum of the electron, neglecting the momentum of the adsorbed photon which is exceedingly small in comparison). In Section 2.4.4, indirect transitions are described in which the excitation of electrons by photons is accompanied by the emission or absorption of phonons (leading to a considerable change in the electron wavevector). These latter, second-order processes can be observed in indirect semiconductors in the absence of direct transitions, i.e., in the range of photon energy above the indirect gap but below the lowest direct gap. To conclude this treatment in Section 2.4.5, the quantum mechanical analog of the Drude free electron expression is derived. The quantum mechanical process is described in terms of intraband transitions involving electrons in the partially filled bands in metals and degenerate semiconductors. This section provides only a concise review of the quantum mechanical processes. For background information and further details, one can consult texts on quantum mechanical principles,[56,57] the quantum theory of solids,[44,45] and the modern theories of the optical properties of solids.[20,32]

As in the discussion of the Lorentz oscillator model of Section 2.3.1, the treatment of this section can be readily generalized to anisotropic crystals of orthorhombic and higher symmetries that are not optically active. In this case, the results must be expressed in the coordinate system of the principal crystallographic axes that simultaneously diagonalizes both real and imaginary parts of the complex dielectric tensor. In addition, the direction of linear polarization (i.e., the direction of $\vec{D}$) must lie along one of the principal axes. Then, the expression for the principal complex dielectric function is of the same form as that presented below, but the relevant quantum mechanical matrix element and the joint density of states will depend on the direction of $\vec{D}$. As an example of the approach that must be taken, anisotropy is included explicitly in the expression for the intraband dielectric function, valid for monovalent metals and degenerate semiconductors (see Section 2.4.5).

2.4.1 Quantum Theory of Absorption and Dispersion

2.4.1.1 Atomic Systems

In performing the quantum mechanical derivation of the dielectric function of a collection of weakly interacting atoms that comprise an

optically isotropic solid, one can begin by considering a single unperturbed atom with eigenstates ϕ_n given by:

$$H_0\phi_n = E_n\phi_n, \tag{2.43}$$

where H_0 is the time-independent Hamiltonian (or quantum mechanical energy operator), E_n is the energy of the n^{th} eigenstate, and $n = 0$ denotes the ground state.[20] (From this point onward, the letter "E" is used both for the electric field and for the energy of an electron or photon; the intended meaning of "E" should be clear from the context.) Application of the perturbation $H'(t)$ associated with the optical electric field yields the time-dependent Hamiltonian $H(t) = H_0 + H'(t)$, and the resulting wavefunction Ψ for the atom is determined using the time-dependent Schrödinger equation:

$$i\hbar\frac{\partial\Psi}{\partial t} = [H_0 + H'(t)]\Psi. \tag{2.44}$$

The solution can be written as:

$$\Psi(\vec{r}, t) = \sum_n \mathfrak{a}_n(t)\,\phi_n(\vec{r})\exp\left(-\frac{iE_nt}{\hbar}\right), \tag{2.45}$$

where $\mathfrak{a}_n(t)$ is the time-dependent probability amplitude for the atom to be found in eigenstate n having energy E_n, and the summaton is over all eigenstates.

Equation (2.44) needs to be solved when the perturbation, treated classically, consists of an optical electric field described by $\vec{E}_{loc}$ with magnitude $E_{0,loc}$. Although several different forms for this perturbation are possible, here it is assumed to be a real quantity, directed along the x axis, and positive maximum at $t = 0$. In addition, the spatial variation of the optical electric field over the atomic volume is neglected, meaning that only electric dipole transitions are to be considered here.[57] Under these conditions, the perturbation can be written as:

$$H'(t) = e\vec{E}_{loc}\cdot\vec{r} = \frac{eE_{0,loc}x}{2}\exp(\gamma_p t)\{\exp(i\omega t) + \exp(-i\omega t)\}, \tag{2.46}$$

which describes the energy operator of the atomic dipole in the electric field $\vec{E}_{loc}$. In Eq. (2.46), $\vec{r}$ and x are quantum mechanical position operators for the electron. This perturbation is allowed to build up gradually on the time scale of the optical frequency through the positive rate

parameter γ_P. Solving Eqs. (2.44) and (2.45) for this form of the perturbation yields:

$$\mathfrak{a}_n(t) = -\frac{e\langle E_{0,loc}\rangle x_{n0}}{2\hbar}\exp(\gamma_P t)\left\{\frac{\exp[i(\omega_{n0}-\omega)t]}{\omega_{n0}-\omega-i\gamma_P}+\frac{\exp[i(\omega_{n0}+\omega)t]}{\omega_{n0}+\omega-i\gamma_P}\right\}, \tag{2.47}$$

where it is assumed that the atom is in its ground state $n = 0$ for $t \to -\infty$, and that first-order perturbation theory is applicable, namely that $\mathfrak{a}_n(t)H'(t)\phi_n(\vec{r})$ is negligible for $n \neq 0$.[57] The $\exp(\gamma_P t)$ factor controls the rate at which the amplitude associated with state n builds up, and provides a phenomenological approach for introducing a finite lifetime associated with the electron transition. With this approach if γ_P is reduced, then the lifetime increases. In addition, in Eq. (2.47), $\omega_{n0} = \omega_n - \omega_0$, where $\omega_n = E_n/\hbar$, $\omega_0 = E_0/\hbar$, and E_0 and E_n are the ground eigenstate and n^{th} excited eigenstate energies. Finally, x_{n0} is the dipole matrix element given by:

$$x_{n0} = \int_{\Delta V_A} \phi_n^*(\vec{r})x\phi_0(\vec{r})d^3\vec{r}, \tag{2.48}$$

where the integral in this expression is taken over the atomic volume ΔV_A.

If the integral of Eq. (2.48) vanishes, then the transition is dipole forbidden; under such circumstances, it may then be necessary to consider the spatial variation of $\vec{E}_{loc}$ over the atomic volume. Introduction of first order terms in $\exp[\pm i(\vec{q}\cdot\vec{r})]$ [see Eq. (2.10a)] in the perturbation theory leads to the much weaker magnetic dipole and electric quadrupole transitions.[57] Finally, it should be noted that the buildup of the perturbation, described by γ_P in Eqs. (2.46) and (2.47), is intended to simulate the intrinsic limiting (i.e., longest possible) lifetime. Normally "extrinsic" mechanisms control the lifetime, e.g., interatomic interactions, as described for the classical model of Eqs. (2.32). These latter broadening mechanisms are difficult to incorporate from a first-principles basis, however, and generally require a more cumbersome mathematical approach (see Section 2.5.4).

The complex polarization induced by the optical electric field is:

$$\vec{P} = \varepsilon_0(\varepsilon - 1)\vec{E} \tag{2.49a}$$

$$= N_a\vec{p} = -N_a e\int_{\Delta V_A}\Psi^*(\vec{r},t)\vec{x}\Psi(\vec{r},t)d^3\vec{r}, \tag{2.49b}$$

where $N_a = N_e/Z$ is the concentration of atoms.[20] By applying Eqs. (2.45) and (2.47) to Eqs. (2.49), setting $\vec{E} = \langle \vec{E}_{loc} \rangle$, and taking the limit $\gamma_P t \to 0$, one can arrive at the following expression for the complex dielectric function:

$$\varepsilon = 1 + \frac{N_a e^2}{2\varepsilon_0 m} \sum_n \frac{f_{n0}}{\omega_{n0}} \left(\frac{1}{\omega_{n0} + \omega + i\gamma_P} + \frac{1}{\omega_{n0} - \omega - i\gamma_P} \right), \quad (2.50a)$$

where f_{n0} is the oscillator strength defined by:

$$f_{n0} = \frac{2m\omega_{n0}}{\hbar} |x_{n0}|^2. \quad (2.50b)$$

Equation (2.50a) has the same form as Eq. (2.32a) for the Lorentz oscillator as can be seen by neglecting terms in $\gamma_P{}^2$ in the former, and making the following parameter associations:

$$\text{Eq. (2.32a)} \to \text{Eqs. (2.50)}$$

$$N_{eb,n} \to N_a f_{n0}, \quad (2.51a)$$

$$\omega_n \to \omega_{n0}, \quad (2.51b)$$

$$\gamma_n \to 2\gamma_P. \quad (2.51c)$$

Thus, the concentration of bound electrons with resonance frequency ω_n in Eq. (2.32a) is associated with the product of the atomic concentration and the oscillator strength, the resonance frequency ω_n in Eqs. (2.32a) simulates the frequency $\omega_{n0} = (E_n - E_0)/\hbar$ connecting the ground and excited states, and the frictional constant γ_n is inversely proportional to the lifetime of the transition, the latter defined according to Eq. (2.46).

A second description of the dielectric function can be obtained in terms of the momentum matrix element by writing the perturbation to first order in the local vector potential $\vec{A}_{loc}$ (of amplitude $A_{0,loc}$ and optical frequency ω) as follows:

$$H'(t) = e\vec{E}_{loc} \cdot \vec{r} = \frac{-e}{m} \vec{A}_{loc} \cdot \vec{p} = \frac{i\hbar e}{m} \vec{A}_{loc} \cdot \vec{\nabla}. \quad (2.52a)$$

Here $\vec{p} = -i\hbar\vec{\nabla}$ is the momentum operator for the electron (not to be confused with the atomic dipole moment of Section 2.3), and the so-called transverse or Coulomb gauge ($\vec{\nabla} \cdot \vec{A} = 0$) is assumed. Then the derivation

of ε is carried out using the following relationships:

$$\vec{J}_{cond} = N_a \vec{j}_{cond} = i\omega\varepsilon_0(1 - \varepsilon)\vec{E} \tag{2.52b}$$

$$= -\frac{N_a e}{2m} \int_{\Delta V_A} [\Psi^*(\vec{r}, t)\vec{p}\Psi(\vec{r}, t) - \Psi(\vec{r}, t)\vec{p}\Psi^*(\vec{r}, t)]d^3\vec{r}$$

$$-\frac{N_a e^2 \vec{A}}{m} \int_{\Delta V_A} \Psi^*(\vec{r}, t)\Psi(\vec{r}, t)d^3\vec{r}, \tag{2.52c}$$

where $\vec{j}_{cond}$ is the conduction current contribution of a single atom.[32,57] The first and second terms in Eq. (2.52c) are called the paramagnetic and diamagnetic currents, respectively. In this derivation of the linear response for a collection of atoms that interact only weakly, the diamagnetic current can be ignored; this is not the case for solids, in general. The approach described by Eq. (2.52c) is very useful for determining the dielectric function associated with interband and intraband transitions in solids (see Sections 2.4.2 and 2.4.5).

Alternatively, in order to obtain the second description of the dielectric function, commutation relations, e.g., $[p_x, x] = -i\hbar$, can be applied directly to yield:

$$|p_{x,n0}|^2 = (m\omega_{n0})^2|x_{n0}|^2, \tag{2.53a}$$

and an equivalent expression for the oscillator strength of Eq. (2.50b):

$$f_{n0} = \frac{2|p_{x,n0}|^2}{m\hbar\omega_{n0}}, \tag{2.53b}$$

where the polarization direction is again assumed to be along $\vec{x}$.[20] In Eqs. (2.53a) and (2.53b), $p_{x,n0}$ is the momentum matrix element given by the x-component of the vector:

$$\vec{p}_{n0} = -i\hbar \int_{\Delta V_A} \phi_n^*(\vec{r})\vec{\nabla}\phi_0(\vec{r})d^3\vec{r}. \tag{2.53c}$$

For atoms with Z electrons, the general f-sum rule is obeyed, given by

$$\sum_n f_{nm} = Z,$$

where f_{nm} represents the oscillator strength for electronic transitions from state m to state n.

Another very useful quantum mechanical approach for describing the optical absorption is derived from an expression for the absorption coefficient associated with electronic transitions from state m to n, given by:

$$\alpha_{nm} = \frac{\hbar\omega_{nm} W_{nm}}{I} = \frac{\hbar\omega_{nm} W_{nm}}{(n\varepsilon_0 c E_0^2/2)}. \tag{2.54}$$

Here $\hbar\omega_{nm} = E_n - E_m$ is the energy difference for the transition from state m to n, and W_{nm} is the transition rate per excitation volume Ω, given[20] by

$$W_{nm} = \frac{1}{\Omega}\frac{d|\mathfrak{a}_{nm}|^2}{dt}.$$

In addition, n and E_0 in the denominator at the right of Eq. (2.54) are the real index of refraction and the amplitude of the optical electric field, respectively. Thus, the numerator of Eq. (2.54) is the rate of energy absorption per volume, and the denominator is the irradiance in the beam denoted I. The quantity $\mathfrak{a}_{nm}$ in the transition rate represents the probability amplitude for finding an electron in state ϕ_n, after having made a transition from state ϕ_m; thus, $|\mathfrak{a}_{nm}|^2$ is the corresponding probability. In fact, $\mathfrak{a}_{nm}$ is a generalization of $\mathfrak{a}_{n0} \equiv \mathfrak{a}_n$ in Eqs. (2.45) and (2.47), the latter being the probability amplitude for finding an electron in state ϕ_n, after having made a transition from the ground state ϕ_0.

2.4.1.2 Solid State Systems

In order to extend the above approaches beyond a collection of weakly interacting atoms to a real solid, the one-electron model with the random phase approximation is usually employed. In this approximation, the influence of the external optical electric field or vector potential on the many-electron wavefunction is obtained by determining the response of a single electron in a self-consistent scalar potential. Local field corrections are needed when the self-consistent potential depends on the external field; such corrections are neglected here. The fact that the one-electron energy states in solids are spread out over bands implies that there will always be a continuum of initial and final states. Thus, such a continuum in turn implies that the valence electrons participating in the optical transitions are usually delocalized throughout the crystal (an exception being the electrons in amorphous solids at energies within the mobility gap). As a result, one must reformulate the model used to determine the fundamental dielectric response of Eqs. (2.50) in which it was assumed that each of the $N_e = ZN_a$ electrons may be excited from a single well-defined ground state to any one of a number of higher energy excited states.

In Section 2.4.2, covalent crystalline solids will be considered in which the one-electron wavefunctions are delocalized Bloch functions specified by the electron wavevector $\vec{k}$ and defined such that the normalization or probability integral is unity when taken over the entire crystal. The Bloch functions can be written as:

$$\phi_{\upsilon}(\vec{r}, \vec{k}_{\upsilon}) = \frac{u_{\vec{k}\upsilon}(\vec{r})\exp(i\vec{k}_{\upsilon} \cdot \vec{r})}{\sqrt{\Omega}}, \tag{2.55}$$

where the index υ distinguishes states in different bands and Ω is the volume of the crystal.[44,45] Here, $u_{\vec{k}\upsilon}$ is the part of the Bloch functions having the periodicity of the lattice, i.e., $u_{\vec{k}\upsilon}(\vec{r}) = u_{\vec{k}\upsilon}(\vec{r} + \vec{R}_l)$, where $\vec{R}_l$ is a lattice vector. In this case, one must distinguish between *direct* interband transitions between Bloch wavefunctions in which $\vec{k}$ is virtually unchanged in the transitions and *indirect* interband transitions in which $\vec{k}$ changes significantly due to an interaction of the electron wave with vibrations of the crystalline lattice. Ionic crystalline solids exhibit behavior that could be considered intermediate between the covalent solids with delocalized electronic wavefunctions and weakly interacting atoms with atomic wavefunctions. If the one-electron wavefunctions are localized as in a highly ionic solid, however, the wavefunctions must be considered as a sum of Bloch functions with a wide range of $\vec{k}$ values. As a result, the transitions in these materials have been called non-direct, i.e., there is no apparent dependence on $\vec{k}$ in the optical absorption process.[20]

Similar such non-direct behavior occurs in amorphous semiconductors, in which the lack of atomic periodicity in the network implies that the electron crystal momentum $\hbar\vec{k}$ is no longer a valid quantum number for describing the one-electron wavefunction. This case will be considered in detail first in order to demonstrate how the continuum of states enters into the calculation of the optical properties, but without the complexity of delocalized Bloch functions.[67] For an amorphous semiconductor, one can write an expression for the absorption coefficient as a function of the energy $E = \hbar\omega$ of each of the photons associated with the quasi-monochromatic electromagnetic wave:

$$\alpha(\hbar\omega) = 2\Omega^2 \int \rho_V(E_m)[\int \alpha(E_n, E_m, \hbar\omega)\rho_c(E_n)dE_n]dE_m, \tag{2.56}$$

where $\alpha(E_n, E_m, \hbar\omega) = \alpha_{nm}$ is defined through Eq. (2.54). In Eq. (2.56), $E_n = \hbar\omega_n$ and $E_m = \hbar\omega_m$ are the energies associated with the excited and initial electronic states, respectively, so that $E_n - E_m = \hbar\omega_{nm}$. In addition, $\rho_v(E_m)dE_m$ and $\rho_c(E_n)dE_n$ are the number of single-spin valence and conduction band states per unit volume, respectively, having energies between E_m and $E_m + dE_m$ and between E_n and $E_n + dE_n$. The factor of two

accounts for the enhancement in absorption due to the fact that there are two spin sub-bands, in general.

Next, Fermi's golden rule for the transition rate per volume in the amorphous semiconductor can be exploited:

$$W_{nm} = \frac{1}{\Omega}\frac{d|\mathfrak{a}_{nm}|^2}{dt} = \frac{2\pi}{\hbar\Omega}|V_{nm}|^2\delta(E_n - E_m - \hbar\omega)$$
$$= \frac{\pi e^2}{2\hbar\Omega}E_0^2|x_{nm}|^2\,\delta(E_n - E_m - \hbar\omega). \qquad (2.57)$$

V_{nm} is the matrix element of the perturbation, which includes factors of E_0, the electric field amplitude, and x_{nm}, the dipole matrix element for the transition from state m to state n, the latter given by Eq. (2.48) with '0' replaced by 'm'.[57] The delta function $\delta(E_n - E_m - \hbar\omega)$ implies that the transition rate from state m to state n vanishes unless the photon energy $E = \hbar\omega$ (unsubscripted) equals the energy difference between the two electronic states. By substituting Eq. (2.57) into Eq. (2.56), one arrives at the following useful result for the photon energy-dependent absorption coefficient:

$$\frac{\alpha(\hbar\omega)n(\hbar\omega)}{\hbar\omega} = \frac{2\pi e^2\Omega}{\varepsilon_0 c\hbar}\int_{E_{0c}-\hbar\omega}^{E_{0y}} |x_{nm}|^2\rho_v(E_m)\rho_c(E_m + \hbar\omega)dE_m, \qquad (2.58)$$

where $n(\hbar\omega)$ is the photon energy-dependent index of refraction. In Eq. (2.58), E_{0v} and E_{0c} denote the electronic energies of the highest (occupied) state of the valence band and the lowest (unoccupied) state of the conduction band, respectively. Thus, $E_g = E_{0c} - E_{0v}$ is the band gap of the amorphous semiconductor, which in elemental tetrahedrally bonded solids arises from the energy separation between σ bonding and σ^* antibonding energy levels.

If one assumes over a limited energy range of E_m that:

(i) the densities of states ρ_v and ρ_c increase parabolically into the bands, separated by the gap E_g,
(ii) the dipole matrix element x_{nm} is independent of the energy of the initial and final states,

and

(iii) absorption involving the band tail and defect states near and below the bandgap of the amorphous semiconductor can be neglected, then one arrives at the simple expression:

$$\sqrt{\frac{\alpha(E)n(E)}{E}} = C(E - E_g); \quad E > E_g, \qquad (2.59)$$

where the un-subscripted E denotes the photon energy ($E = \hbar\omega$), C is a constant, and $\alpha(E) = 0$ for $E \leq E_g$.[67] One can derive an alternative expression to Eq. (2.59) by assuming instead that the momentum matrix element $p_{x,nm}$ given by Eq. (2.53c) is independent of the initial and final states. The result in this case is:

$$\sqrt{\alpha(E)n(E)E} = C'(E - E_g); \quad E > E_g, \tag{2.60}$$

along with $\alpha(E) = 0$ for $E \leq E_g$. The expression of Eq. (2.60), called the Tauc Law, has been used widely to determine the optical band gap for tetrahedrally bonded amorphous semiconductor thin films such as hydrogenated amorphous silicon (a-Si:H) and germanium (a-Ge:H) from reflectance and transmittance measurements over a relatively narrow range of E.[68] However, when higher photon energy data (for example, from spectroscopic ellipsometry) are employed in the case of a-Si:H, it has been found that the relationship of Eq. (2.59) provides a better overall fit.[46,67]

2.4.2 Direct Interband Transitions in Solids

The approach associated with Eq. (2.52c) in Section 2.4.1, can be applied assuming transitions between delocalized one-electron states described by Bloch wavefunctions. As a result, one can derive the following expression for the dielectric function of an optically isotropic crystalline solid as a function of the photon energy:

$$\varepsilon(E) = 1 + \frac{e^2\hbar^2}{\varepsilon_0 m^2} \sum_{\vec{k},c,v} \frac{|\vec{e} \cdot \vec{p}_{cv}(\vec{k})|^2}{[E_{cv}(\vec{k})]^2} \times \left(\frac{1}{E_{cv}(\vec{k}) - E - i\Gamma} + \frac{1}{E_{cv}(\vec{k}) + E + i\Gamma} \right). \tag{2.61}$$

In this equation, the electronic transitions occur from valence band (v) to conduction band (c) states each characterized by the same electron wavevector $\vec{k}$, which is unchanged (i.e., conserved) in the transition as demonstrated further below.[69] In addition, $E = \hbar\omega$ is the photon energy, $\Gamma = \hbar\gamma_p$ is the broadening energy, and $E_{cv}(\vec{k}) = E_c(\vec{k}) - E_v(\vec{k})$ is the energy difference between the pair of conduction and valence band states. Finally, $\vec{p}_{cv}(\vec{k})$ is the momentum matrix element for the transitions between valence and conduction band states, and $\vec{e}$ is the polarization

direction of the optical electric field. In fact, the second term in Eq. (2.61) is non-resonant and describes the low energy behavior of the dielectric function. Although the step-by-step details of the derivation of Eq. (2.61) are beyond the scope of this chapter, a more extensive explanation of its interpretation is required.

Basic similarities are evident between the form of Eqs. (2.50) for a collection of weakly interacting atoms with the electrons localized at atomic sites and the form of Eq. (2.61) for a collection of strongly interacting (i.e., covalently-bonded) atoms with the electrons delocalized throughout the crystal. One similarity between Eqs. (2.50) and (2.61) that may not be justified, however, in the extension from the weakly interacting to strongly interacting collections of atoms is the assumption of so-called Lorentzian broadening in Eq. (2.61), e.g., the broadening simulated in Eq. (2.50) by the exponential buildup of the optical electric field through the rate parameter γ_P. Non-Lorentzian broadening mechanisms are sure to dominate in solids, however, such as electron-phonon scattering and electron-impurity scattering, or statistical variations in the local environments of the electrons. Such mechanisms are usually better simulated by Gaussian broadening. Due to the nontransparent mathematical formulas that result from incorporating Gaussian broadening, however, the approach of Section 2.4.1.1 will be continued here and throughout Section 2.4. The phenomenological approach for incorporating Gaussian broadening mechanisms will be delayed until Section 2.5, where the modeling of optical functions is discussed in detail.

In considering next the differences between Eqs. (2.50) and Eq. (2.61), two details are immediately obvious. First, the oscillator strength is now expressed in terms of the momentum matrix element $\vec{p}_{cv}(\vec{k})$, and second, the polarization vector $\vec{e}$ of the electric field is now included explicitly. However, one aspect that makes Eq. (2.61) fundamentally different from its atomic analog is the interpretation of the summation. In deriving Eqs. (2.50) for an atomic system, one assumes that the transitions occur from a single ground state, identical for each atom. In Eq. (2.61), in contrast, one views the transitions as occurring from a set of one-electron Bloch states in *completely occupied* valence bands (v). (Partially filled bands leading to intraband transitions will be considered in Section 2.4.5.) Thus, in deriving Eq. (2.61), one uses an expression analogous to Eq. (2.52c) in which (i) the atomic wavefunction is replaced by a Bloch function, and (ii) the factor of N_a is replaced by a summation over separate terms, each one connecting an occupied Bloch state with another state. In fact, for transitions between any pair of *occupied* states (designated v_1 and v_2), the (v_1,v_2) term will cancel the (v_2,v_1) term, so that one need not be concerned with the exclusion principle. Thus for simplicity, the sum in Eq. (2.61) can be interpreted as being over pairs of occupied

valence band (v) and empty conduction (c) band electronic states. Finally, it should be noted that Eq. (2.61) arises solely from the paramagnetic conduction current, which is the first term in the expression analogous to Eq. (2.52c); the diamagnetic current vanishes in summing all possible transitions from completely filled bands (see Section 2.4.5).

A second aspect that makes Eq. (2.61) fundamentally different is the interpretation of the momentum matrix element $\vec{p}_{cv}(\vec{k})$, which provides the selection rules for the transition. The matrix element $\vec{p}_{cv}(\vec{k})$ is defined by Eq. (2.53c), but with wavefunctions $\phi_\upsilon(\vec{r},\vec{k}_\upsilon)$ given by the Bloch functions of Eq. (2.55) with $\upsilon = (v,c)$. Here v represents an occupied valence band state and c represents an unoccupied conduction band state.[20] The resulting expressions are as follows:

$$\vec{p}_{cv} = -\frac{i\hbar}{\Omega}\int_{\Omega}[u^*_{\vec{k}c}(\vec{r})\exp(-i\vec{k}_c\cdot\vec{r})]\vec{\nabla}[u_{\vec{k}v}(\vec{r})\exp(i\vec{k}_v\cdot\vec{r})]d^3\vec{r} \tag{2.62a}$$

$$= -\frac{i\hbar}{\Omega}\{\sum_L \exp[i(\vec{k}_v - \vec{k}_c)\cdot\vec{R}_L]\}\int_{cell} u^*_{\vec{k}c}\vec{\nabla}u_{\vec{k}v}d^3\vec{r}. \tag{2.62b}$$

The integrals in Eqs. (2.62a) and (2.62b) are over the excited volume Ω of the crystal, and a single unit cell of the crystal, respectively. Here $\vec{k}_v$ and $\vec{k}_c$ are the electron wavevectors for the occupied valence and unoccupied conduction band states and $u_{\vec{k}v}$ and $u_{\vec{k}c}$ are the parts of the Bloch functions having the periodicity of the lattice. The summation in Eq. (2.62b) over the unit cells of the crystal, identified by the index L and the position vector $\vec{R}_L$, vanishes unless $\vec{k}_v - \vec{k}_c = \vec{K}$, where $\vec{K}$ is a reciprocal lattice vector, defined by $(\vec{K}\cdot\vec{R}_L)/2\pi =$ integer. When $\vec{k}_v - \vec{k}_c = \vec{K}$, the summation equals the number of unit cells in the crystal. In the reduced zone scheme, $\vec{K} = 0$ and $\vec{k}_v = \vec{k}_c$, so that the transition can be described by a single value of $\vec{k} \equiv \vec{k}_v = \vec{k}_c$ in the first Brillouin zone of reciprocal space. Such transitions in which $\vec{k}_v = \vec{k}_c$ are called direct interband transitions. If the integral in Eq. (2.62b) vanishes, then the direct transitions are forbidden; otherwise they are allowed. For forbidden transitions, it is no longer valid to neglect the spatial dependence $[\exp(\pm i\vec{q}\cdot\vec{r})]$ of the vector potential, as has been done in the derivations of ε up to this point, and under these circumstances transitions analogous to the magnetic dipole and electric quadrupole transitions in atomic systems may become important.

Since $\vec{k}$ is conserved in the allowed transitions, then one can convert the summation in Eq. (2.61) to an integral over the k_x, k_y, and k_z values in the first Brillouin zone of reciprocal space. These integration variables can be converted in turn to the energy E' and two wavevector variables k_1 and k_2 that sweep out the points on a constant energy surface, S defined by

$E' = E_{cv}(\vec{k})$, in the first zone. (The prime is added to distinguish the integration variable E' from the photon energy E.) With this transformation one obtains the expression:

$$\varepsilon(E) = 1 + \frac{e^2\hbar^2}{4\pi^3\varepsilon_0 m^2}\sum_{c,v}\int dE'\int_S dk_1 dk_2 \frac{|\vec{e}\cdot\vec{p}_{cv}(\vec{k})|^2}{[E_{cv}(\vec{k})]^2|\vec{\nabla}_{\vec{k}}E_{cv}(\vec{k})|}\times \left(\frac{1}{E_{cv}(\vec{k}) - E - i\Gamma} + \frac{1}{E_{cv}(\vec{k}) + E + i\Gamma}\right), \tag{2.63}$$

where the summation over pairs of valence (v) and conduction (c) bands remains.[69] In Eq. (2.63), $|\vec{\nabla}_{\vec{k}}E_{cv}(\vec{k})|^{-1}$ is introduced in the integrand as the Jacobian of the transformation. If $\Gamma \ll E_{cv}(\vec{k})$, then the first term in parentheses under the integrand acts as a delta function centered at $E = E_{cv}(\vec{k})$. Under these conditions, the contribution to the integral is dominated by the energy difference $E' = E_{cv}(\vec{k}) \approx E$ (so the distinction between E and E' now disappears), and the following expression for $\varepsilon_2(E)$ can be derived:

$$\varepsilon_2(E) = \frac{e^2\hbar^2}{4\pi^2\varepsilon_0 m^2 E^2}\sum_{c,v}\int_S dk_1 dk_2 \frac{|\vec{e}\cdot\vec{p}_{cv}(\vec{k})|^2}{|\vec{\nabla}_{\vec{k}}E_{cv}(\vec{k})|}. \tag{2.64}$$

Here the integration is now over the surface with an energy difference $E' = E_{cv}(\vec{k})$ given by the specific value of the photon energy E.

Equation (2.64) is useful in calculating absorption spectra.[20] For example, in the case of a semiconductor with both valence and conduction bands that are parabolic and isotropic in $\vec{k}$-space, one has:

$$E_{cv}(\vec{k}) = E_g + \frac{\hbar^2 k^2}{2\mu_{cv}}, \tag{2.65a}$$

where

$$\frac{1}{\mu_{cv}} = \frac{1}{m_e} + \frac{1}{m_h}. \tag{2.65b}$$

Here μ_{cv}, m_e, and m_h are the reduced effective mass and the electron and hole effective masses, respectively. If one assumes that the momentum matrix element is nearly constant over the range of $\vec{k}$ of interest, and

defines a mean-square value of $\langle p_{cv}^2 \rangle = (1/3)\,|\vec{e} \cdot \vec{p}_{cv}(\vec{k})|^2$, then the absorption spectrum versus photon energy E is given by:

$$\alpha(E)n(E)E = \frac{3(2\mu_{cv})^{3/2}e^2\langle p_{cv}^2 \rangle}{2\pi\varepsilon_0\hbar^2 m^2 c}(E - E_g)^{1/2}; \quad E > E_g, \tag{2.66}$$

along with $\alpha(E) = 0$ for $E \leq E_g$. If the oscillator strength [from Eq. (2.53c)] is set at unity, the reduced effective mass at $m/2$, and the index of refraction n at 4, then an estimate of $\alpha = 2 \times 10^4$ cm^{-1} is obtained for $E - E_g = 0.01$ eV. This is in reasonable agreement with experimental results for Ge, for example. Although Eq. (2.66) is designed to be used in such applications near the absorption onset for the lowest energy direct transitions in semiconductors, the one-electron model breaks down under these circumstances. In fact, the inclusion of the effects of many-body interactions are required here for a complete understanding of the shape of the experimental absorption spectra. Specifically, a Coulomb interaction binds the excited electron and the hole left behind in the valence band, forming an exciton. The optical behavior of excitons will be referred to briefly in Section 2.4.3; a more extensive review is given by Cardona.[32]

2.4.3 Band Structure and Critical Points in Solids

One can relate the dielectric function more closely to the band structure by noting that the joint density of states, $J_{cv}(E')$, defined as the number of single-spin (i.e., not including the factor of two for spin) electronic states per volume per energy interval, separated by an energy E' is given by:

$$J_{cv}(E') = \frac{1}{8\pi^3}\int_S \frac{dk_1 dk_2}{|\vec{\nabla}_{\vec{k}} E_{cv}(\vec{k})|}. \tag{2.67a}$$

Here the integration is performed over the constant energy surface S in $\vec{k}$-space with $E_{cv}(\vec{k}) = E'$.[20] If the dependence of the momentum matrix element on k_1 and k_2 can be neglected, then Eq. (2.63) can be rewritten as:

$$\varepsilon(E) = 1 + \frac{2e^2\hbar^2}{\varepsilon_0 m^2}\sum_{c,v}\int J_{cv}(E')\frac{|\vec{e} \cdot \vec{p}_{cv}(E')|^2}{E'^2} \times \left(\frac{1}{E' - E - i\Gamma} + \frac{1}{E' + E + i\Gamma}\right)dE'. \tag{2.67b}$$

Again if $\Gamma \ll E_{cv}(\vec{k})$, then the first term in parentheses under the integrand acts as a delta function centered at $E' = E_{cv}(\vec{k}) = E$ and the contribution to the integral is dominated by the energy difference $E' \approx E$. As a result, an expression for $\varepsilon_2(E)$ corresponding to Eq. (2.64) can be derived:

$$\varepsilon_2(E) = \frac{2\pi e^2 \hbar^2}{\varepsilon_0 m^2 E^2} \sum_{c,v} |\vec{e} \cdot \vec{p}_{cv}(E)|^2 J_{cv}(E). \tag{2.67c}$$

This expression shows clearly the close relationship between the imaginary part of the dielectric function and the joint density of electronic states.

The joint density of states in Eq. (2.67a), and hence the dielectric functions of Eqs. (2.67b) and (2.67c) exhibit relatively sharp structure when $\vec{\nabla}_{\vec{k}} E_{cv}(\vec{k}) = 0$, and points in $\vec{k}$-space where this condition is met are called critical points or van Hove singularities. At the photon energy E corresponding to a critical point, the primary contribution to the dielectric function in Eq. (2.63) arises from the region of $\vec{k}$-space centered on the critical point, and this leads to the structure in the dielectric function. In the vicinity of a semiconductor critical point, one can replace the normally complicated variation $E_{cv}(\vec{k})$ in Eq. (2.63) with a local parabolic expansion:

$$E_{cv}(\vec{k}) = E_g + \frac{\hbar^2}{2}\left(\frac{k_x^2}{\mu_{xx}} + \frac{k_y^2}{\mu_{yy}} + \frac{k_z^2}{\mu_{zz}}\right), \tag{2.68}$$

where $\vec{k}$ is expressed in a coordinate system with the origin at the critical point, oriented so as to diagonalize the inverse of the reduced effective mass tensor, $[\mu]^{-1}$.

By substituting Eq. (2.68) into the integral of Eq. (2.63), neglecting the non-resonant term, and assuming that the matrix element is only a weak function of $\vec{k}$ (and can be brought outside the integral), one finds:

$$\varepsilon(E) = \frac{Q}{E^2} i^{n_c+1} U_x K_{cy} K_{cz} (E - E_g + i\Gamma)^{-1/2}; \quad 1D, \tag{2.69a}$$

$$\varepsilon(E) = \frac{Q}{E^2} i^{n_c+2} U_x U_y K_{cz} \ln(E - E_g + i\Gamma); \quad 2D, \tag{2.69b}$$

$$\varepsilon(E) = \frac{2\pi Q}{E^2} i^{n_c+1} U_x U_y U_z (E - E_g + i\Gamma)^{1/2}; \quad 3D. \tag{2.69c}$$

Here, expressions are given for one-, two-, and three-dimensional critical points for which a pair, one, or none of the effective masses μ_{jj} is very large, respectively.[69] In addition in Eqs. (2.69), $Q = (e^2\hbar^2/4\pi^2\varepsilon_0 m^2)|\vec{e} \cdot \vec{p}_{cv}(\vec{k})|^2$; $U_j = (2|\mu_{jj}|/\hbar^2)^{1/2}$, $(j = x, y, z)$; and n_c is the critical point type, designated by the number of effective masses that are negative. For a three-dimensional critical point, the following assignments can be made: ($n_c = 0 \Rightarrow$ minimum; $n_c = 1 \Rightarrow$ saddle point; $n_c = 2 \Rightarrow$ saddle point; $n_c = 3 \Rightarrow$ maximum). These are called the M_0, M_1, M_2, and M_3 critical point types, respectively. For a two-dimensional critical point, the following assignments are made: ($n_c = 0 \Rightarrow$ minimum; $n_c = 1 \Rightarrow$ saddle point; $n_c = 2 \Rightarrow$ maximum). Finally, for a one-dimensional critical point: ($n_c = 0 \Rightarrow$ minimum; $n_c = 1 \Rightarrow$ maximum). K_{cj} $(j = y, z)$ are the cutoff lengths in the Brillouin zone that are imposed to restrict the limits of integration in Eq. (2.63) along the z-direction for a two-dimensional critical point (in which case $\mu_{zz} \to \infty$) and along the y- and z-directions for a one-dimensional critical point (in which case $\mu_{yy} \to \infty$ and $\mu_{zz} \to \infty$). In fact, Eqs. (2.69) are clearly not consistent with the Kramers-Kronig relationships [Eqs. (2.29)] and are only valid for calculating difference or derivative spectra (i.e., in so-called critical point analyses) in which the weaker, smoothly varying background in ε from other regions of $\vec{k}$-space completely disappears.

A general form of the dielectric function associated with a critical point can be written as:

$$\varepsilon(E) = A\Gamma^p \exp(i\varphi)\,(E - E_g + i\Gamma)^{-p}, \tag{2.70}$$

where A is an amplitude factor, φ is the phase projection factor, and $-p$ is the exponent. The exponent $-p$ takes on the values of $-1/2$, 0 (logarithmic), and $1/2$ for 1D, 2D, and 3D critical points, respectively. With Eq. (2.70) one can also describe the critical point lineshape for discrete excitons, by setting $p = 1$. In this case, the phase factor is interpreted as a coupling parameter between the exciton and the overlapping interband continuum. Again it must be emphasized that the functional form of Eq. (2.70) is not Kramers-Kronig consistent and is valid only in critical point analyses. Finally, Eqs. (2.69) and (2.70) are based on the assumption of Lorentzian broadening with an energy width Γ; Gaussian broadening, which is perhaps more realistic, will be treated in Section 2.5.

Table 2-2 includes the functional form of $J_{cv}(E)$ in the neighborhood of one-, two-, and three-dimensional critical points of different types having allowed transitions as calculated from Eq. (2.67a); Fig. 2.4 depicts schematics of the shapes of these critical points.[32] For transitions that are forbidden at the critical point, but allowed in moving away from the point, the momentum matrix element in Eqs. (2.67b) and (2.67c) is replaced by a factor of $|E - E_g|$. This factor arises from a first-order expansion of the matrix element about the critical point.

Table 2-2. The form of the joint density of electronic states J_{cv} versus photon energy E [see Eqs. (2.67) and Fig. 2.4] for allowed transitions in the neighborhood of band structure critical points of different types. The dimensionality of the critical point is determined by the number of principal components of the reduced effective mass tensor μ_{jj} ($j = x, y, z$) that are infinite. The type of the critical point is determined by the number of components of the diagonalized effective mass tensor that are negative. Weakly varying background contributions to the joint density of states are assumed in each case. In the expressions for $J_{cv}(E)$, c_{mn} and c_{mn}' (m = 0,1,2,3; n = 1,2,3) are constants independent of photon energy.

Dim.		Type	μ_{xx}	μ_{yy}	μ_{zz}	$J_{cv}(E)$	
						$E < E_g$	$E > E_g$
3D	M_0	minimum	+	+	+	c_{03}	$c_{03} + c_{03}'(E - E_g)^{1/2}$
3D	M_1	saddle	+	+	−	$c_{13} - c_{13}'(E_g - E)^{1/2}$	c_{13}
3D	M_2	saddle	+	−	−	c_{23}	$c_{23} - c_{23}'(E - E_g)^{1/2}$
3D	M_3	maximum	−	−	−	$c_{33} + c_{33}'(E_g - E)^{1/2}$	c_{33}
2D	M_0	minimum	+	+	∞	c_{02}	$c_{02}' > c_{02}$
2D	M_1	saddle	+	−	∞	$c_{12} - c_{12}' \ln(E_g - E)$	$c_{12} - c_{12}'' \ln(E - E_g)$
2D	M_2	maximum	−	−	∞	c_{22}	$c_{22}' < c_{22}$
1D	—	minimum	+	∞	∞	c_{01}	$c_{01} + c_{01}'(E - E_g)^{-1/2}$
1D	—	maximum	−	∞	∞	$c_{11} + c_{11}'(E_g - E)^{-1/2}$	c_{11}

2.4.4 Indirect Interband Transitions in Solids

The interactions of an electron with the periodic array of atoms in a crystalline solid are described, within the one-electron model, by introducing concepts such as the electron crystal momentum $\hbar\vec{k}$, the effective mass, and the one-electron band structure of allowed energy states vs. $\vec{k}$. Then the interaction of a classical electromagnetic wave with an electron in a perfect lattice leads to direct transitions between electron states in which $\vec{k}$ is conserved [see Eqs. (2.61) and (2.62)]. If one considers possible interactions of the electron with imperfections of the lattice, then $\vec{k}$-conservation in the electron transitions is relaxed, and this gives rise to indirect transitions.[23] Quantized lattice vibrations, or phonons, constitute an important source of imperfections and are unavoidable at finite temperature even in a perfect crystal. Furthermore, even in the limit of 0 K, zero-point motion occurs; in addition, the electron undergoing a transition can interact with the lattice to generate a phonon. Specifically, transitions

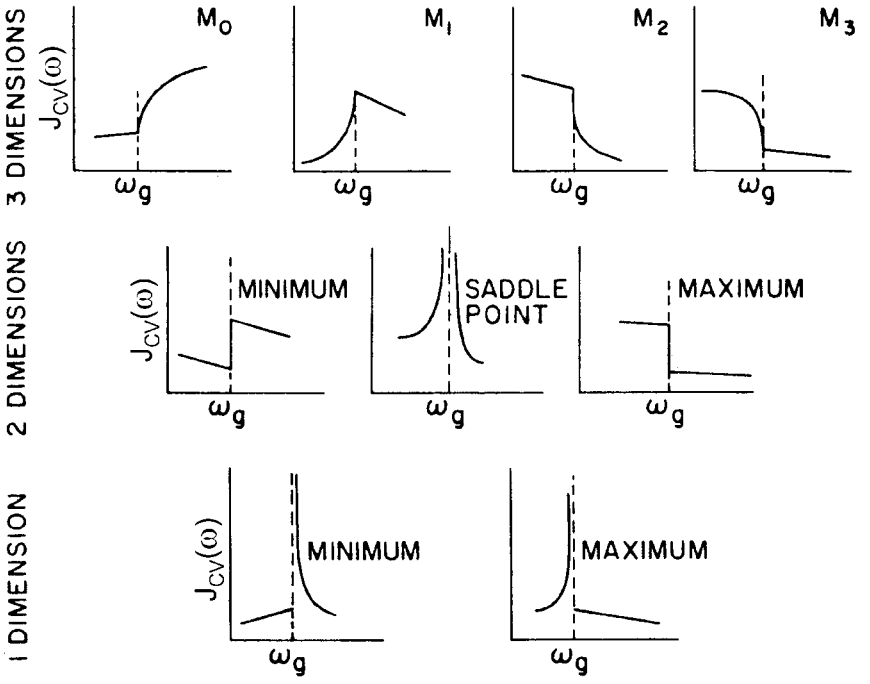


Figure 2.4 Behavior of the joint density of states, as well as the imaginary part of the dielectric function in the absence of broadening [see Eq. (2.67c)], versus optical frequency in the neighborhood of one-, two-, and three-dimensional critical points of different types. The critical point is placed at $\omega_g = E_g/\hbar$, and the functional dependences are given in Table 2-2. (After Ref. 32.)

of an electron from an initial state $\phi_v(\vec{r}, \vec{k}_v)$ to a final state $\phi_c(\vec{r}, \vec{k}_c)$ with $\vec{k}_v \neq \vec{k}_c$ can occur if a phonon of wavevector $\vec{k}_p = \pm(\vec{k}_c - \vec{k}_v) + \vec{K}$ is emitted $(-)$ or absorbed $(+)$. Here $\vec{K}$ is a reciprocal lattice vector. In addition to conservation of crystal momentum, as embodied in this expression, overall conservation of energy also applies, given by $\hbar\omega = E_{cv} - (\pm\hbar\omega_{p\vec{k}})$, where $\hbar\omega$ is the energy of the photon and $\hbar\omega_{p\vec{k}}$ is the energy of the emitted $(-)$ or absorbed $(+)$ phonon.

Because such indirect processes are second-order, involving perturbations described by the electron-photon and electron-phonon coupling Hamiltonians, the resulting transitions will be observed experimentally only when no direct (first-order) processes occur at the same photon energy. Thus, the practical situation of interest is one that occurs in an indirect semiconductor when the photon energy $\hbar\omega$ is below the minimum direct gap (so that no direct transitions can occur), but above the minimum indirect gap.

To describe the situation mathematically, one must resort to second-order perturbation theory. It is sufficient here to consider only (i) transitions from a single valence band maximum at $\vec{k} = 0$ to a single conduction band minimum away from the Brillouin zone center, and (ii) single phonon absorption or emission. In addition, all other valence and conduction bands are assumed to be distant in energy so that only two intermediate states designated '*a*' and '*b*' need to be considered as shown in Fig. 2.5. Under these circumstances the transition rate per unit volume is given by:

$$W_{cv} = \frac{\pi e^2 A_0^2}{2\Omega m^2 \hbar^2} \left(\frac{|V_{ca}^{phon}|^2 \, |\vec{e} \cdot \vec{p}_{av}|^2}{(E_{av} - E)^2} + \frac{|V_{bv}^{phon}|^2 \, |\vec{e} \cdot \vec{p}_{cb}|^2}{(E - E_{cb})^2} \right)$$

$$\times \{\delta(E_{cv} - \hbar\omega_{p\vec{k}} - \hbar\omega) + \delta(E_{cv} + \hbar\omega_{p\vec{k}} - \hbar\omega)\}. \quad (2.71)$$

In this expression, A_0 is the amplitude of the vector potential, and the two terms incorporating the matrix elements represent the two possible pathways via the intermediate conduction band state '*a*' and the intermediate valence band state '*b*'.[20] In these terms, $|V_{ca}^{phon}|^2$ and $|V_{bv}^{phon}|^2$ are the matrix elements of the electron-phonon coupling Hamiltonian which, unlike electron-photon coupling, must include spatial dispersion. If the initial state of the electron is characterized by $\vec{k}_v = 0$, then both these matrix

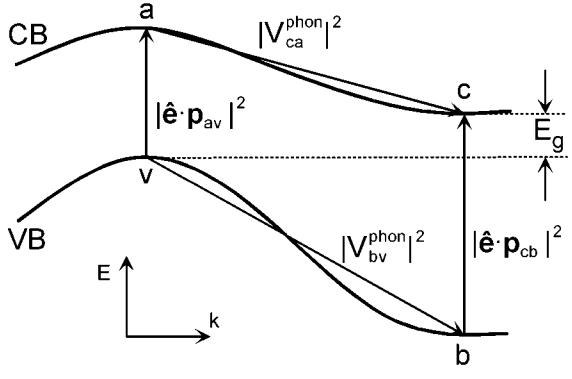


Figure 2.5 Semiconductor band diagram showing two possible pathways for the indirect excitation of an electron from the valence band (VB) maximum *v* to the conduction band (CB) minimum *c*. In one case, the electron is excited in a virtual direct transition to *a* followed by phonon-assisted scattering to *c*. In the other case, an electron is excited in a direct transition from *b* to *c*, and the hole left behind is scattered to *v*. The matrix elements describing the electron-photon and electron-phonon coupling for the transitions are included from Eq. (2.71).

elements vanish unless crystal momentum is conserved in accordance with $\vec{k}_p = \pm\vec{k}_c + \vec{K}$. The delta-function terms in Eq. (2.71) describe energy conservation corresponding to absorption ($-\hbar\omega_{p\vec{k}}$) and emission ($+\hbar\omega_{p\vec{k}}$) of phonons, where $E_{cv} = E_g$ is the indirect gap. Normally, one of the two matrix element terms will dominate due to the squared-energy-difference factors in the denominators. These factors reflect the fact that the transition rates for the two pathways increase as the square of the lifetimes in the intermediate states. According to the uncertainty principle, these lifetimes are given by $\tau_a \sim \hbar/(E_{av} - E)$ and $\tau_b \sim \hbar/(E_{cb} - E)$, for states '$a$' and '$b$'. Here, $E_{av} = E_a - E_v$, and $E_{cb} = E_c - E_b$, where E_a and E_b are the electron energies for the two intermediate states.

In practice, transitions according to Eq. (2.71) occur between two distributions of states near the valence band maximum and the conduction band minimum with a joint density of states $J_{cv}(E') = \rho_v(E')\rho_c(E' + \hbar\omega \pm \hbar\omega_{p\vec{k}})$, where ρ_v and ρ_c are typically parabolic functions. Then by integrating over all possible transitions in an approach similar to that of Eq. (2.56) one can arrive at the following practical formulas that incorporate the absorption coefficient [assuming the first matrix element term in Eq. (2.71) dominates]:

$$\alpha(E)n(E)E(E_{av} - E)^2 = \frac{C}{\exp(\beta_{\vec{k}}) - 1}(E + \hbar\omega_{p\vec{k}} - E_g)^2; \quad E > E_{cv} - \hbar\omega_{p\vec{k}}, \tag{2.72a}$$

$$\alpha(E)n(E)E(E_{av} - E)^2 = \frac{C'\exp(\beta_{\vec{k}})}{\exp(\beta_{\vec{k}}) - 1}(E - \hbar\omega_{p\vec{k}} - E_g)^2; \quad E > E_{cv} + \hbar\omega_{p\vec{k}}. \tag{2.72b}$$

In these expressions, $E = \hbar\omega$ is the photon energy, C and C' are constants, and $\beta_{\vec{k}} \equiv \hbar\omega_{p\vec{k}}/k_BT$, where k_B is Boltzmann's constant.[32] For photon energies below the specified ranges, $\alpha(E) = 0$. Equation (2.72a) is appropriate for phonon absorption and the temperature (T) dependent term arises from the phonon concentration N_p which has been extracted from the electron-phonon matrix element. Similarly, Eq. (2.72b) is appropriate for phonon emission with probability proportional to $N_p + 1$. Over a sufficiently narrow range one can neglect all photon-energy dependent terms on the left except for α in applying these formulas to identify indirect transitions.

Finally, some qualifications are required before accepting Eqs. (2.72) as a valid description of the indirect absorption onset. First, in deriving

Eqs. (2.72), one assumes that the momentum matrix element connecting the valence and conduction band states near $\vec{k} = 0$ is independent of energy. This is invalid when the dominant direct transition here is forbidden.[20] A derivation in this case leads to analogous equations to those of Eqs. (2.72) but with the quadratic dependence $(E \pm \hbar\omega_{p\vec{k}} - E_g)^2$ replaced by the cubic dependence $(E \pm \hbar\omega_{p\vec{k}} - E_g)^3$. Second, when excitons are formed, the results for α near the absorption onset of the indirect gap given by Eqs. (2.72) are modified, as in the case of the direct transitions.[32]

2.4.5 Intraband Transitions in Metals

In the final part of this section, intraband transitions will be briefly considered from the standpoint of the quantum theory of solids. For the purposes of this discussion, it is assumed that the interband contribution is non-dispersive over the photon energy range of interest, so that this energy range must lie below the minimum interband transition energy. This situation holds over certain ranges for monovalent metals and degenerate semiconductors.

Starting from Eq. (2.52c) as a foundation, the following result can be derived for the more general case of an optically anisotropic crystal:

$$\varepsilon_{jj} = (\varepsilon_{0s})_{jj} - \frac{e^2/4\pi^3\varepsilon_0\hbar^2}{\omega[\omega + i(2\Gamma_{mj}/\hbar)]} \int \frac{\partial^2 E(\vec{k})}{\partial k_j^2} d^3\vec{k}, \qquad (2.73a)$$

where $j = (x, y, z)$ describes the coordinate system that diagonalizes the real and imaginary parts of the complex dielectric tensor.[32] Thus, a crystal of orthorhombic or higher symmetry is assumed. The first term on the right in Eq. (2.73a), $(\varepsilon_{0s})_{jj}$, originates from the interband term of Eq. (2.61) when $E = \hbar\omega \ll E_{cv}(\vec{k})$, for all $\vec{k}$. The energy-dependent second term originates from the diamagnetic current [see Eq. (2.52c)] and is non-zero for metals and degenerate semiconductors having partially filled conduction bands. The integral in the second term of Eq. (2.73a) is performed over all occupied states in $\vec{k}$-space, and because it vanishes for completely filled bands, this second term does not appear in Eq. (2.61). In the prefactor of the second term, $\Gamma_{mj} = \hbar(\gamma_{Pm})_j$ has been introduced to ensure a finite lifetime for the transition (even in the absence of scattering by phonons and defects, the typical lifetime limiting mechanism for electrons in metals). Here $(\gamma_{Pm})_j$ describes the exponentially increasing magnitude of the j^{th} component of the vector potential as defined in Eq. (2.46). Finally, the wavevector $\vec{k}$ under the integral in Eq. (2.73a) is measured using the same coordinate system that diagonalizes the complex dielectric tensor.

Two important steps are required to derive the second term in Eq. (2.73a) from the diamagnetic current. First, the zero-order (or unperturbed) wavefunctions are incorporated to obtain a first-order result in the perturbation expansion. Second, the effective mass theorem, also called the $\vec{k} \cdot \vec{p}$ sum rule, is applied given by:

$$\frac{1}{\hbar^2}\frac{\partial^2 E_c(\vec{k})}{\partial k_j^{\,2}} = \frac{1}{m} + \frac{2}{m^2}\sum_{\upsilon \neq c}\frac{[\vec{p}_{c\upsilon}(\vec{k})]_j[\vec{p}_{\upsilon c}(\vec{k})]_j}{E_c(\vec{k}) - E_\upsilon(\vec{k})}. \tag{2.73b}$$

Here, 'c' designates an occupied state in the partially filled conduction band and υ designates a state in another band.[32] The quantity $[\vec{p}_{c\upsilon}(\vec{k})]_j$ is the j^{th} vector component of the momentum matrix element taken between the states $(c, \vec{k})$ and $(\upsilon, \vec{k})$ defined as in Eq. (2.62a), and $E_c(\vec{k})$ and $E_\upsilon(\vec{k})$ are the electron energies of these states. In fact, the quantity on the left-hand side is the j^{th} diagonal element of the inverse effective mass tensor. The effective mass theorem can be derived from a comparison of the second-order terms in a Taylor series expansion of the electron energy about a point in $\vec{k}$-space with the corresponding second-order terms in a perturbation theory analysis.[45]

It is of interest to evaluate Equation (2.73a) for the practical situation of a cubic crystal in which there is a single set of equivalent, parabolic conduction band minima, so that in the vicinity of each:

$$E(\vec{k}) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_{e,xx}} + \frac{k_y^2}{m_{e,yy}} + \frac{k_z^2}{m_{e,zz}}\right). \tag{2.74}$$

With a cubic crystal, the coordinate system for $\vec{k}$ has now been fixed in turn at each of the conduction band minima so as to coincide with the principal directions of the inverse effective mass tensor. Thus, $(m_{e,\,jj})^{-1}$; $j = (x, y, z)$, denote the principal components of the inverse effective mass tensor. Substituting Eq. (2.74) into Eq. (2.73a), yields:

$$\varepsilon(\omega) = \varepsilon_{0s} - \frac{N_{ec}e^2/\varepsilon_0 m_o}{\omega[\omega + i(2\Gamma_m/\hbar)]}, \tag{2.75a}$$

where N_{ec} is the electron concentration in the partially filled conduction band and m_o is often called the optical effective mass, given by

$$\frac{1}{m_o} = \frac{1}{3}\left(\frac{1}{m_{e,xx}} + \frac{1}{m_{e,yy}} + \frac{1}{m_{e,zz}}\right).$$

Equation (2.75a) is similar in form to the first two terms of Eqs. (2.38), derived classically. In these classical formulas, if one were to make the

following associations:

$$\text{Eqs. (2.38)} \rightarrow \text{Eq. (2.75a)},$$

$$N_{ef} \rightarrow N_{ec}, \tag{2.75b}$$

$$m \rightarrow m_o, \tag{2.75c}$$

$$\gamma_f \rightarrow 2\Gamma_m/\hbar, \tag{2.75d}$$

then the quantum mechanically derived expression of Eq. (2.75a) results.

2.5 Modeling the Optical Properties of Solids

In this section, examples will be presented that illustrate the theoretical derivations and concepts of Sections 2.3 and 2.4. In the first part, representative results are given for the classical Lorentz oscillator and Drude expressions in terms of the real and imaginary parts of the complex dielectric functions (ε_1, ε_2), the real and imaginary parts of the complex index of refraction (n, κ), as well as the absorption coefficient and normal-incidence reflectance (α, R). The results for (α, R) provide an intuitive understanding of the physical manifestations of the optical functions. The theoretical results from the Lorentz oscillator and Drude expressions for the dielectric functions are compared qualitatively and quantitatively to experimental results for a selection of different materials, including amorphous silicon (a-Si:H) in thin film form,[46] aluminum (Al) and silver (Ag) in particulate film form,[47,48] bulk single-crystal gallium arsenide (GaAs),[49,50] and thin film polycrystalline Ag.[36] In the second part of this section, more detailed expressions developed on the basis of quantum mechanical principles are employed specifically to fit a similar group of materials, including amorphous silicon-germanium (a-$Si_{1-x}Ge_x$:H) and amorphous silicon-carbon (a-$Si_{1-x}C_x$:H) alloy thin films,[51,52] bulk single-crystal Si and GaAs,[53,54] epitaxial mercury cadmium telluride ($Hg_{1-x}Cd_xTe$) alloy layers,[55] and thin film Al.[47]

2.5.1 Classical Lorentz Oscillator Models

2.5.1.1 Optical Characteristics: Semiconductors, Insulators, and Confined Electron Systems

Figure 2.6 shows the photon energy ($E = \hbar\omega$) dependence of the complex dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$ for the Lorentz oscillator model,

calculated from Eqs. (2.32) with a single term in each summation. Thus single-electron "two-level atoms" ($N_e = N_a$) are assumed for simplicity. Each electron is assumed to be bound to its atom with a resonant frequency of ω_0 in the near-ultraviolet range; as a result, there is no free-electron or Drude component to the dielectric function. The behavior in Fig. 2.6 exhibits some of the optical characteristics of a semiconductor, in particular an amorphous semiconductor as will be seen shortly, owing to the large value of γ (or the short transition lifetime) selected for the particular example of Fig. 2.6. In fact, the values for the parameters of the Lorentz oscillator are chosen for qualitative consistency with the measured dielectric function of amorphous silicon (a-Si:H), namely, a resonance energy of $E_0 \equiv \hbar\omega_0 = 3.85$ eV, a broadening energy of $\Gamma \equiv \hbar\gamma = 2.00$ eV, and a transition strength of $A^2 \equiv (e^2\hbar^2 N_e/\varepsilon_0 m) = (15.1 \text{ eV})^2$.

In the limit of low optical wave frequencies in Fig. 2.6, i.e., for $\omega \ll \omega_0$ (or $E \ll E_0$), and when $\gamma < \omega_0$ (or $\Gamma < E_0$), ε_1 approaches a constant given by $\varepsilon_1 = 1 + (e^2 N_e/\varepsilon_0 m\omega_0^2) = 1 + (A/E_0)^2$. The value of this

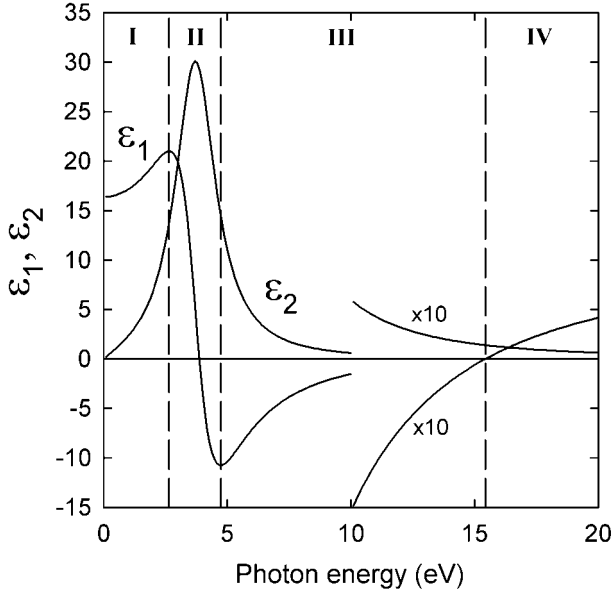


Figure 2.6 Photon energy dependence of the real (ε_1) and imaginary (ε_2) parts of the complex dielectric function (ε) for a single Lorentz oscillator with resonance energy, $\hbar\omega_0 = 3.85$ eV; broadening, $\hbar\gamma = 2.00$ eV; and amplitude, $A^2 \equiv e^2\hbar^2 N_e/\varepsilon_0 m = (15.1 \text{ eV})^2$. Normal dispersion in ε_1 is observed in photon energy regimes I and III−IV, below and above the resonance energy, respectively. In regime II, anomalous dispersion is observed. The boundary between regimes III and IV is placed at the plasma energy, $\hbar\omega_p$, defined where $\varepsilon_1 = 0$.

constant is 16.4 for the selected parameters in Fig. 2.6. When $\omega \ll \omega_0$ and $\gamma < \omega_0$, the real part of the dielectric function increases with increasing ω according to

$$\varepsilon_1(\omega) \approx 1 + \frac{e^2 N_e}{\varepsilon_0 m} \frac{1}{\omega_0^2 - \omega^2} = 1 + \frac{A^2}{E_0^2 - E^2}.$$

This is the regime of so-called "normal dispersion" in ε_1 that extends to

$$\omega_- = \omega_0[1 - (\gamma/\omega_0)]^{1/2} \{\text{or } E_- \equiv \hbar\omega_- = E_0[1 - (\Gamma/E_0)]^{1/2} = 2.67 \text{ eV}\}.$$

Above ω_-, ε_1 decreases with ω, first crossing unity at $\omega = \omega_0$ (or $E = E_0 = 3.85$ eV) and then crossing zero from above at

$$\omega \approx \omega_0[1-(\varepsilon_0 m/e^2 N_e)\gamma^2]^{-1/2} \{\text{or } E \approx E_0[1-(\Gamma/A)^2]^{-1/2} = 4.13 \text{ eV}\}.$$

[The latter approximation is valid when $(\varepsilon_0 m/e^2 N_e)\gamma^2 = (\Gamma/A)^2 \ll 1$.] The regime in which ε_1 decreases with ω is the regime of so-called "anomalous dispersion" in ε_1 that extends to

$$\omega_+ = \omega_0[1 + (\gamma/\omega_0)]^{1/2} \{\text{or } E_+ \equiv \hbar\omega_+ = E_0[1 + (\Gamma/E_0)]^{1/2} = 4.75 \text{ eV}\}.$$

When $\omega_0 \gg \gamma$, the regime of anomalous dispersion in ε_1 exhibits a width of $\omega_+ - \omega_- \approx \gamma$. In spite of the fact that ω_0 and γ are of similar order for the chosen parameters in Fig. 2.6, $\hbar(\omega_+ - \omega_-)$ yields a value of 2.08 eV, not far from the approximate value $\Gamma = 2.00$ eV. Above ω_+, ε_1 resumes normal dispersion, crossing zero from below at the plasma frequency of $\omega_p = [\omega_0^2 + (e^2 N_e/\varepsilon_0 m)]^{1/2}$ {or at the plasma energy of $E_p \equiv \hbar\omega_p = (E_0^2 + A^2)^{1/2} = 15.6$ eV}. For $\omega \gg \omega_0 \gg \gamma$, ε_1 approaches unity asymptotically as

$$\varepsilon_1(\omega) \sim 1 - (e^2 N_e/\varepsilon_0 m\omega^2) = 1 - (A/E)^2.$$

In the limit of decreasing frequency in Fig. 2.6, i.e., for $\omega \ll \omega_0$ (or $E \ll E_0$), ε_2 approaches zero asymptotically as $\varepsilon_2(\omega) \sim e^2 N_e \gamma\omega/\varepsilon_0 m\omega_0^4 = A^2\Gamma E/E_0^4$. In this aspect, the Lorentz oscillator function differs from typical semiconductor characteristics in which $\varepsilon_2(E)$ approaches zero rapidly as E approaches the bandgap E_g from above. Furthermore, in Fig. 2.6, ε_2 increases with increasing ω, reaching its maximum value of $\varepsilon_2 \approx e^2 N_e/\varepsilon_0 m\gamma\omega_0 = A^2/\Gamma E_0$ (=29.6) at $\omega = \omega_0$. For $\omega \gg \omega_0$, ε_2 approaches zero asymptotically as $\varepsilon_2(\omega) \sim N_e e^2\gamma/\varepsilon_0 m\omega^3 = A^2\Gamma/E^3$. Finally, when $\omega_0 \gg \gamma$, the full width at half-maximum $\Delta\omega$ of the resonance band in $\varepsilon_2(\omega)$ is also given by $\Delta\omega = \gamma$ (or $\Delta E = \Gamma$).

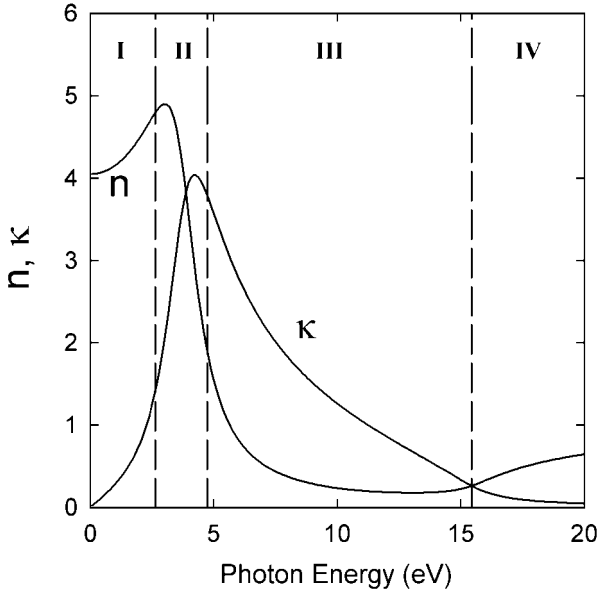


Figure 2.7 Photon energy dependence of the real (n) and imaginary (κ) parts of the complex index of refraction (N) for the Lorentz oscillator of Fig. 2.6. The four regimes are delineated as described in Fig. 2.6.

In Fig. 2.7, the corresponding frequency dependence of the complex index of refraction $N = n + i\kappa$ is plotted for the same single Lorentz oscillator. The index of refraction n and extinction coefficient κ exhibit features similar to those of ε_1 and ε_2 in Fig. 2.6 with the exception that n remains positive and the regime of anomalous dispersion in n is extended to higher ω. Figure 2.8 includes both the absorption coefficient and normal-incidence reflectance. The former is calculated from the extinction coefficient and wavelength according to Eq. (2.12b). The latter is calculated from the index of refraction and extinction coefficient, using the expression $R = [(n - 1)^2 + \kappa^2]/[(n + 1)^2 + \kappa^2]$, assuming that the index of refraction of the ambient medium is unity.

In Figs. 2.6 through 2.8, four regimes are delineated as described by Wooten and classical interpretations are suggested as follows.[20] In regime I corresponding to $\omega < \omega_-$(or $E < E_-$), i.e., the first regime of normal dispersion in ε_1, the electrons respond as if they were strongly bound, and the solid exhibits relatively weak, but increasing absorption and reflectance. The reflectance in this regime is lower for solids with a lower electron density N_e and higher electronic "spring constant" $m\omega_0^2$. Regime II is the regime of anomalous dispersion in ε_1 with $\omega_0 - (\gamma/2) \leq \omega \leq \omega_0 + (\gamma/2)$

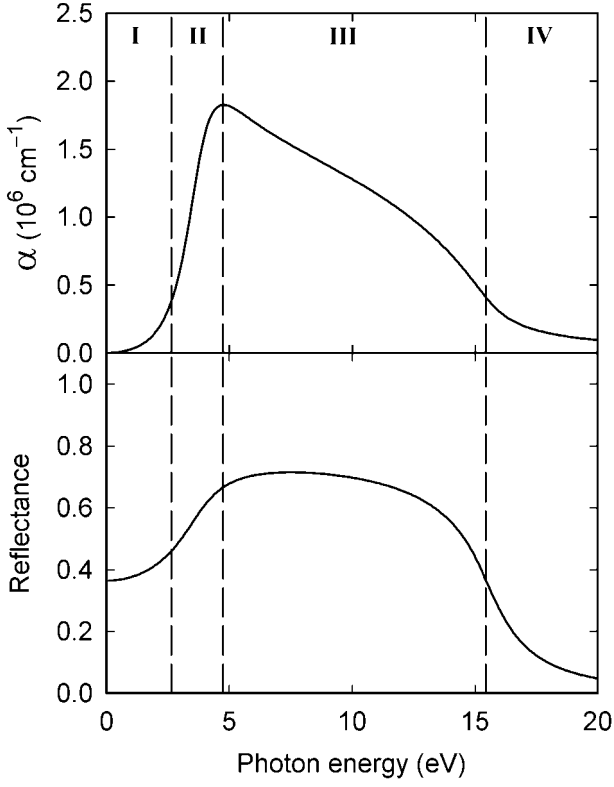


Figure 2.8 Photon energy dependence of the absorption coefficient (α, top) and the normal incidence reflectance (bottom) for a hypothetical material whose dielectric function is given by the Lorentz oscillator of Fig. 2.6. In the calculation of the reflectance, a single interface between the ambient (with index of refraction $n_a = 1$) and the material is considered. The four regimes are delineated as described in Fig. 2.6. (After Ref. 36)

[or $E_0 - (\Gamma/2) \leq E \leq E_0 + (\Gamma/2)$]. This regime is characterized by a response consistent with weakly bound conduction electrons in that a sharply increasing absorption and reflectance is observed as a function of frequency. In regime III, $\omega > \omega_+$ and the electrons respond as if they were free electrons in a metal [see Eqs. (2.35) in Section 2.3]. As a result, the solid has a metallic-like reflectance in this regime, but gradually decreasing absorption (as well as reflectance) with increasing ω. The onset of regime IV occurs when $\varepsilon_1 = 0$, which defines the plasma frequency ω_p (or plasma energy E_p). In regime IV, both the reflectance and absorption are relatively low and decrease gradually to zero.

2.5.1.2 Examples: Semiconductors, Insulators, and Confined Electron Systems

Figures 2.9 and 2.10 show the dielectric functions obtained by ex situ and real time spectroscopic ellipsometry (SE), respectively, for three thin film materials, hydrogenated amorphous silicon (a-Si:H)[46] and discontinuous elemental aluminum (Al) and silver (Ag),[47,48] for which single broad resonances are apparent, similar to that in Fig. 2.6.

Although the dielectric function of the a-Si:H is qualitatively similar to that of the single oscillator in Fig. 2.6, a precise fit cannot be obtained for at least two reasons associated with the oversimplified classical Lorentz model. First of all, solids in general are more realistically described in terms of a large collection of oscillators enumerated by n, as in Eqs. (2.32), with closely spaced resonance frequencies ω_n spread over bands, and each of the individual resonances has a width γ_n and electron concentration $N_{eb,n}$ associated with it. In the specific case of a-Si:H, there is a single broad band of resonances typical of an amorphous solid, that can be understood from the standpoint of quantum mechanics in terms of

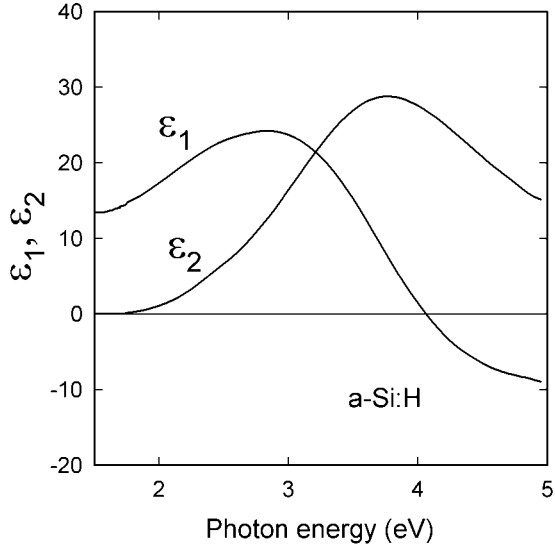


Figure 2.9 Real and imaginary parts of the dielectric function (ε_1, ε_2) of hydrogenated amorphous silicon measured at room temperature by combining spectroscopic ellipsometry, transmittance and reflectance spectroscopy, and dual beam photoconductive spectroscopy on a thin film prepared on glass by plasma-enhanced chemical vapor deposition. (Data from Ref. 46.)

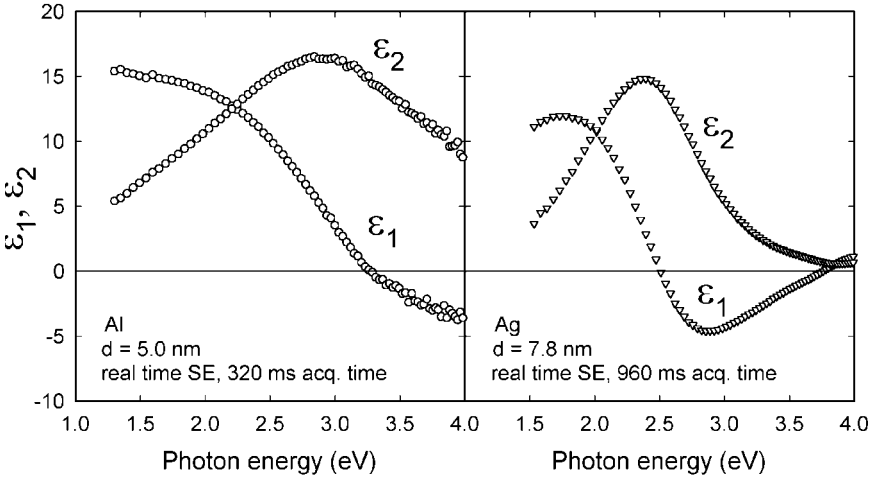


Figure 2.10 Real and imaginary parts of the dielectric functions (ε_1, ε_2) of thin films consisting of isolated aluminum (left) and silver (right) particles on oxide-covered Si substrates, measured by spectroscopic ellipsometry (SE) in real time during preparation at room temperature. The film thicknesses at the measurement times were 5.0 nm and 7.8 nm for the Al and Ag films, respectively. The real time SE acquisition times for the full spectra are provided. (Unpublished data from H.V. Nguyen and R.W. Collins.)

the excitation of valence electrons (four per Si atom) from a band of occupied σ bonding energy states across an energy gap (~1.65 eV) to a band of unoccupied σ^* antibonding states.[70] Second, even in amorphous solids, an energy gap E_g develops between valence and conduction band states. This gap is manifested by a decrease in $\varepsilon_2(E)$ with decreasing E toward E_g at a rate much faster than the corresponding linear decrease to $E = 0$ in the Lorentz oscillator model. In fact, amorphous solids exhibit absorption tails, so-called "Urbach tails", whereby $\alpha(E)$ decreases exponentially with decreasing E below E_g. Although the Urbach tails fall off much more rapidly than the Lorentz oscillator tail, such tails in amorphous solids are far more extensive, in fact, than those observed in crystalline solids. A comprehensive optical model to describe amorphous semiconductors will be presented in Section 2.5.3.

The dielectric functions in Fig. 2.10 were obtained by real time SE from thin films of aluminum and silver in the initial stages of growth on thermal oxide-covered silicon wafer substrates when the films consist of an array of discontinuous particles (5.0 nm and 7.8 nm thick, respectively).[47,48] Such results are of interest here because they demonstrate that

electronic processes other than band-to-band transitions can also give rise to Lorentz oscillator characteristics. In Fig. 2.10, the resonances can be understood physically in terms of the forced collective oscillation of conduction electrons confined within each metal particle, whereby the restoring force is provided by the rigid array of positive nuclei (see Section 2.3.2).

A mathematical expression can be derived for the effective dielectric function of the discontinuous metal film, starting from the generalized Maxwell-Garnett effective medium theory, given by

$$\varepsilon = 1 + \frac{Q(\varepsilon_p - 1)}{1 + F(\varepsilon_p - 1)}, \tag{2.76a}$$

where ε_p is the dielectric function associated with the bulk metal from which the particles are formed, Q is the volume fraction associated with the particles, and F is an interparticle screening factor.[71–73] In this expression, the host in which the particles are embedded is the ambient medium with $\varepsilon_h = 1$. If the Drude free electron expressions of Eqs. (2.35) are substituted into Eq. (2.76a) for ε_p, then a Lorentz oscillator expression results:

$$\varepsilon(E) = 1 + \frac{A^2}{(E_0^2 - E^2) - i\Gamma E}, \tag{2.76b}$$

with amplitude factor $A = \sqrt{Q}\hbar\omega_p$, resonance energy $E_0 = \sqrt{F}\hbar\omega_p$, and broadening $\Gamma = \hbar\gamma_f$. In these expressions ω_p and γ_f are the plasma frequency and damping parameter of the bulk metal from Eqs. (2.35). Equation (2.76b) is more general than the oscillator result from Eq. (2.39) in that no specific particle shape is assumed; in addition, Eq. (2.76b) is valid for a collection of interacting particles. In fact, Eqs. (2.39) and (2.76b) yield the same result when the isotropic (spherical particle) formula for F is used,[71] namely $F = (1 - Q)/3$, in the limit $Q \ll 1$. Figure 2.11 shows an example fit to experimental data in ε acquired by real time SE during Ag film nucleation on a crystalline Si substrate in a sputter-deposition process.[48] Also shown is the ε_2 spectrum for a much thicker, fully coalesced silver film that emphasizes the absence of the resonance feature in this case. Deviations of the data from the fit in Fig. 2.11 can be attributed to a particle size distribution that generates a distribution in the screening parameter F.

As an application of the Lorentz oscillator analysis in real time SE, Fig. 2.12 shows the evolution of the resonance energy E_0 and the broadening parameter Γ in Eq. (2.76b) as functions of particle film thickness during Ag growth.[48] At the lowest thickness of 1.5 nm, $E_0 = 2.8$ eV and

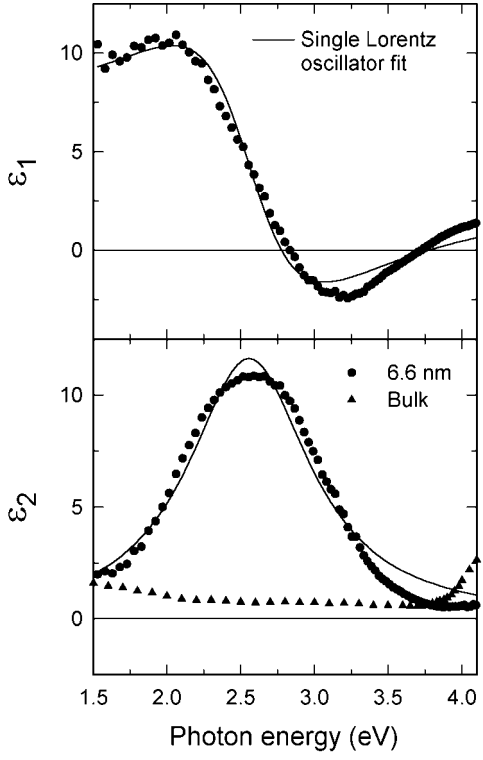


Figure 2.11 Real and imaginary parts of the dielectric function (ε_1, ε_2) of a thin film consisting of isolated silver particles on an oxide-covered silicon wafer substrate at room temperature. The results were measured by SE in real time during a magnetron sputtering process. The film thickness at the measurement time was 6.6 nm as determined in a three-parameter fitting procedure that includes the reflectance spectrum along with the SE data. The solid line is a best fit to the Lorentz oscillator equation [see Eqs. (2.76)] using parameters depicted in Fig. 2.12. Also shown for comparison is the imaginary part of the dielectric function for bulk Ag (triangles). (After Ref. 48.)

$\Gamma = 0.96$ eV. As d increases above 5 nm, E_0 decreases throughout the visible range to $E_0 = 1.6$ eV at 13 nm. These best fit values for E_0 in Fig. 2.12, along with the free electron plasma energy of $E_p = 8.8$ eV for bulk Ag, lead to screening parameter values $F = (E_0/E_p)^2$ that range from 0.1 at 1.5 nm thickness to 0.04 at 13 nm. This range of values is lower than would be expected based on the isotropic version of the Maxwell-Garnett effective medium theory with $F = (1 - Q)/3$. As a result, the particle film may in fact be optically anisotropic with different principal dielectric functions $\varepsilon_{\|}$

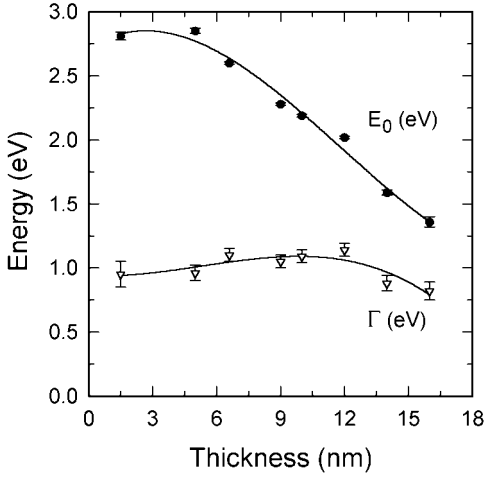
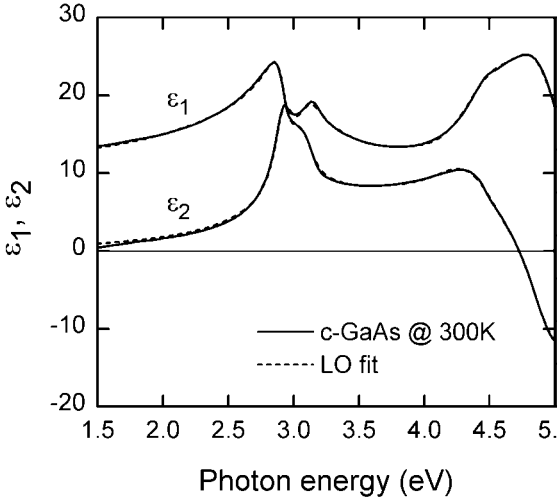


Figure 2.12 Resonance energy E_0 and broadening parameter Γ in the Lorentz oscillator equation [Eq. (2.76b)] versus film thickness for the Ag particle film of Fig. 2.11, as deduced from the evolution of the dielectric function during growth. The dielectric function was extracted in an analysis of spectra in the ellipsometric angles (ψ, Δ) and polarized reflectance, collected in real time during Ag film growth. (After Ref. 48.)

and $\varepsilon_\perp$ characterized by different values of $F_\parallel$ and $F_\perp$, for electric field directions parallel and perpendicular to the plane of the film, respectively. However, for these films, the SE spectra are dominated by the characteristics of $\varepsilon_\parallel$, and are not significantly influenced by $\varepsilon_\perp$.[74,75] Also in Fig. 2.12, a near constant broadening parameter of $\Gamma \sim 1$ eV is observed, independent of thickness (and thus particle size), suggesting a very short relaxation time of $\tau = \gamma_f^{-1} = 1.4 \times 10^{-15}$ s compared with the bulk value for silver of 4×10^{-14} s. This result could be attributed to a nearly constant electron mean free path of $L \sim 1$ nm independent of particle size. One can thereby conclude that defects within the particles, not at particle surfaces, limit the free electron relaxation time for the particulate films.[47,48]

In contrast to the results for a-Si:H and metal particle films, the dielectric function of bulk crystalline gallium arsenide (GaAs) in Fig. 2.13 exhibits characteristics that can be associated with a collection of resonances.[49] In GaAs, the bands of resonances are narrower, however, and tend to be confined to discrete frequency regions typical of a crystalline solid. For this material, an excellent simulation of the data above the lowest direct gap of ~1.5 eV can be obtained with seven Lorentz oscillators (broken line in Fig. 2.13).[33–35] Without a quantum mechanical description of the



Lorentz Osc. (eV)	Critical Point Energies (eV)
2.925	E_1=2.916
3.121	$E_1+\Delta_1$=3.154
3.294	–
3.857	–
4.514	E_0'=4.456
4.870	E_2=4.971
5.300	–

Figure 2.13 Real and imaginary parts of the dielectric function (ε_1, ε_2) of single crystal GaAs at room temperature as measured by spectroscopic ellipsometry (solid line). The broken line, nearly coinciding with the data on this scale, is a fit using seven Lorentz oscillators. The table at the right shows the best-fit resonance energies in eV, as well as the critical point energies obtained by fitting the second derivative of the dielectric function (see Fig. 2.33). Lorentz oscillator energies that do not correspond to critical points serve to simulate the broad background in ε in regions between critical points. (Data from Ref. 49.)

electronic states in the solid, however, the positions of these resonances cannot be understood. In fact, the resonances correspond to collections of interband transitions, i.e., the excitation of electrons from states in the filled valence band across an energy gap $\hbar\omega_n$ to states in the empty conduction band. As discussed in detail in Section 2.4, the magnitude of the photon momentum $\hbar q$ is very small compared to the range of electron momenta in the crystal, $0 \leq \hbar k \leq (2\pi/a_c)\hbar$ (a_c: unit cell length), and because only an electron is excited by the photon in a direct transition, the transition does not lead to a significant change in the electron wavevector $\vec{k}$. Such transitions are direct or vertical and are strongest in regions of $\vec{k}$-space where the energy surfaces $E_v(\vec{k})$ and $E_c(\vec{k})$ defining the valence and conduction bands are parallel over an appreciable range of $\vec{k}$. Indirect transitions can also occur between bands, involving the emission or absorption of quanta of lattice vibrational energy in conjunction with the excitation of electrons by photons, but their probability is much lower (refer to Section 2.4).

Figure 2.14 provides an example of the application of the multiple oscillator simulation of crystalline semiconductors as applied in a real time SE study of GaAs sputter-etching by Ar ions.[50] In the real time SE analysis, the overlayer thickness and composition, and the near-surface GaAs temperature are all extracted as a function of time, as the potential that accelerates the Ar ions in the beam is stepped from 50 to 500 V. At low accelerating potentials (< 200 V), prior to the complete removal of the native oxide, the overlayer is modeled as a mixture of oxide and amorphous GaAs; at high potentials (≥ 200 V), the overlayer is modeled as a mixture of a-GaAs and voids. The GaAs temperature can be incorporated as a free parameter in least-squares regression analysis by constructing a database as follows. First, the dielectric functions of GaAs measured at different temperatures are each fit to a sum of N_o Lorentz oscillators ($N_o = 7$ in Fig. 2.13). Then, the $3N_o + 1$ oscillator parameters (3 parameters per oscillator plus the constant contribution to ε_1) obtained from such fits are described as p^{th}-order polynomial functions of temperature. As a result, if the temperature is specified, the oscillator parameters corresponding to that temperature can be determined from the $(p + 1)(3N_o + 1)$ polynomial coefficients, and then the dielectric function can be determined from the oscillator parameters. For the least-squares regression analysis of Fig. 2.14, a table of 44 polynomial coefficients (with $p = 1$ and $N_o = 7$) is sufficient for this purpose.

As noted in Section 2.1, when the dielectric function is fitted with such a large number of free parameters (22 for the analysis of Figs. 2.13 and 2.14), care must be taken to ensure that the oscillator model is justified from a statistical standpoint. As a result, confidence limits on the oscillator parameters should be determined and also included in the evaluation of the polynomial coefficients. Obviously, the confidence limits on the polynomial coefficients improve when the dielectric function database includes a large number of spectra (versus temperature in this case), and this becomes possible when the database is established by the techniques of real time SE. With narrower confidence limits on the polynomial coefficients, any interpolations and extrapolations employing the database are more likely to be valid (see Section 2.1).

Finally, it should be noted that the Lorentz oscillator describes physical situations other than classical electronic vibrations. For example, the vibrations of ions in ionic crystals and the orientations of permanent dipoles generate optical property characteristics similar to those of Figs. 2.6 through 2.8 in the infrared and microwave ranges of the electromagnetic spectrum.

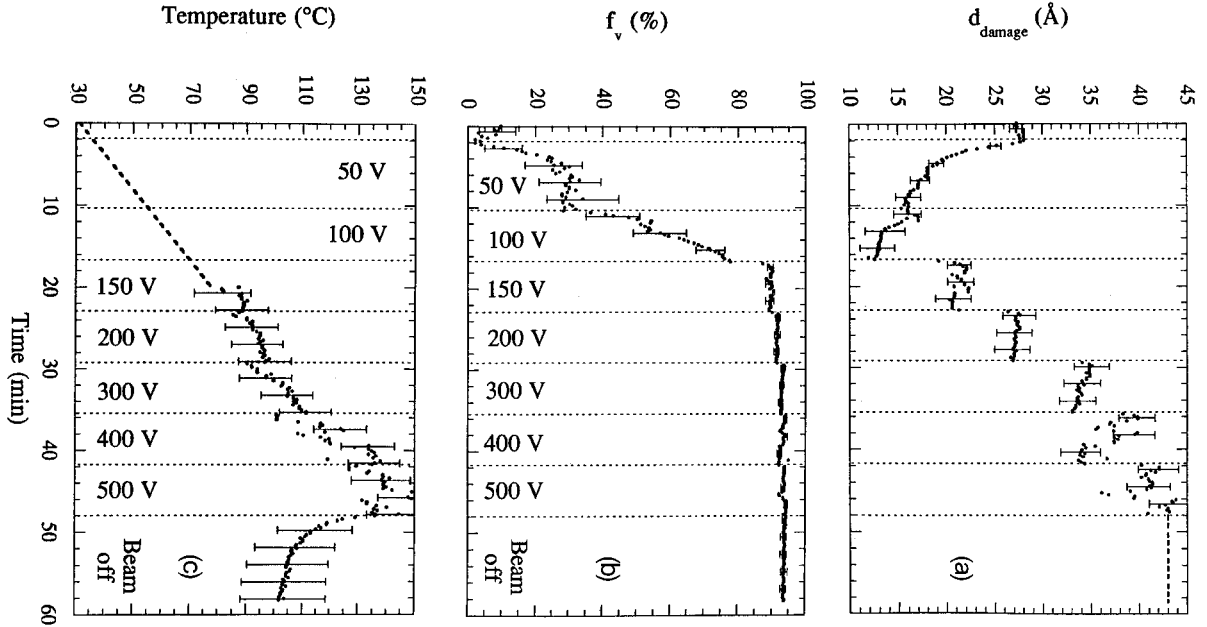


Figure 2.14 (a) Thickness of the surface damage layer (or pure oxide layer for $t < 2$ min), (b) composition of the damage layer, expressed as the volume fraction of amorphous GaAs in oxide (or in void for $t > 23$ min), and (c) near-surface temperature, all obtained from a least-squares regression analysis of real time spectroscopic ellipsometry data collected during ion beam etching of a GaAs substrate surface. At $t = 2$ min the ion beam was ignited, and at each of the vertical lines the accelerating voltage of the ion beam was increased as indicated. The horizontal broken lines indicate fixed values in the analysis, either a constant value for the damage layer when $t > 48$ min or linearly increasing values for the temperature when $2 < t < 20$ min. The temperature was fixed over this range to avoid correlations when the composition of the damage layer changes rapidly. (After Ref. 50.)

2.5.2 Classical Drude Models

2.5.2.1 Optical Characteristics: Metals and Free Electron Systems

Figures 2.15 through 2.17 show the complex dielectric function and index of refraction, the absorption coefficient, and the normal-incidence reflectance calculated for two different hypothetical metallic materials using Eqs. (2.12b), (2.13d), (2.13e), and (2.35). For both materials, the same

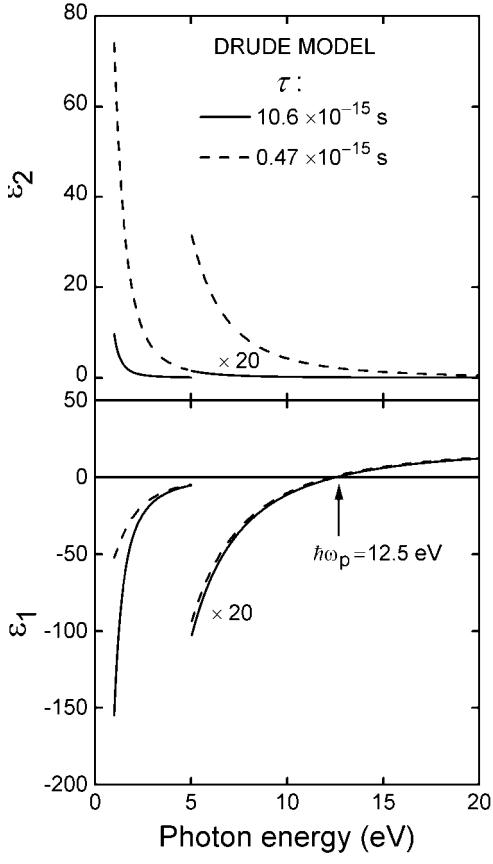


Figure 2.15 Photon energy dependence of the real and imaginary parts of the dielectric functions (ε_1, ε_2) for hypothetical free electron metals, according to the Drude model. In this case, the plasma energy is fixed at $\hbar\omega_p = 12.5$ eV, and two values of the free electron relaxation time τ are assumed, $\tau = 10.6 \times 10^{-15}$ s (solid line) and $\tau = 0.47 \times 10^{-15}$ s (broken line). Interband transitions are assumed to be absent.

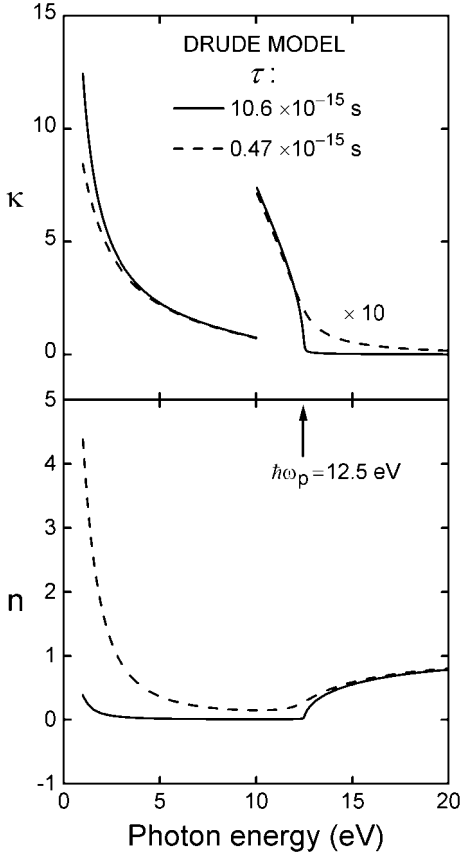


Figure 2.16 Photon energy dependence of the real (n) and imaginary (κ) parts of the complex index of refraction (N), deduced from the two dielectric functions for the free electron metals given in Fig. 2.15.

plasma energy of $E_p = 12.5$ eV is incorporated into the formulas, but two different values of γ_f are assumed. The E_p value and the longer value of $\gamma_f^{-1} = \tau = \tau_B = 10.6 \times 10^{-15}$ s is characteristic of the free electron component of fully coalesced (B: bulk) aluminum films. The much shorter value of $\tau = 0.47 \times 10^{-15}$ s is extracted from the formula $\tau^{-1} = \tau_B{}^{-1} + (v_F/L)$ using $L = 1$ nm as the mean free path that the electron travels before scattering, as has been deduced for aluminum particle films.[47]. In the formula for τ, v_F is set to 2.0×10^8 cm/s, the velocity of electrons at the Fermi level in aluminum.

The frequency dependences associated with the Drude model for the optical characteristics with $\omega > 0$ in Figs. 2.15 through 2.17, are similar to those of the single Lorentz oscillator model for the optical characteristics

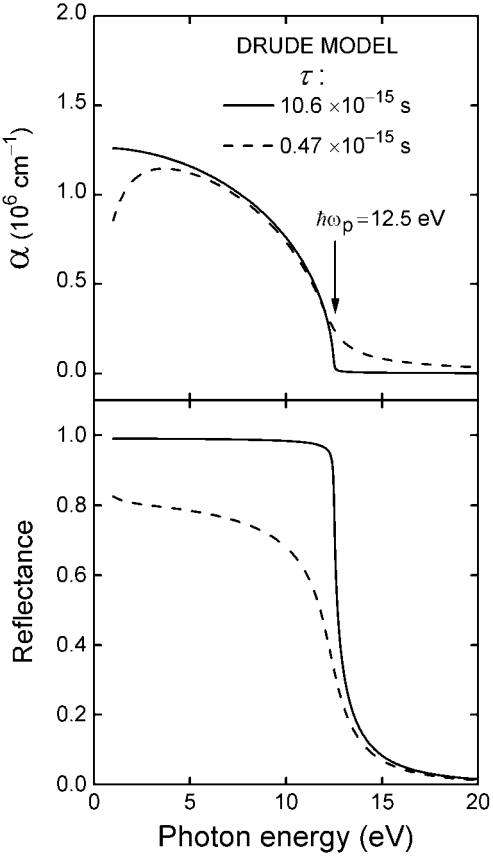


Figure 2.17 Photon energy dependence of the absorption coefficient (α, top) and normal incidence reflectance (bottom), for the free electron metals whose dielectric functions are given in Fig. 2.15.

for $\omega > \omega_0$ in Figs. 2.6 through 2.8. This similarity arises because the Drude formulas for the optical characteristics can be obtained directly from the single Lorentz oscillator formulas by setting $E_0 = \hbar\omega_0 = 0$. Here, the emphasis is on the marked changes in the optical characteristics in Figs. 2.16 and 2.17 at the plasma energy $E_p = \hbar\omega_p$ given by $\varepsilon_1(E_p) = 0$, where $E_p^2 = N_{ef} e^2\hbar^2/\varepsilon_0 m$ when $\omega\tau \gg 1$. For the larger value of τ, the real part of the index of refraction increases abruptly from near zero toward unity, the extinction and absorption coefficients drop abruptly to near zero values, and the reflectance decreases abruptly from near unity toward zero as E passes through E_p. At the smaller τ value, these marked changes occur more gradually, and the reflectance below the plasma frequency is lower due to the higher dissipation rate.

At the plasma frequency, the wavelength associated with the electric field in the metal tends to infinity, meaning that the entire electron population oscillates in concert, driven by the electric field. Such oscillations do not result in charge density fluctuations in the bulk of the material, and so are not to be confused with plasmons (see Section 2.3.3).[20] Although plasmons cannot be excited by incident transverse electromagnetic fields with $\vec{\nabla} \cdot \vec{E} = 0$, they can be excited by the longitudinal fields associated with incident charged particles. In optically isotropic materials in which the longitudinal and transverse dielectric functions are equal, plasmons can be excited at the plasma frequency given by, $\omega_p{}^2 = N_e e^2/\varepsilon_0 m$ (assuming $\omega\tau \gg 1$), as measured by the transverse optical probe, and they persist even after the longitudinal field is removed. A more detailed discussion of plasmons was presented in Section 2.3.3.

2.5.2.2 Examples: Metals and Free Electron Systems

Experimental results for the dielectric function and reflectance of fully coalesced silver films are shown in Figs. 2.18 and 2.19, respectively, and provide an instructive example of the complexity inherent in the optical properties of real metals.[20,36] The feature of greatest interest in Fig. 2.19 is the rapid decrease in the reflectance with increasing photon energy just below 4 eV where ε_1 in Fig. 2.18 vanishes. This feature identifies the plasma energy ($\hbar\omega_p$) as 3.9 eV. However, this value is much smaller than 9.2 eV, which is calculated on the basis of the free electron model. The abrupt increase in reflectance above 3.9 eV, along with the features in ε_1 and ε_2 in this region, reveal contributions to the optical properties of silver, not only from the free (i.e., intraband) electron transitions, but also from a collection of bound (i.e., interband) electron transitions. This can be seen clearly by the broken lines in Fig. 2.18 that represent a decomposition of the dielectric function into contributions from non-interacting free (f) and bound (b) electrons, according to $\varepsilon = \varepsilon_f + \varepsilon_b$.

In the dielectric function decomposition of Fig. 2.18, Eqs. (2.35) are used for the free electron contribution ε_f with parameters obtained by fitting the data for ε in Fig. 2.18 below 3.75 eV.[36] In the fit, however, a static dielectric constant ε_{0s} is included in the expression for ε_{1f} [see Eq. (2.35a)], rather than unity. As noted in Section 2.3.1, this accounts for the higher energy bound electron transitions that occur above the fitted spectral range. The best fit value of $\varepsilon_{0s} - 1$ is then included as part of ε_{1b} in the decomposition of Fig. 2.18. Resorting to a quantum mechanical interpretation now, the bound electron resonances near 4 eV and above arise from direct interband transitions in which electrons are excited from filled d-bands to empty states in the partially filled conduction band (see inset,

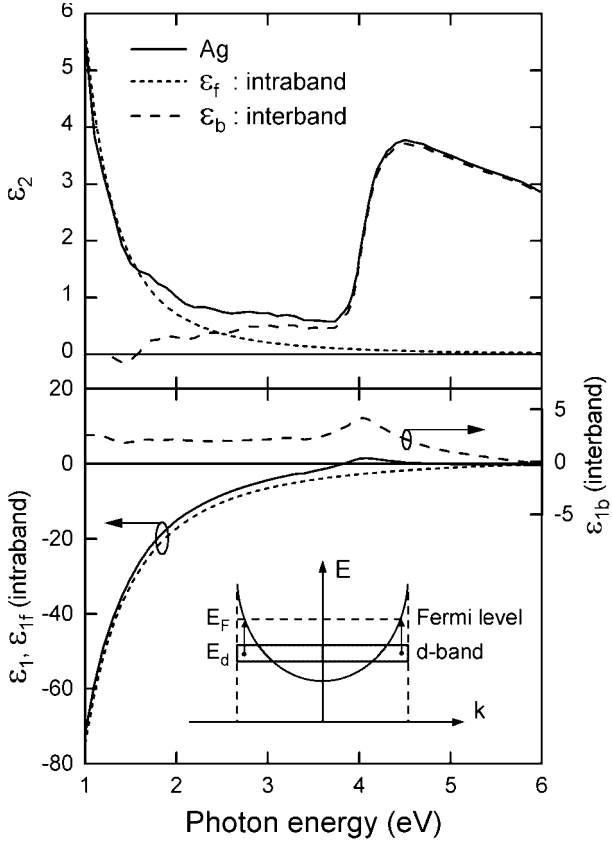


Figure 2.18 Real and imaginary parts of the dielectric function of silver (solid line). The short- and long-dashed lines represent the approximate intraband and interband components of the dielectric function, respectively. The appropriate intraband parameters are obtained in a best fit of ε for $\hbar\omega < 3.75$ eV, assuming it follows the Drude relationship of Eqs. (2.35), but with a photon energy-independent interband contribution ($\varepsilon_{0s} - 1$) added to the right-hand side in Eq. (2.35a). When plotting the two components, however, this contribution is included in the interband component. The inset shows the direct interband transitions on a simplified electron band structure diagram. (Data from Ref. 13.)

Fig. 2.18). In contrast, the free electron contribution can be attributed to the intraband transitions in which electrons are excited from below the Fermi level to above, within the same, partially filled conduction band. On the basis of the decomposition in Fig. 2.18, it is clear that the contribution to ε_1 from the bound electrons leads to a shift in the energy at which $\varepsilon_1 = 0$ (i.e., the plasma energy) to much lower values. This can be

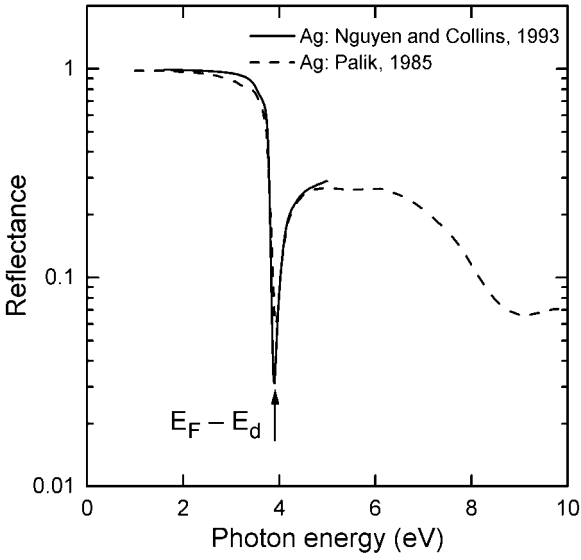


Figure 2.19 Normal incidence reflectance of silver plotted versus photon energy, including handbook data (broken line; from Ref. 13) and results deduced from SE measurements on a sample with a higher quality surface (solid line; from H.V. Nguyen and R.W. Collins, unpublished data). Both results reveal a sharp minimum at the (screened) plasma frequency ($\hbar\omega_p = 3.9$ eV) as well as the onset of interband reflection associated with (d-band)-to-(Fermi level) transitions (see inset of Fig. 2.18).

understood physically by a screening of the free electrons by the bound electrons in the d-band.[20]

A similar explanation may also apply to the optical behavior for the aluminum and silver particle films of Fig. 2.10. For these films, the observed resonance energies in the dielectric function, $E_0 = \hbar\omega_0 = 2.9$ and 2.4 eV, are well below the values of $\hbar\omega_{pp} = \hbar\omega_p/\sqrt{3} = 7.2$ and 5.3 eV, calculated for widely separated ($Q \ll 1$) spherical ($F = 1/3$) particles [see Eq. (2.39) in Section 2.3.2 and Eq. (2.76a) in Section 2.5.1.2]. In fact, these lower resonance energies may result from screening of the fields within the particles by the response of neighboring particles (which exist at high density on the substrate surface). This interpretation is supported by the observation that $E_0 = \hbar\omega_0$ decreases with increasing volume fraction Q in the discontinuous film (see Fig. 2.12). In Fig. 2.10 (right) for the aluminum particulate film, the bound electrons associated with the interband transitions near 1.5 eV and above (see Section 2.5.4.2) also play a role by providing a pathway for energy dissipation from the free electron oscillations in the particles. This decreases τ in Eq. (2.39) [and also increases Γ in Eq. (2.76b)] and, as a result, increases the width of the resonance compared to that of the Ag particulate films.

2.5.3 Generalized Quantum Mechanical Models

As is evident from the discussion of the previous section, it is not possible to understand the optical properties of solids solely on the basis of classical concepts. In each specific example above, quantum mechanical concepts were invoked in order to account for the detailed characteristics. In many cases, such concepts can be introduced in general terms without evaluation of the specific band structures of the solids and the matrix elements of the associated electronic transitions. Most often the purpose of such generalized models is to simulate the measured dielectric functions in terms of a number of physically meaningful, wavelength-independent parameters.[29,30] The simulations are then used to model the dielectric functions obtained for a system of bulk solids or thin films having different material characteristics, for example different compositions, temperatures, doping levels, polycrystalline grain sizes, etc. Then the best fit wavelength-independent parameters are extracted versus one or more material characteristics, thus providing a database for future analysis of such materials by SE. An example of this modeling approach was given earlier (see Fig. 2.14), in which case the near-surface temperature of GaAs was determined from real time SE measurements performed during ion etching.

2.5.3.1 Optical Characteristics: Amorphous and Nano/Microcrystalline Materials

As has been described in Sections 2.3 and 2.4, the imaginary part of the dielectric function $\varepsilon_2(E)$ of a material can be divided into two components, one generated by excitation of intraband electronic transitions that occur in metals and degenerate semiconductors [designated $\varepsilon_{2f}(E)$] and the other generated by excitation of interband electronic transitions [designated $\varepsilon_{2b}(E)$], such that $\varepsilon_2(E) = \varepsilon_{2f}(E) + \varepsilon_{2b}(E)$. In fact, $\varepsilon_{2f}(E)$ can be expressed analytically by applying the formula of Eq. (2.75a), and $\varepsilon_{2b}(E)$ can often be written as a sum of generalized resonance features.[76] In addition, if a suitable expression for $\varepsilon_2(E)$ is available spanning the full energy range, then $\varepsilon_1(E)$ can be written as a Kramers-Kronig transformation of $\varepsilon_2(E)$ given by Eqs. (2.29). These considerations lead to two generalized expressions of the form:

$$\varepsilon_2(E) = \frac{E_p^2 \Gamma_f}{E(E^2 + \Gamma_f^2)} + G(E) \sum_{j=1}^{N_Q} L_j(E), \tag{2.77a}$$

$$\varepsilon_1(E) = \varepsilon_{0s} - \frac{E_p^2}{E^2 + \Gamma_f^2} + \frac{2}{\pi} \sum_{j=1}^{N_Q} P \int_0^\infty \frac{E' G(E') L_j(E')}{E'^2 - E^2} dE'. \tag{2.77b}$$

In these expressions, (E_p, Γ_f) are the plasma and broadening energies that describe the intraband electron excitations, given from Eq. (2.75a) by $E_p = (e^2\hbar^2 N_{ec}/\varepsilon_0 m_o)^{1/2}$ and $\Gamma_f = 2\Gamma_m$. In addition in Eqs. (2.77), $L_j(E)$ is a lineshape function corresponding to the j^{th} interband transition, and $G(E)$ modifies the shape of the imaginary part of the dielectric function near the absorption onset and below it as well. This function is defined such that $G(E) \to 1$ for $E \gg E_g$ and $G(E) \propto \varepsilon_2(E)$ for $E \approx E_g$, where E_g designates the lowest optical band gap. In fact, $G(E)$ is incorporated to simulate quantum mechanical effects due to the electronic joint densities of states and the matrix element of the transitions. Finally, $\varepsilon_{0s} \geq 1$ is the constant contribution to the real part of the dielectric function that incorporates transitions above the upper energy limit of measurement and not included in the sum of Eq. (2.77a).

Next, the application of specific forms in Eqs. (2.77) relevant to amorphous and nano/microcrystalline thin film materials will be described. For these materials, the Lorentz oscillator lineshape function provides a suitable description of the optical features in ε_2 above the bandgap owing to the extensive broadening that occurs in these materials. This function can be expressed in terms of photon energy as:

$$L_j(E) = \frac{A_{0j} E_{0j} \Gamma_j E}{(E_{0j}^2 - E^2)^2 + \Gamma_j^2 E^2}, \tag{2.78}$$

where $(A_{0j}, E_{0j}, \Gamma_j)$ describe the oscillator strength, resonance energy, and broadening energy for the j^{th} transition. In this equation, in order to be consistent with an analogous development in the literature, A_{0j} is defined according to $A_{0j} \equiv (A_j^2/E_{0j})$, where $A_j^2 \equiv (e^2\hbar^2 N_{ej}/\varepsilon_0 m)$ for a classical atomic oscillator.[51] For amorphous semiconductors, a single oscillator function modified by $G(E)$ is suitable in Eq. (2.77a) so that $N_o = 1$. For nano/microcrystalline semiconductors, multiple oscillators are required to simulate the broadened critical point features as necessary.[52,76] For semiconductors in the absence of degenerate doping, the free electron terms in Eqs. (2.77) can be neglected.

In order to develop appropriate functional forms for $G(E)$, it is of interest to determine possible shapes of $\varepsilon_2(E)$ [or $\alpha(E)$] in the photon energy region near the energy gap where the Lorentz oscillator function inevitably fails to provide an adequate fit. Just above the gap in the band-to-band absorption region, Eqs. (2.59) and (2.60) provide two possible forms for $\varepsilon_2(E)$ based on the assumptions of parabolic densities of valence and conduction band states and either constant dipole or constant momentum matrix elements versus photon energy, respectively.[67] As an example, Fig. 2.20 shows plots that compare the two different expressions for the absorption onset of a-Si:H using as data the dielectric function given

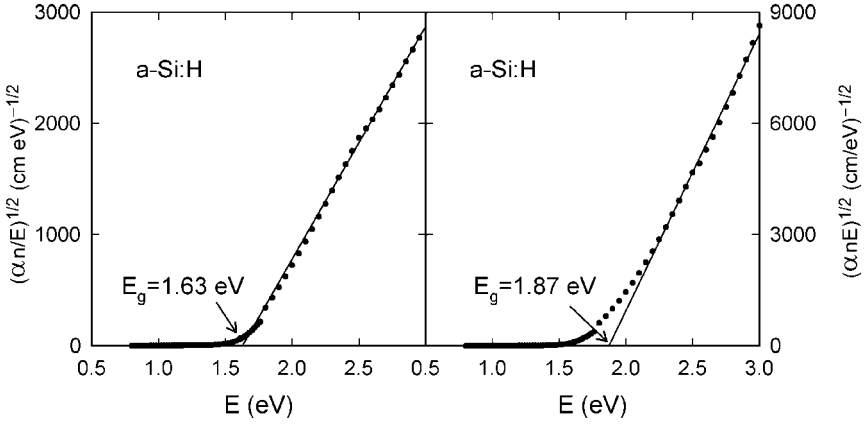


Figure 2.20 Optical spectra for hydrogenated amorphous silicon from Fig. 2.9, plotted as $(\alpha n/\hbar\omega)^{1/2}$ vs. photon energy (left) and as $(\alpha n\hbar\omega)^{1/2}$ vs. photon energy $E = \hbar\omega$ (right). For parabolic band densities of states and constant dipole (left) or momentum (right) matrix elements, the plot should be linear and extrapolate to the band gap at $\alpha = 0$. (Data from Ref. 46.)

previously in Fig. 2.9.[46] These data were obtained ex situ at room temperature by supplementing SE measurements at high photon energies with transmittance and reflectance spectroscopy at intermediate energies and with photoconductive spectroscopy at low energies. It is clear that the expression of Eq. (2.59), equivalent to $\varepsilon_2(E) \propto (E - E_g)^2$ and derived assuming a constant dipole matrix element, provides the widest range of linearity in the experimental data (see Fig. 2.20, left panel). In contrast, the Tauc expression of Eq. (2.60), equivalent to $\varepsilon_2(E) \propto (E - E_g)^2/E^2$ and derived assuming a constant momentum matrix element, exhibits continuous curvature with only narrow ranges of linearity (see Fig. 2.20, right panel). As a result, Eq. (2.60) does not provide a unique value of the gap; this so-called Tauc gap depends on the energy range of the data used for extrapolation. Weak absorption is present below the energy gaps in Fig. 2.20 irrespective of the plotting method and can be described unambiguously in terms of an Urbach tail given by $\alpha(E) = \alpha_0 \exp[(E - E_t)/E_u]$ for $E < E_t$, where E_u is the Urbach absorption tail slope and $E_t > E_g$ is the transition energy below which the Urbach behavior is observed. Both band-to-band and Urbach tail characteristics can be combined into a simple piece-wise expression as described next.

In spite of the poor linearity of the Tauc gap plot in Fig. 2.20 (right), the associated band edge result $G(E) = (E - E_g)^2/E^2$ (for $E > E_g$) has been applied previously in deriving the so-called Tauc-Lorentz expression for $\varepsilon(E)$ of an amorphous semiconductor.[51] In fact, this form for $G(E)$ is

very convenient since it obeys the requirement that $G(E) \to 1$ for $E \gg E_g$, without further manipulation. Incorporation of an Urbach tail then leads to the piecewise expression

$$G(E) = G_T(E) = \begin{cases} (E_1/E)\exp[(E - E_t)/E_u], & E < E_t; \quad (a) \\ [(E - E_g)/E]^2, & E > E_t; \quad (b) \end{cases} \tag{2.79}$$

where E_1 is defined such that $\varepsilon_2(E)$ for the two functions is continuous at the transition energy $E = E_t$. To obtain the conventional Tauc-Lorentz expression without the Urbach tail, one can set $E_t = E_g$ and $E_1 = 0$ (or $E_u \to 0$) so that $G(E) = 0$ for $E \leq E_g$.[51]

An alternative approach to Eqs. (2.79) is based on the constant dipole matrix element result, which requires further manipulation to obtain $G(E)$, in particular, the addition of another free parameter E_P to ensure that $G(E) \to 1$ for $E \gg E_g$:

$$G(E) = G_C(E) = \begin{cases} (E_1/E)\exp[(E - E_t)/E_u], & E < E_t; \quad (a) \\ (E - E_g)^2/[E_P^2 + (E - E_g)^2], & E > E_t. \quad (b) \end{cases} \tag{2.80}$$

Here E_P (not to be confused with the plasma energy) defines a second transition energy (in addition to E_t) given by $E_P + E_g$. This transition energy separates the absorption onset behavior [$E < (E_P + E_g)$] from the Lorentz oscillator behavior [$E > (E_P + E_g)$].[52,76] Specifically, if $E \approx E_g$ [so that $(E - E_g) \ll E_P$], then $G(E) \to [(E - E_g)/E_P]^2$, whereas if $E \gg (E_P + E_g)$ [so that $(E - E_g) \gg E_P$] then $G(E) \to 1$, as required for the Kramers-Kronig integration to extract $\varepsilon_1(E)$. In fact, different functional forms for $G(E)$ can be applied to modify the oscillator function $L(E)$, depending on the desired shape of the absorption onset.

2.5.3.2 Examples: Amorphous and Nano/Microcrystalline Semiconductors

For the amorphous and nano/microcrystalline silicon-based semiconductors to be considered here, the expressions for $\varepsilon(E)$ in Eqs. (2.77) do not require the free-electron terms.[76] This leads to the following form for $\varepsilon_2(E)$:

$$\varepsilon_2(E) = \begin{cases} G(E), & E < E_t; \quad (a) \\ G(E)\sum_{j=1}^{N_O} L_j(E); & E > E_t. \quad (b) \end{cases} \tag{2.81}$$

With this expression given as two parts, the variation of the Lorentz oscillator sum over the energy range of the Urbach tail ($0 \leq E < E_t$) is now neglected—not only in $\varepsilon_2(E)$ as indicated in Eq. (2.81a), but also in the Kramers-Kronig integral for $\varepsilon_1(E)$ [see Eq. (2.82b) below]. Considering for the moment only one term in this sum ($N_o = 1$), the Kramers-Kronig integration of Eq. (2.77b) for $\varepsilon_1(E)$ can be split into three terms, ε_{0s}, $I_U(E)$, and $I_L(E)$, for the constant, the Urbach tail (sub-gap), and the modified oscillator (above-gap) contributions, respectively:

$$\varepsilon_1(E) = \varepsilon_{0s} + I_U(E) + I_L(E) \tag{2.82a}$$

$$= \varepsilon_{0s} + \frac{2E_1}{\pi} P \int_0^{E_t} \frac{\exp[(E' - E_t)/E_u]}{E'^2 - E^2} dE' + \frac{2}{\pi} P \int_{E_t}^{\infty} \frac{E' G(E') L(E')}{E'^2 - E^2} dE'. \tag{2.82b}$$

In Eq. (2.82b), E_1 is defined so that $\varepsilon_2(E)$ is continuous at E_t; thus

$$E_1 = E_t G(E_t) L(E_t). \tag{2.82c}$$

For the amorphous silicon-based semiconductors, only one oscillator is required, and so Eqs. (2.81) with $N_o = 1$ followed by Eq. (2.82) are directly applicable. In addition, for these materials, Eq. (2.80b) can be employed for $G(E')$ (for $E' > E_t$) in I_L. In contrast for nano/microcrystalline silicon films, two oscillators are required, one to fit the strongly broadened E_0'-E_1 critical point complex near 3.4 eV and the other to fit the E_2 critical point near 4.2 eV. Multiple oscillators can be incorporated easily by generalizing the approach of Eqs. (2.82). Thus, two integrals $I_{L1}(E)$ and $I_{L2}(E)$ are used in Eqs. (2.82), each having the same gap function $G(E')$, but different Lorentz oscillator functions $L_1(E')$ and $L_2(E')$ with parameters (A_{0j}, E_{0j}, Γ_j; $j = 1,2$). In this case the gap function is given for simplicity by Eq. (2.79b). Such a form is possible, not only for so-called "non-direct gap" amorphous semiconductors when the momentum matrix element is constant [see Eq. (2.60)], but also for an indirect gap semiconductor when $(E_{av} - E)^2$ is approximately constant [see Eqs. (2.72)]. However, it should be emphasized that absorption below the direct gap in an indirect gap semiconductor such as Si is too weak to be detected accurately by SE except in the case of Si nanocrystals for which $\vec{k}$-conservation is relaxed.[77]

Starting with the case of nano/microcrystalline silicon, substituting Eq. (2.79b) into the third term of Eq. (2.82b), and setting $I_{TL} = I_L$ [where 'T' designates the indirect gap expression for G of Eqs. (2.72) which is approximately equivalent, in fact, to the Tauc gap expression of

Eq. (2.60)] yields the following expressions:

$$I_{TL}(a_{iT};\, i = 0,1,2,3;c_{0T},d_{0T}) =$$

$$\frac{2A_0E_0\Gamma}{\pi}[a_{3T}\{\zeta^2 I_{1T} - \ln[L_D(E_t)]^{1/4}\} + a_{2T}(I_{0AT} + I_{0BT}) + a_{1T}I_{1T}$$

$$+ a_{0T}\frac{I_{0AT} - I_{0BT}}{E_0^2} - c_{0T}\ln|E - E_t| - d_{0T}\ln(E + E_t)], \tag{2.83a}$$

where

$$I_{1T} = \frac{1}{2\chi\Gamma}\left\{\pi - 2\arctan\left(2\frac{E_t^2 - \zeta^2}{\chi\Gamma}\right)\right\}, \tag{2.83b}$$

$$I_{0AT} = \frac{1}{2\Gamma}\left\{\pi - \arctan\left(\frac{2E_t + \chi}{\Gamma}\right) + \arctan\left(\frac{-2E_t + \chi}{\Gamma}\right)\right\}, \tag{2.83c}$$

$$I_{0BT} = \frac{1}{4\chi}\ln\frac{E_t^2 + E_0^2 + \chi E_t}{E_t^2 + E_0^2 - \chi E_t}, \tag{2.83d}$$

and

$$\zeta = \sqrt{E_0^2 - (\Gamma^2/2)}, \tag{2.83e}$$

$$\chi = \sqrt{4E_0^2 - \Gamma^2}, \tag{2.83f}$$

$$c_{0T} = \frac{EG_T(E)}{2L_D(E)}, \tag{2.83g}$$

$$d_{0T} = \frac{-(E + E_g)^2}{2EL_D(E)}, \tag{2.83h}$$

$$a_{3T} = -(c_{0T} + d_{0T}), \tag{2.83i}$$

$$a_{2T} = -E(c_{0T} - d_{0T}), \tag{2.83j}$$

$$a_{1T} = -(E^2 - 2\zeta^2)(c_{0T} + d_{0T}), \tag{2.83k}$$

$$a_{0T} = 1 - E(E^2 - 2\zeta^2)(c_{0T} - d_{0T}), \tag{2.83l}$$

$$G_T(E) = \frac{(E - E_g)^2}{E^2}, \tag{2.83m}$$

$$L_D(E) = (E^2 - E_0^2)^2 + \Gamma^2E^2. \tag{2.83n}$$

The subscript $j = 1$ on the oscillator parameters (A_{01}, E_{01}, Γ_1) is suppressed in Eqs. (2.83), since only one oscillator is being considered at a time. In addition, in Eqs. (2.83) the fully expanded form of the expression for $\varepsilon_1(E)$ is presented, rather than the simplified version given by Jellison and Modine.[51] With this expanded form, it is easier to adapt the expressions when $G(E)$ is modified. For this reason, a functional description of I_{TL} is provided in terms of the energy dependent coefficients, written as $I_{TL}(a_{iT}; i = 0, 1, 2, 3; c_{0T}, d_{0T})$ in Eq. (2.83a). Finally, it is important to emphasize that Eqs. (2.83) can be applied for the cases in which an Urbach tail is either absent [using $E_t = E_g$ and $E_1 = 0$, so $I_U(E) = 0$] or present [using $I_U(E)$ from Eq. (2.85) below and $E_t > E_g$].[76]

Proceeding with the case of amorphous silicon-based materials, substituting Eq. (2.80b) into the third term of Eq. (2.82b) and setting $I_{CL} = I_L$ yields,

$$I_{CL}(a_{iC}; i = 0,1,2,3; b_{0C}, b_{1C}, c_{0C}, d_{0C}) = I_{TL}(a_{iC}; i = 0,1,2,3; c_{0C}, d_{0C}) + \frac{2A_0E_0\Gamma}{\pi}[b_{1C}\{E_gI_{0C} - \ln[(E_t - E_g)^2 + E_P^2]^{1/2}\} + b_{0C}I_{0C}], \quad (2.84a)$$

where

$$I_{0C} = \frac{1}{E_P}\left\{\frac{\pi}{2} - \arctan\frac{E_t - E_g}{E_P}\right\}, \quad (2.84b)$$

and

$$c_{0C} = \frac{EG_C(E)}{2L_D(E)}, \quad (2.84c)$$

$$d_{0C} = \frac{-E(E + E_g)^2}{2L_D(E)[(E + E_g)^2 + E_P^2]}, \quad (2.84d)$$

$$b_{0C} = \frac{Y^4F^2\{L_D(E)[E^{-1}(c_{0C} - d_{0C}) + 2E_gK^2Y^{-4}(c_{0C} + d_{0C})] - 1\}}{(K^2 - F^2)F^2Y^4 + E_0^4Y^4 + 4E_g^2F^2K^4}, \quad (2.84e)$$

$$b_{1C} = \frac{2E_gK^2b_{0C} - L_D(E)(c_{0C} + d_{0C})}{Y^4}, \quad (2.84f)$$

$$a_{3C} = -(b_{1C} + c_{0C} + d_{0C}), \quad (2.84g)$$

$$a_{2C} = -[b_{0C} + 2E_gb_{1C} + E(c_{0C} - d_{0C})], \quad (2.84h)$$

$$a_{1C} = -[2E_gb_{0C} - (K^2 - F^2)b_{1C} + (E^2 - 2\zeta^2)(c_{0C} + d_{0C})], \quad (2.84i)$$

$$a_{0C} = 1 + (K^2 - F^2)b_{0C} + 2E_gK^2b_{1C} - E(E^2 - 2\zeta^2)(c_{0C} - d_{0C}), \quad (2.84j)$$

$$F^2 = E_P^2 + E_g^2, \tag{2.84k}$$

$$K^2 = 2F^2 + 2\zeta^2 - 4E_g^2, \tag{2.84l}$$

$$Y^4 = E_0^4 + F^2(K^2 - F^2) - 4E_g^2K^2, \tag{2.84m}$$

$$G_C(E) = \frac{(E - E_g)^2}{[E_P^2 + (E - E_g)^2]}. \tag{2.84n}$$

Again, these expressions can be applied for the cases in which an Urbach tail is either absent [using $E_t = E_g$ and $E_1 = 0$, so $I_U(E) = 0$] or present [using $I_U(E)$ from Eq. (2.85) below and $E_t > E_g$].[76]

In fact, when an Urbach tail is incorporated, the following expression can be derived for $I_U(E)$ in Eq. (2.82b):

$$I_U(E) = \frac{E_1}{\pi E}\left\{\exp\left(\frac{E - E_t}{E_u}\right)\left[\mathrm{Ei}\left(\frac{E_t - E}{E_u}\right) - \mathrm{Ei}\left(\frac{-E}{E_u}\right)\right] - \exp\left(\frac{-(E + E_t)}{E_u}\right)\left[\mathrm{Ei}\left(\frac{E_t + E}{E_u}\right) - \mathrm{Ei}\left(\frac{E}{E_u}\right)\right]\right\}, \tag{2.85}$$

where $\mathrm{Ei}(x)$ designates the exponential integral defined by

$$\mathrm{Ei}(x) = \int_{-\infty}^{x} \frac{\exp(t)}{t}\,dt.$$

Algorithms for evaluation of this integral can be found elsewhere.[78] The addition of the Urbach tail to the formulations of Eqs. (2.82) through (2.84) introduces difficulties in the evaluation of the two Kramers-Kronig integrals of Eq. (2.82b). $\mathrm{Ei}(x)$ diverges when $x \to 0$ leading to apparent divergences when $E \to 0$ or E_t. These divergences in the different terms cancel for the most part, and any residual features due to the discontinuity of the higher derivatives of ε_2 with respect to energy are so weak as to be inconsequential in fitting experimental data.

In summary, Eqs. (2.80) and (2.81) can be combined with $N_o = 1$ to obtain $\varepsilon_2(E)$ for a-Si:H and its alloys, and Eqs. (2.80), (2.82), (2.83a-f), (2.83n), (2.84), and (2.85) can be combined to obtain $\varepsilon_1(E)$ from $\varepsilon_2(E)$. The latter steps are quite tedious; however, once they are programmed in a least-squares regression analysis of SE data, fitting of SE data becomes quite routine. The model for the optical properties of a-Si:H and its alloys to be described in the following paragraphs involves determining eight free parameters in the most general case, two associated with the Urbach tail

(E_u, E_t), two associated with the absorption onset function (E_g, E_P), three associated with the single oscillator (A_0, E_0, Γ), and ε_{0s}. In such fitting, however, ε_{0s} can be fixed at unity, thereby reducing the number of free parameters to seven. In cases in which the Urbach tail is neglected or not included, the number of free parameters is then reduced to five. In addition, Eqs. (2.79) and (2.81) can be combined with $N_o = 2$ to obtain $\varepsilon_2(E)$ for nano/microcrystalline silicon, and Eqs. (2.79), (2.82), (2.83), and (2.85) can be combined to obtain $\varepsilon_1(E)$ from $\varepsilon_2(E)$. Such a model for nano/microcrystalline silicon requires at most ten free parameters, two associated with the Urbach tail (E_u, E_t), one associated with the absorption onset (E_g), six associated with the two Lorentz oscillators (A_{0j}, E_{0j}, Γ_j; $j = 1,2$), along with ε_{0s}. Omission of the Urbach tail, as is done in fitting SE data for nano/microcrystalline silicon, reduces the number of parameters to eight.

As examples of the modeling of a-Si:H-based films, Figs. 2.21 (a-c) show the optical properties of 500 to 1000 nm thick samples of a-Si:H, a-$Si_{1-x}Ge_x$:H, and a-$Si_{1-x}C_x$:H expressed as (ε_1, ε_2) and (n, $\log\alpha$) versus photon energy (points).[76] These results were obtained ex situ at room temperature by combining (i) SE measurements at high photon energies with (ii) transmittance and reflectance spectroscopy (R) at intermediate energies and finally (iii) dual-beam photoconductive spectroscopy (DBPC) at low energies. The DBPC technique provides only a relative measurement of α; as a result the normalization constant must be determined either by ensuring an energy overlap with the T&R data or by using the extrapolation method described below. The thin films were prepared on glass substrates using H_2-dilution to optimize material performance as the intrinsic layer in p-i-n (or n-i-p) solar cell devices. For the a-Si:H sample, a full data set is available over the range of photon energy from 1.5 to 4.8 eV, and a seven-parameter fit to the data is obtained; (ε_{0s} is fixed at unity). For the other two samples, the experimental data sets are incomplete. First, the a-$Si_{1-x}Ge_x$:H sample is thinner than the others; as a result, it was not possible to obtain a region of photon energy overlap between the DBPC and T&R measurements. To solve this problem, the DBPC normalization constant is found by comparing the DBPC spectra to an extrapolated result. In the extrapolation, E_u is determined directly from the slope of the DBPC data, and E_t is obtained from the expression $E_t = E_g + 2E_u$ (see discussion below). All other parameters in the extrapolation are determined in the best fit to the T&R and SE measurements alone. For the a-$Si_{1-x}C_x$:H, index of refraction measurements by T&R were unavailable. The fits to the incomplete data sets in both cases (solid lines) suggest that, because the optical model is consistent with the Kramers-Kronig relationships, such interpolations and extrapolations of the model will be valid in regimes where data are unavailable.

Generally, in order to classify the optical properties of a-Si:H-based alloys relevant for the design of a wide variety of devices, one quotes a nominal value of the optical gap that is easily accessible from transmission

and reflection (T&R) measurements. In Fig. 2.22, the seven parameters $\{(A_0, E_0, \Gamma), (E_g, E_P), (E_u, E_t)\}$ extracted in the best fits to data such as those of Fig. 2.21 are plotted as functions of the optical gap E_g(T&R) obtained according to the method assuming parabolic bands and a constant dipole

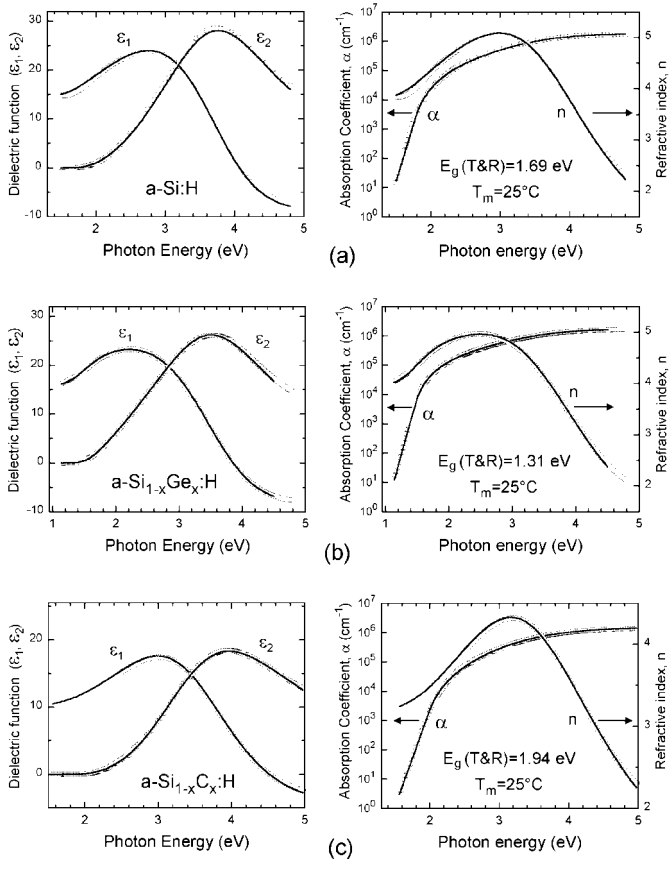


Figure 2.21 Optical properties of 0.5 to 1.0 μm thick samples of (a) a-Si:H, (b) a-$Si_{1-x}Ge_x$:H, and (c) a-$Si_{1-x}C_x$:H expressed as (ε_1, ε_2) (left) and (n, logα) (right) versus photon energy (points). These results were obtained ex situ at room temperature by combining dual beam photoconductivity, transmittance and reflectance (T&R) spectroscopy, and spectroscopic ellipsometry measurements. The samples are identified according to their band gaps E_g(T&R) as deduced from T&R measurements according to the procedure of Fig. 2.20 (left). The solid lines are fits to the data using a seven parameter optical model, including Urbach sub-gap, parabolic-band near-gap, and Lorentz above-gap modeling components. These results also demonstrate the ability to apply the model for the purposes of extrapolation or interpolation into photon energy regions where data are unavailable. (After Ref. 76.)

matrix element method [see Fig. 2.20 (left)].[76] In general, well-defined variations occur in Fig. 2.22 between the best fit parameters and E_g(T&R). Several of these variations provide physical insights into the nature of the materials. For example, the variations of E_u and Γ show that the highest ordering of the amorphous network is obtained for a-Si:H (minimum E_u

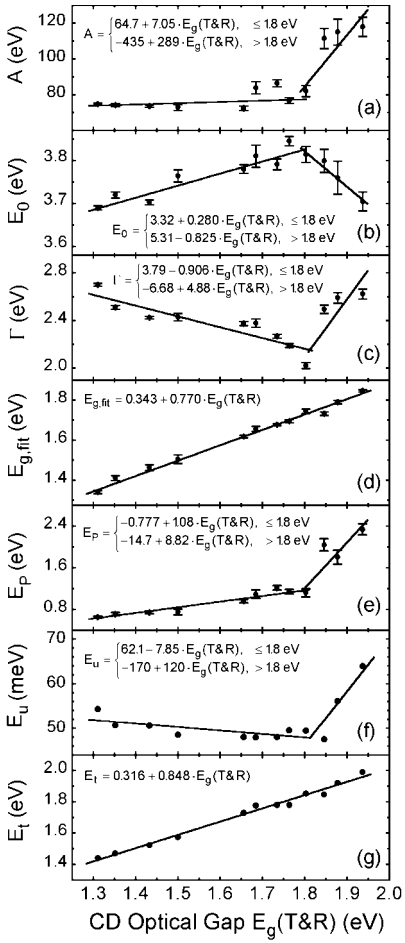


Figure 2.22 The seven free parameters {(A, E_0, Γ), (E_g, E_P), (E_u, E_t)} [(a–g), respectively], in the analytical expression for the dielectric function, obtained in the best fits to data such as those of Fig. 2.21, plotted versus the optical gap E_g(T&R), obtained from transmittance and reflectance spectra according to the method of Fig. 2.20 (left). Results from several photovoltaic-quality a-Si:H, a-$Si_{1-x}Ge_x$:H, and a-$Si_{1-x}C_x$:H films are correlated, suggesting the piecewise linear functions shown. Such functions provide the room temperature optical properties of photovoltaic-quality materials based on a specification of the optical gap as obtained by T&R. (After Ref. 76.)

and Γ); the ordering is reduced as Ge or C is added to make narrower or wider gap alloys. As expected, the best single-branch linear variations are obtained for E_g and E_t. First, it is expected that the gap values E_g obtained from the complete fits as in Fig. 2.21 should be in approximate agreement with the values E_g(T&R) obtained as in Fig. 2.20 (left). Second, one would expect that $E_t \approx E_g + 2E_u$ in order to maintain approximate continuity of the first derivative of $(\varepsilon_1, \varepsilon_2)$ at E_t, as is shown from a mathematical analysis of Eqs. (2.79) and (2.80).

Figure 2.22 is important because the expressions for the linear fits listed there can provide the optical properties (n, logα) for a device-quality amorphous semiconductor alloy of any arbitrary optical gap as determined from T&R data. The procedure is to (i) specify E_g(T&R), (ii) substitute this value into the appropriate linear relationship to deduce each of the seven parameters $\{(A_0, E_0, \Gamma), (E_g, E_P), (E_u, E_t)\}$, and (iii) substitute the seven parameters into the analytical formulas for the dielectric function, obtained by combining Eqs. (2.81) through (2.85). Figure 2.23

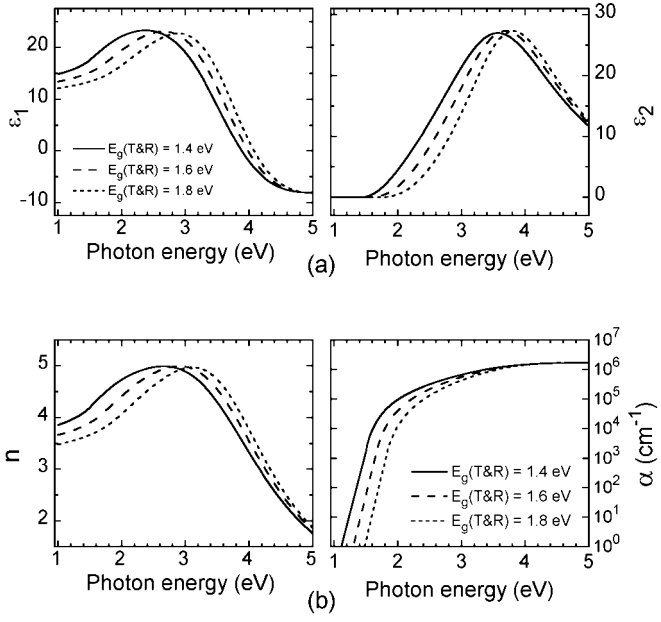


Figure 2.23 Room temperature optical properties including (a) $(\varepsilon_1, \varepsilon_2)$ and (b) (n, logα) for hypothetical a-Si:H alloys computed on the basis of a single specification of the optical band gap E_g(T&R) as determined from a conventional transmittance and reflectance measurement. These results were deduced from the relationships in Fig. 2.22 and from the analytical expression for the dielectric function in Eqs. (2.82)–(2.85). (After Ref. 76.)

shows some examples of the final results of this three-step process for the determination of (n, logα) given E_g(T&R).[76]

Figures 2.24 through 2.26 are presented as examples of the necessary ingredients for modeling real time SE data collected during plasma-enhanced chemical vapor deposition (PECVD) of microcrystalline Si, denoted μc-Si:H.[79] Under a wide range of PECVD conditions, such films evolve from an a-Si:H structure through a mixed-phase amorphous + microcrystalline (a + μc) structure, and finally stabilize in a single

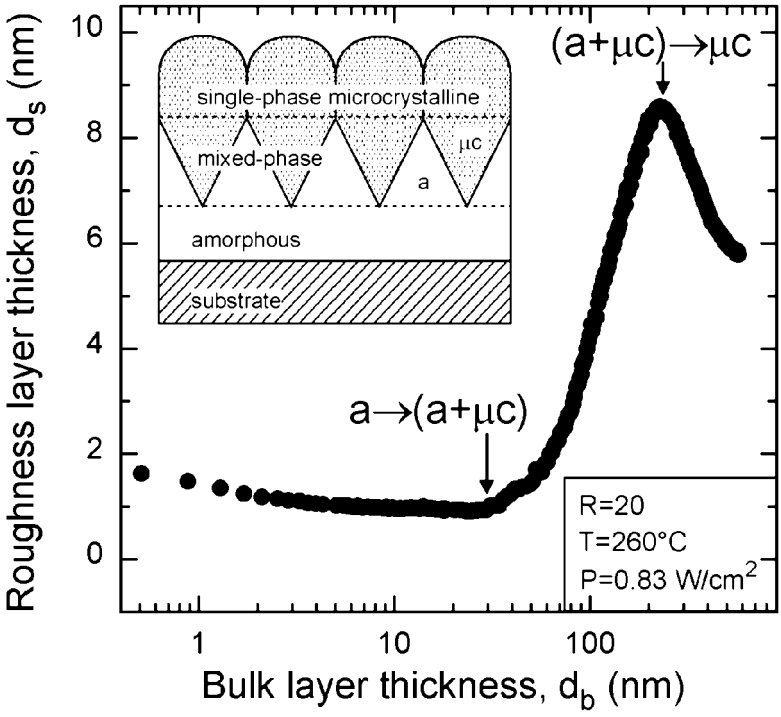


Figure 2.24 Results from real time spectroscopic ellipsometry depicting the dependence of the surface roughness layer thickness on the accumulated bulk ayer thickness for a silicon film prepared on oxide-covered crystalline silicon by plasma-enhanced chemical vapor deposition. This film evolves during growth from the amorphous phase to a microcrystalline phase with increasing thickness. The inset shows a simplified schematic of the film structure and the arrows denote the amorphous-to-(mixed-phase microcrystalline) nucleation transition and the (mixed-phase)-to-(single-phase) microcrystalline coalescence transition. Measurements of the surface roughness thickness are required in order to extract the dielectric functions at any stage of the growth process, as can be inferred from results in Figs. 2.25 and 2.26. (After Ref. 79.)

phase μc-Si:H structure (see inset of Fig. 2.24). This structural evolution is seen most clearly in the surface roughness layer thickness plotted as a function of the bulk layer thickness in Fig. 2.24, in which the surface roughness layer is modeled using the Bruggeman effective medium theory as a 0.5/0.5 vol. fraction mixture of the subsurface layer material and void. After a critical bulk layer thickness of about 30 nm, microcrystals nucleate from the a-Si:H, and grow preferentially, leading to the an increase in surface roughness thickness. After a second critical bulk layer thickness of 220 nm, the crystalline inclusions make contact and coalesce, leading to surface smoothening and single-phase μc-Si:H.

To analyze the real time SE data for such a complicated film structure, the deposition is divided into three regimes.[79] These regimes include: (i) single-phase a-Si:H growth applying a thickness-independent dielectric function, (ii) mixed-phase (a + μc)-Si:H growth applying a dielectric function determined from the dielectric functions and volume fractions of separate a-Si:H and μc-Si:H phases using the Bruggeman effective medium approximation, and (iii) single-phase μc-Si:H growth returning again to a thickness-independent dielectric function. In the first regime, the time dependences of the surface roughness and bulk layer thicknesses are determined along with the a-Si:H dielectric function. These results are obtained by applying a two-layer least-squares regression analysis, coupled with numerical inversion at a bulk layer thickness of 20 nm. A similar approach is applied in the third regime to extract the surface roughness thickness and the dielectric function of the growing μc-Si:H film; however, here the complicated substrate information is buried within a pseudo-substrate approximation. A full virtual interface analysis is applied in the intervening mixed-phase regime in which the evolution of the surface roughness thickness and μc-Si:H volume fraction is determined while the deposition rate is fixed at a value obtained using a simplified two-layer roughness/bulk model throughout the growth.

Figure 2.25 shows the volume fraction of the crystalline material plotted versus the projected bulk layer thickness, along with the behavior predicted from a cone growth model (see the inset in Fig. 2.24). Figure 2.26 compares the dielectric functions of the resulting a-Si:H and μc-Si:H obtained in the first and third analysis regimes. The solid lines are fits to the dielectric functions performed using the five-parameter, one-oscillator model for a-Si:H with $G(E) = G_C(E)$ as in Eq. (2.80*b*) and using an eight-parameter, two-oscillator model for μc-Si:H with $G(E) = G_T(E)$ as in Eq. (2.79b) and with ε_{0s} variable. In the fits to both dielectric functions, the Urbach tail contribution is omitted since only the SE data are available. Such data do not provide sufficient sensitivity to low values of ε_2 to detect the Urbach tail.

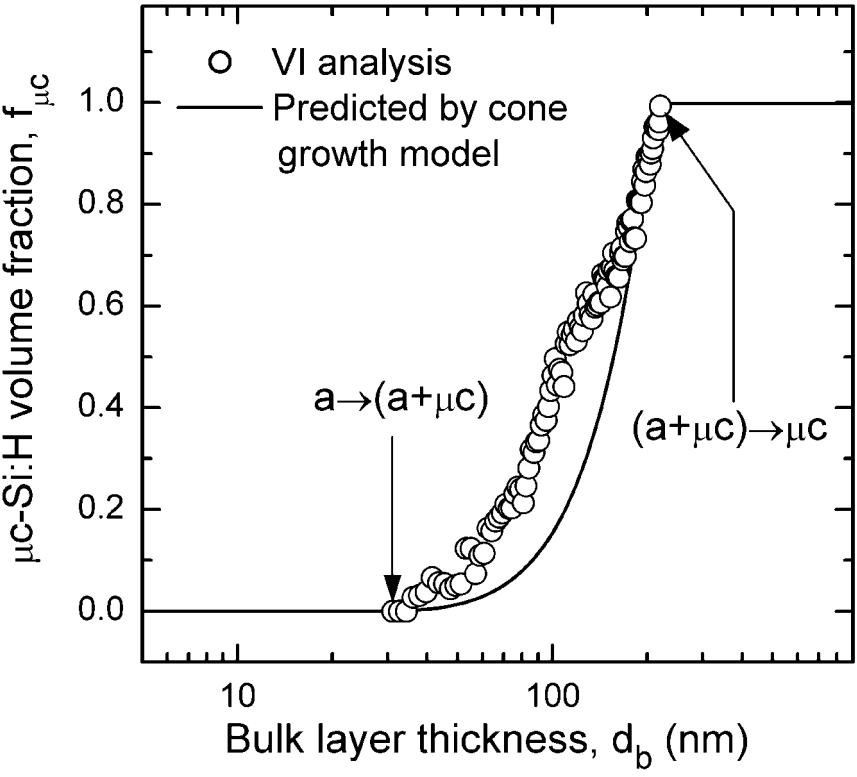


Figure 2.25 Results from real time spectroscopic ellipsometry for the evolution of the microcrystalline silicon volume fraction in the bulk layer during plasma-enhanced chemical vapor deposition for the film of Fig. 2.24 that evolves from an amorphous structure to a microcrystalline structure. These results were obtained using a two-layer virtual interface approximation. The solid line shows the result that would be predicted from a simple geometrical model of inverted cones as in the inset of Fig. 2.24. (After Ref. 79.)

In a final application of such formulas, Fig. 2.27 shows the temperature dependence of the eight parameters that define the dielectric function of a 30 nm thick p-type μc-Si:H film prepared on an a-Si:H surface held at 200° C. The sample structure is typical of that used in a-Si:H-based solar cells.[76] The results of Fig. 2.27 were obtained by extracting the dielectric functions from successive numerical inversions of real time SE data collected after deposition and upon cooling the μc-Si:H film from the substrate temperature of 200° C. In the numerical inversions, the microstructural parameters extracted from the analysis of real time SE data collected at the very end of film growth

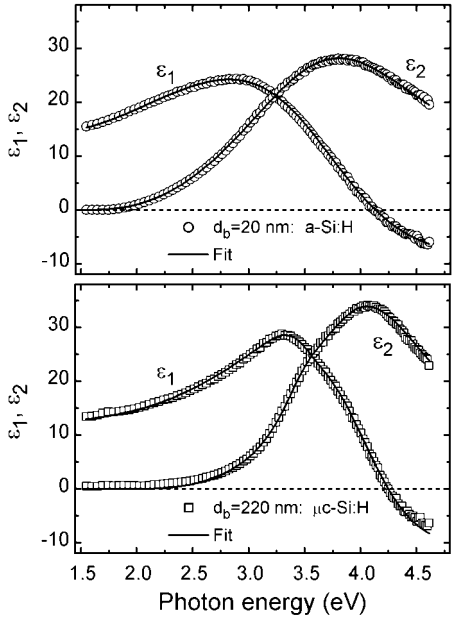


Figure 2.26 Real and imaginary parts of the dielectric functions (ε_1, ε_2) in the amorphous (top) and microcrystalline (bottom) phases for the silicon film of Figs. 2.24 and 2.25 obtained by real time SE during film growth. The dielectric functions of near surface material were obtained for the amorphous phase just before the onset of microcrystalline silicon nucleation and for the microcrystalline phase just after microcrystallite coalescence. The solid lines represent fits using Eqs. (2.80)–(2.82) with a single Lorentz oscillator for the a-Si:H dielectric function and using Eqs. (2.79), (2.81), and (2.82) with two Lorentz oscillators for the μc-Si:H dielectric function. (After Ref. 79.)

(as in Fig. 2.24) are assumed to remain constant during the cooling process. In fact, Fig. 2.27 also demonstrates the usefulness of real time SE in developing optical property databases versus a material characteristic such as temperature. In this example, the linear coefficients provided below each curve in Fig. 2.27 are obtained with confidence from fits to ~50 dielectric function spectra, and thus are well-suited for the purposes of interpolation and limited extrapolation in order to compute the dielectric functions for specified temperatures. An important application of this capability is the ability to deduce the dielectric function of the μc-Si:H appropriate for ex situ measurements at room temperature. Another application includes the possibility of assessing the effect of temperature on the optical performance of solar cells that incorporate such layers.

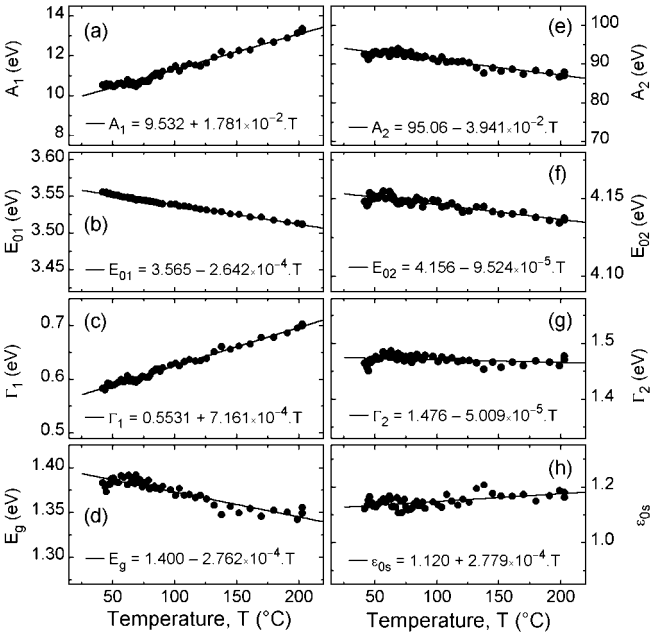


Figure 2.27 Best fit parameters in an analytical expression for the dielectric function of a microcrystalline silicon film plotted versus measurement temperature. The analytical expression employs eight free parameters, including two Lorentz oscillators, a Tauc absorption onset with a bandgap E_g, and an energy-independent contribution to ε_1, ε_{0s}. These results were obtained in situ by real time spectroscopic ellipsometry continuously during sample cooling from the deposition temperature of 200° C to 40° C. (After Ref. 76.)

2.5.3.3 Optical Characteristics: Poly- and Single-Crystalline Semiconductors

In Section 2.5.1, applications of the multiple Lorentz oscillator model were described for fitting the dielectric functions of crystalline semiconductors, and results for GaAs were presented in Figs. 2.13 and 2.14 as an example. Such an approach has several drawbacks as described next. First, the critical point features in the dielectric functions of crystalline solids arise from a superposition of electronic transitions between pairs of bands often covering appreciable regions of the Brillouin zone in $\vec{k}$-space. As a result, more complicated functional forms than that of the classical damped harmonic oscillator are required to characterize these features, as has been derived in detail in Section 2.4. In fact, if the Lorentz oscillator

expression is used to simulate such a critical point feature in the dielectric function, the best fit broadening parameter is inevitably larger than the true width of the feature due to the presence of the additional electronic transitions occurring over a range of photon energies. Because of this latter problem, broad Lorentz oscillator features are sometimes required in simulations to fit the dielectric function over regions of photon energy where no critical points are present. Second, the use of several Lorentz oscillators in such simulations, generally leads to appreciable ε_2 values below the lowest direct gap due to the accumulated tails of the oscillators. In contrast, the actual ε_2 values measured on crystalline semiconductors below the lowest direct gap are negligible by comparison. Third, the Lorentz oscillator expression lacks a key feature of the more realistic expressions for ε in Eqs. (2.61) and (2.63), namely the $[E_{cv}(\vec{k})]^{-2}$ factor that arises because the Bloch functions are eigenfunctions of the crystal momentum operator. Fourth, as noted in Section 2.4, the broadening of electronic transitions in solids is often more closely fit using a Gaussian lineshape function (or even a superposition of Gaussian and Lorentzian lineshape functions), rather than the pure Lorentzian lineshape that was introduced in Section 2.4 by assuming an exponential onset of the optical field.

(a) The Generalized Oscillator Model

Leng and coworkers started with the following Lorentz oscillator equation for modeling the dielectric functions of crystalline semiconductors:

$$\varepsilon(E) = \varepsilon_{0s} + \frac{1}{2}\sum_j \frac{A_j^2}{E_{0j}}\left(\frac{1}{E_{0j} - E - i\Gamma_j} + \frac{1}{E_{0j} + E + i\Gamma_j}\right). \quad (2.86)$$

This expression is most closely related to the atomic oscillator result of Eq. (2.50), where as usual A_j is an amplitude, E_{0j} is the resonance energy, and Γ_j is the broadening parameter associated with the j^{th} oscillator (all in units of energy).[53] Leng et al. proceeded to address the above limitations of the classical oscillator model through three key modifications of Eq. (2.86).

The first modification of Eq. (2.86) involves generalizing the two lineshape functions so that they can simulate the effects of the $\vec{k}$-space integration:

$$\varepsilon(E) = \varepsilon_{0s} + \sum_j C_{0j}\{[(E_{0j} - E - i\Gamma_j)^{-p_j}\exp(i\varphi_j)] + [(E_{0j} + E + i\Gamma_j)^{-p_j}\exp(-i\varphi_j)]\}. \quad (2.87)$$

In this expression, different exponents $-p_j$ are possible, simulating different critical point dimensionalities. In addition, different phases φ_j are possible, representing different critical point types, i.e., minima, maxima,

and saddle points [see Eqs. (2.69) and Fig. 2.4]. Although Eq. (2.87) is Kramers-Kronig consistent as long as $p_j > 0$, it does not obey the plasma sum rule given by Eq. (2.41). In other words, it no longer follows the high energy asymptotic behavior $\varepsilon_1(E) \sim 1-(A/E)^2$ and $\varepsilon_2(E) \sim A^2\Gamma/E^3$, where $A \equiv (e^2\hbar^2 N_e/\varepsilon_0 m)^{1/2}$, as described in Section 2.5.1.

Even with the modified expression of Eq. (2.87), the second difficulty inherent in the Lorentz oscillator model remains, namely the excess optical absorption at energies below the lowest direct gap. In the case of nano/microcrystalline semiconductors, this difficulty is addressed by premultiplying all lineshape functions $L_j(E)$ by a common absorption onset function $G(E)$ in Eqs. (2.77), (2.81), and (2.82). However, this leads to very tedious Kramers-Kronig integrals whose evaluation was presented in Eqs. (2.83) and (2.84). To address this difficulty in another way, Eq. (2.87) is expanded in a power series in E that converges for all photon energies lower than the lowest resonance energy, i.e., below the lowest direct band gap (assuming $\Gamma_j \ll E_{0j}$ for all j):

$$\varepsilon(E) = \varepsilon_{0s} + \sum_j \sum_{n=0}^{\infty} \binom{-p_j}{n} C_{0j}\{[(-1)^n (E_{0j} - i\Gamma_j)^{-p_j - n} \exp(i\varphi_j)] + [(E_{0j} + i\Gamma_j)^{-p_j - n} \exp(-i\varphi_j)]\} E^n. \tag{2.88}$$

Here, the first factor in parentheses after the two summations represents the general binomial coefficient. The factor $(-1)^n$ in the first term of the summation ensures that the even contributions $n = 0, 2, 4 \ldots$ are purely real, and that the odd contributions are purely imaginary.

By subtracting the $n = 0$ and 1 contributions of the power series from the sum in Eq. (2.87), the low energy asymptotic behavior of Eq. (2.87) now follows $\varepsilon_1(E) \sim \varepsilon_{0s} + C'_{02}E^2$ and $\varepsilon_2(E) \sim C'_{03}E^3$, where the C'_{0n} values are constants.[53] With this modified low energy behavior, the third limitation can now be corrected by introducing the E^{-2} factor for consistency with the quantum mechanical expressions of Eqs. (2.61) and (2.63), for example. As a result, Kramers-Kronig consistency is now recovered (as long as $p_j > 0$) without losing the suppression of absorption below the band gap. The final result can be written as follows:

$$\begin{aligned}\varepsilon(E) = \varepsilon_{0s} + \\ \frac{1}{E^2}\sum_j C_{0j}\{&[(E_{0j} - E - i\Gamma_j)^{-p_j}\exp(i\varphi_j)] \\ &+ [(E_{0j} + E + i\Gamma_j)^{-p_j}\exp(-i\varphi_j)] - 2\mathrm{Re}[(E_{0j} - i\Gamma_j)^{-p_j}\exp(i\varphi_j)] \\ &- 2ip_j E\,\mathrm{Im}[(E_{0j} - i\Gamma_j)^{-p_j-1}\exp(i\varphi_j)]\}.\end{aligned} \tag{2.89}$$

An application of this formula will be demonstrated later in which the dielectric function of crystalline silicon is simulated using a total of four oscillators with 21 free parameters:$\{\varepsilon_{0s}, (C_{0j}, E_{0j}, \Gamma_j, \varphi_j, p_j); j = 1, ..., 4\}$.

(b) The Generalized Critical Point Model

A second even more general approach for modeling the dielectric function of crystalline semiconductors has been developed by Johs et al.[55] The analytical model includes m energy-bounded, Gaussian-broadened polynomials and P unbounded Gaussian-broadened single resonances due to discrete electronic transitions outside the measured range of energy. The starting expression is given by:

$$\varepsilon(E) = 1 + i\sum_{j=1}^{m}\int_{E_{\min}}^{E_{\max}} W_j(E')\Phi_j(E, E', \sigma_j)dE' + i\sum_{j=m+1}^{j=m+P} A_j\int_0^{\infty}\delta(E' - E_j)\Phi_j(E, E', \sigma_j)dE'. \quad (2.90a)$$

The following three paragraphs describe the different features of Eq. (2.90a).

First, the function $\Phi_j(E, E', \sigma_j)$ under the two integrals describes the broadening of the j^{th} polynomial or resonance. In general, this function is given by:

$$\Phi_j(E, E', \gamma_j) = \int_0^{\infty} ds\ \exp\{i[E - E' + i\gamma_j(s)]s\} - \int_0^{\infty} ds\ \exp\{i[E + E' + i\gamma_j(s)]s\}, \quad (2.90b)$$

where s is an integration variable with units of inverse energy, $\gamma_j = \Gamma_j$ for Lorentzian broadening mechanisms (adopted so far in this chapter for mathematical transparency), and $\gamma_j = 2\sigma_j^2 s$ for Gaussian broadening mechanisms.[54] In the case of Gaussian broadening, one can transform to the unitless variables

$$\xi_{1j} = \frac{E - E'}{2\sqrt{2}\sigma_j} \quad \text{and} \quad \xi_{2j} = \frac{E + E'}{2\sqrt{2}\sigma_j},$$

where the subscript j indicates that the transformation depends in general on the polynomial through the broadening parameter σ_j. Solving Eq. (2.90b)

via Laplace transforms for the case of Gaussian broadening one arrives at the following expressions:

$$\Phi_j(E, E', \sigma_j) = \sqrt{\frac{\pi}{8\sigma_j^2}}\{\exp(-\xi_{1j}^2)\mathrm{erfc}(-i\xi_{1j}) - \exp(-\xi_{2j}^2)\mathrm{erfc}(-i\xi_{2j})\} \tag{2.90c}$$

$$= \sqrt{\frac{\pi}{8\sigma_j^2}}\{\Psi(\xi_{1j}) - \Psi(\xi_{2j})\}. \tag{2.90d}$$

Here erfc designates the complementary error function given by $\mathrm{erfc}(z) = 1 - \mathrm{erf}(z)$, where

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z \exp(-t^2)dt \quad \text{and} \quad \Psi(z) \equiv \exp(-z^2)\mathrm{erfc}(-iz).$$

Returning to the description of Eq. (2.90a), ($E_{\min}$, $E_{\max}$) are the minimum and maximum photon energies that bound the full set of polynomials and $W_j(E')$ is the j^{th} polynomial segment of order N_j, given by:

$$W_j(E') = \sum_{k=0}^{N_j} p_{jk}E'^k\,\Theta(E' - E_{Lj})\,\Theta\,(E_{Uj} - E') \tag{2.90e}$$

$$= \sum_{k=0}^{N_j} q_{1jk}\xi_{1j}^k\,\Theta\left(\frac{E - E_{Lj}}{2\sqrt{2}\sigma_j} - \xi_{1j}\right)\Theta\left(\xi_{1j} - \frac{E - E_{Uj}}{2\sqrt{2}\sigma_j}\right) \tag{2.90f}$$

$$= \sum_{k=0}^{N_j} q_{2jk}\xi_{2j}^k\,\Theta\left(\xi_{2j} - \frac{E + E_{Lj}}{2\sqrt{2}\sigma_j}\right)\Theta\left(\frac{E + E_{Uj}}{2\sqrt{2}\sigma_j} - \xi_{2j}\right). \tag{2.90g}$$

In Eq. (2.90e), p_{jk} is the k^{th} order coefficient for polynomial j, and $\Theta(t)$ is the unit step function centered at $t = 0$ such that $\Theta(t) = 0$ for $t < 0$ and $\Theta(t) = 1$ for $t > 0$. E_{Lj} and E_{Uj} are the lower and upper photon energy limits for the j^{th} polynomial, such that $E_{L1} = E_{\min}$ and $E_{Um} = E_{\max}$. Corresponding polynomial sums are given in Eqs. (2.90f) and (2.90g) in which the integration variable E' has been transformed to the unitless variables

$$\xi_{1j} = \frac{E - E'}{2\sqrt{2}\sigma_j} \quad \text{and} \quad \xi_{2j} = \frac{E + E'}{2\sqrt{2}\sigma_j}.$$

Finally, the second integration in Eq. (2.90a) represents a collection of P discrete electronic transitions, represented by broadened delta functions

having amplitudes, transition energies, and broadening parameters of $\{(A_j, E_j, \sigma_j); j = m + 1, ..., m + P\}$ with E_j either below E_{min} or above E_{max}. This term generates energy-dependent contributions to $\varepsilon_1(E)$ and tails in $\varepsilon_2(E)$ that extend into the spectral range spanned by the polynomials.

Combining all these results together using the unitless variables leads to the following expression:

$$\varepsilon(E) = 1 - i\sqrt{\pi}\sum_{j=1}^{m}\sum_{k=0}^{N_j}\left[q_{1jk}\left\{\Xi_k\left(\frac{E - E_{Uj}}{2\sqrt{2}\sigma_j}\right) - \Xi_k\left(\frac{E - E_{Lj}}{2\sqrt{2}\sigma_j}\right)\right\} - q_{2jk}\left\{\Xi_k\left(\frac{E + E_{Uj}}{2\sqrt{2}\sigma_j}\right) - \Xi_k\left(\frac{E + E_{Lj}}{2\sqrt{2}\sigma_j}\right)\right\}\right] + i\sum_{j=m+1}^{m+P}\sqrt{\frac{\pi A_j^2}{8\sigma_j^2}}\left\{\Psi\left(\frac{E - E_j}{2\sqrt{2}\sigma_j}\right) - \Psi\left(\frac{E + E_j}{2\sqrt{2}\sigma_j}\right)\right\}, \quad (2.91a)$$

where q_{1jk} and q_{2jk} are the coefficients of the polynomials of Eqs. (2.90f) and (2.90g). The final summation incorporates the amplitudes A_j of the resonances outside the energy range. In addition, the function Ξ_k for the k^{th} polynomial is defined generally by:

$$\Xi_k(z) = \int_0^z t^k \Psi(t) dt, \quad (2.91b)$$

where t is an integration variable and the function $\Psi(z)$ is given by $\Psi(z) \equiv \exp(-z^2)\ \mathrm{erfc}(-iz)$.

The significant differences in form between Eq. (2.91a) and the theoretical expressions derived earlier [see, e.g., Eq. (2.67b)]—and the consequent loss of mathematical transparency—arise from the fact that Gaussian rather than Lorentzian broadening is used. Johs et al. find that, although closed-form integration is not possible under the Gaussian broadening scheme, more rapid computation is possible owing to the ability to exploit the normalized energy variables that yield simple one-dimensional look-up tables for the complicated functions Ξ_k.[55] In order to make connection with the previous derivations, it is of interest to explore the forms of Eqs. (2.90a) and (2.91a) when Lorentzian broadening is assumed. As noted above, for Lorentzian broadening $\gamma_j(s) = \Gamma_j$ in Eq. (2.90b), and the application of a Laplace transform leads to the following result:

$$\Phi_j(E, E', \Gamma_j) = -i\left(\frac{1}{E' - E - i\Gamma_j}\right) + \left(\frac{1}{E' + E + i\Gamma_j}\right). \quad (2.92a)$$

Ultimately, this provides an expression for the dielectric function analogous to Eq. (2.91a) given by:

$$\varepsilon(E) = 1 + \sum_{j=1}^{m} \sum_{k=0}^{N_j} p_{jk} \int_{E_{Lj}}^{E_{Uj}} E'^k \left(\frac{1}{E' - E - i\Gamma_j} + \frac{1}{E' + E + i\Gamma_j} \right) dE'$$

$$+ \sum_{j=m+1}^{m+P} \frac{2A_j E_j}{E_j^2 - E^2 - 2i\Gamma_j E}. \qquad (2.92b)$$

The first term of this equation is similar in form to Eq. (2.67b) with the exception that

$$\sum_{c,v} J_{cv}(E') \frac{\left| \vec{e} \cdot \vec{p}_{cv}(E') \right|^2}{E'^2}$$

is replaced by the piecewise polynomial function

$$\sum_j W_j(E')$$

[see Eq. (2.90e)]. The second summation in Eq. (2.92b) includes a superposition of P Lorentz oscillators of amplitude $2A_jE_j$, resonance energy E_j, and broadening parameter $2\Gamma_j$. In the denominator of each summand, a Γ_j^2 term is neglected.

Finally, further details related to the implementation of the approach of Eq. (2.91a) or (2.92b) will be reviewed.[55] First, the m polynomial segments are collected in groups of four, each group centered at a critical point. As a result, if there are N_c critical points, then there will be $m = 4N_c$ polynomial segments. The last and first polynomial segments of two neighboring ensembles overlap in order to account for the electronic transitions that occur between critical points. This approach has the advantage of avoiding the arbitrary addition of resonance features for this purpose, as is necessary when using a superposition of Lorentz oscillators to simulate the dielectric functions of single crystal semiconductors. Figure 2.28 shows a typical such ensemble of four polynomials used in a simulation of the dielectric function in the region around a single critical point. The two extremum energies $E_{\min}$ and $E_{\max}$, along with the critical point energy E_c and center amplitude A are specified absolutely, and the other amplitudes and energies are specified relative to these. As a result, in fitting dielectric function spectra for slightly different material characteristics (e.g., composition or temperature) the critical point ensemble can be

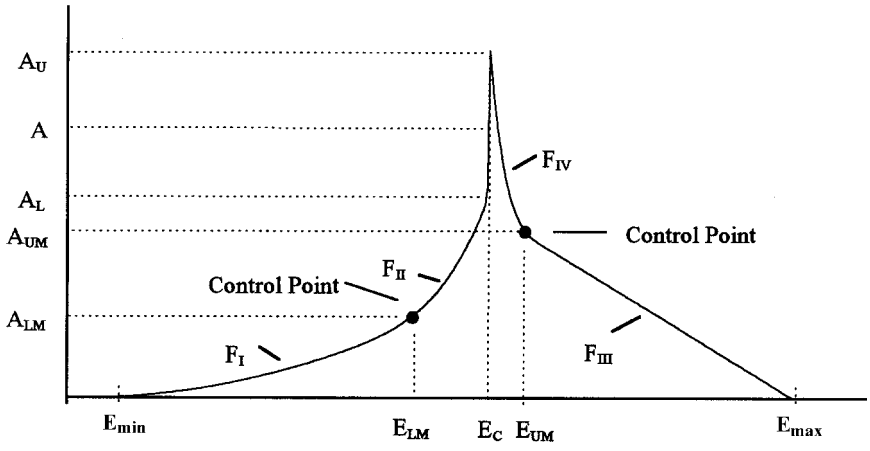


Figure 2.28 A generalized critical point structure in the absence of broadening. This structure consists of four polynomials, two on each side of the critical point energy E_c. In typical fits, the two outermost polynomials are of second order whereas the two innermost polynomials are of fourth order. At the control points, the inner and outer polynomials are continuous as are their first and second derivatives. A discontinuity is possible at the critical point that can simulate the superposition of critical points of different dimensionality. (After Ref. 55.)

shifted in energy or changed in amplitude without affecting its overall shape. The polynomials and their first and second derivatives are assumed to be continuous at the control points. At the critical point energy, however, a discontinuity is allowed in order to provide the capability of simulating the step-like absorption at two-dimensional critical points. Typically, the outer polynomials of the ensemble are chosen to be of second-order (i.e., $N_j = 2; j = 1,4$) whereas the inner polynomials are chosen to be of fourth order (i.e., $N_j = 4, j = 2,3$).

2.5.3.4 Examples: Crystalline Semiconductors

(a) Generalized Oscillator Model

In the first application of Eq. (2.89), Leng and coworkers have simulated the room temperature dielectric function of single crystal silicon.[53] Among the simulations investigated, that selected as providing the overall best fit includes a sum of four generalized oscillators in Eq. (2.89), corresponding to the E_0', E_1, E_2, and E_1' critical points in the silicon band structure. As noted earlier, each generalized oscillator requires five free parameters in the fit, namely the amplitude C_{0j}, phase φ_j,

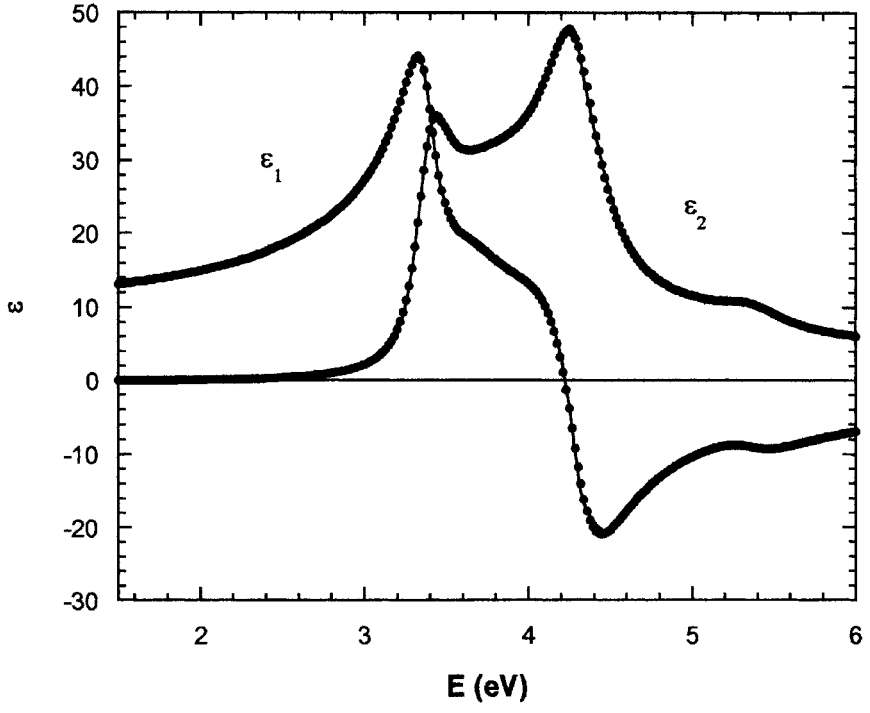


Figure 2.29 Dielectric function for crystalline silicon determined experimentally using SE in conjunction with an in situ cleaning procedure (solid lines). Also shown is the best fitting dielectric function using the generalized oscillator model of Eq. (2.89) (i.e., fully modified) incorporating four oscillators (points). The best fit parameters are given in Table 2-3 and the difference between the two spectra are given in Fig. 2.30d. (After Ref. 53.)

resonance energy E_j, broadening parameter Γ_j, and exponent p_j ($j = 1, ..., 4$). Figure 2.29 shows a comparison of the experimental data for the dielectric function of silicon (solid line) and this 21-parameter best fit (points). The 20 parameters associated with the four oscillators are given in Table 2-3, and the best fit value of the final parameter ε_{0s} in Eq. (2.89) is -0.047 (a non-physical result). With this parameter set, the data and fit are indistinguishable on the scale of Fig. 2.29. Also given in Table 2-3 are the resonance energies and broadening parameters as deduced by Lautenschlager et al. from a fit to the second derivative of the dielectric function, applying Eq. (2.70).[22] These results demonstrate the correspondence between each generalized oscillator and a band structure critical point.

Table 2-3. Best-fit values of the oscillator parameters $\{C_{0j}, \varphi_j, p_j, E_{0j}, \Gamma_j; j = 1, \ldots 4\}$ used by Leng et al. in a simulation of the dielectric function of crystalline silicon applying the generalized four-oscillator model of Eq. (2.89) (results indicated below as GO). The associated simulations in this case are given in Figs. 2.29 and 2.30d. Also shown are the results of the critical point analysis of dielectric function data by Lautenschlager et al.[22] applying Eq. (2.70) (results indicated as CP). In the latter analysis, all critical points are fit to a 2D lineshape (ln function) except E_1 which is fit to an excitonic lineshape ($p_j = 1$).

	Oscillator Amplitude C_{0j} (eV^{2-p})		Phase φ_j (rad)		Exponent $-p_j$		Energy E_{0j} (eV)		Broadening Γ_j (eV)	
j	CP	GO	GO	CP	GO	CP	GO	CP	GO	CP
1	E_0'	56.32	−0.4589	0.20	−0.8241	ln	3.3800	3.320	0.1150	0.070
2	E_1	240.90	−0.4110	0.05	−0.3965	−1	3.6266	3.396	0.3079	0.098
3	E_2(X)	125.46	0.3307	2.10	−0.9504	ln	4.2906	4.270	0.2030	0.087
4	E_1'	16.66	0.2816	1.45	−0.9761	ln	5.3825	5.317	0.2410	0.132

In Fig. 2.30, spectra describing the differences between the best fit ε_1 and ε_2 and the corresponding experimental data are plotted, comparing analytical models of the silicon dielectric function having different levels of complexity.[53] (Note that the scale in the original figure is erroneous; the upper plot limit of 6 eV should be replaced by 5 eV and the scale adjusted accordingly.) The curves designated (a) in Fig. 2.30 result from a reduced-range fit (1.5–5.1 eV) using an 11-parameter model of two generalized oscillators without the low energy subtraction and without the E^{-2} factor. The two oscillators in this model simulate the dominant features, the E_0'–E_1 critical point complex near 3.4 eV and the E_2 critical point near 4.2 eV. Clearly the best-fit ε_2 spectrum for this model overestimates the experimental spectrum at low photon energies, a pervasive problem when employing multiple oscillators in fits of semiconductor dielectric functions below the lowest direct band gap. The curves (b) in Fig. 2.30 result from a similar fit (1.5–5.1 eV) using the 11-parameter model of two generalized oscillators; however, the first-order term of the power series in ε vs. photon energy E is subtracted from the two oscillator sum. This model is similar to that of Eq. (2.89) except that the zero-order term in ε_1 is not subtracted and the E^{-2} factor is not included. As a result, ε_2 is suppressed to third order in E at low energies, and this accounts for the improvement in the fit to ε_2. Curves (c) result from a reduced-range fit (1.5–5.1 eV) with two generalized oscillators and with the full modifications in accordance with

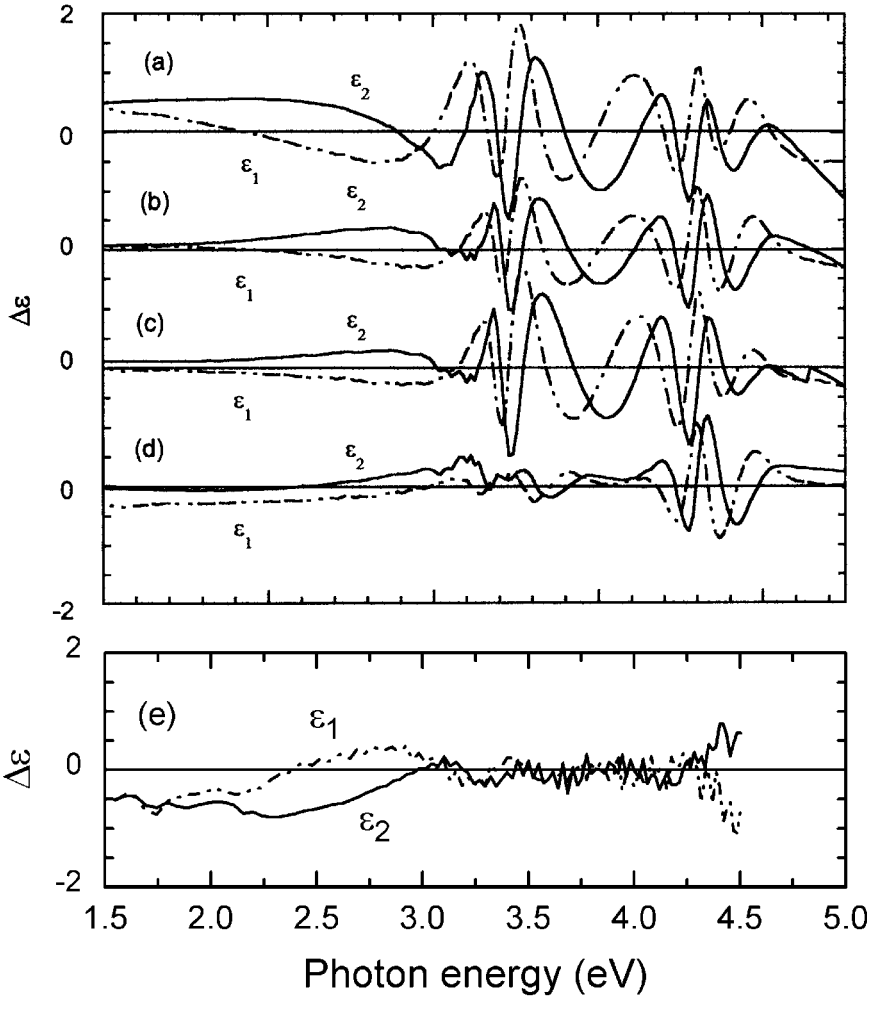


Figure 2.30 Difference between the simulated and experimental dielectric function spectra for crystalline silicon using the following different models for the dielectric function: (a) the generalized oscillator model of Eq. (2.87) (i.e., without any modifications) incorporating two oscillators; (b) generalized oscillator model, modified to suppress ε_2 at low energy, incorporating two oscillators; (c) generalized oscillator model of Eq. (2.89) (i.e., fully modified) incorporating two oscillators; (d) as in (c) but incorporating four oscillators. Also shown in (e) is the difference between the simulated and experimental dielectric function spectra for the microcrystalline silicon film of Fig. 2.26. In this case, the model applies Eqs. (2.79), (2.81), and (2.82) and incorporates two Lorentz oscillators. (After Ref. 53).

Eq. (2.89). Curves (d) represent the difference in Fig. 2.29 between the 21-parameter, full-range fit (1.5–6.0 eV) and the experimental data. A comparison of the curves in (d) with those in (a) through (c) demonstrates that two oscillators are required centered near 3.4 and 3.6 eV to obtain an adequate fit over the energy range of the E_0'–E_1 complex. For the best fit model of Fig. 2.30 and Table 2-3, however, further statistical analysis is required in order to assess whether the added complexity of the generalized oscillators is required for all four critical point structures.

Based on the results of Figs. 2.29 and 2.30, the generalized oscillator model should also provide suitable simulations of the dielectric functions of polycrystalline silicon materials having grain sizes $L_g > 100$ nm, in which case the dielectric function features remain sharp.[80] For much smaller grain sizes, particularly those in the nano/microcrystalline regime with $L_g > 10$ nm, however, two effects generally occur.[52,76] First, the critical point structures are broadened due to the relaxation of $\vec{k}$-conservation. Such an effect is a manifestation of electron scattering at grain boundaries that also reduces the excited state lifetime. This mechanism has been proposed to lead to an increase in the critical point broadening parameters with grain size L_g in accordance with $\Gamma(L_g) = \Gamma_0 + (C_g/L_g)$, where Γ_0 is the broadening parameter in the single crystal and C_g is a constant of the form $C_g \sim \hbar v$, with v being the excited state electron velocity.[81] Second, the weak absorption observed for single crystal Si in Fig. 2.29 over the range from 1.5 to 3.0 eV is enhanced in nano/microcrystalline Si also due to the relaxation of $\vec{k}$-conservation (providing the crystals are not so small as to lead to quantum size effects).

As a result of these considerations, the full complexity of four generalized oscillators is not required for nano/microcrystalline silicon as demonstrated in section 2.5.3.1. Furthermore, the modification of the oscillator model by the elimination of the first-order power series term in $\varepsilon_2(E)$ is not necessary either, due to the increase in absorption at low energies in the nano/microcrystalline materials. For these materials, the absorption onset function $G(E)$ in Section 2.5.3.1 provides the required flexibility to model this gradually increasing absorption. In Fig. 2.30(e), the differences between the analytical model and the experimental data are shown for the microcrystalline silicon film of Fig. 2.26. In this model, two Lorentz oscillators are used with absorption suppressed according to the Tauc gap expression for $G(E)$.[79] In this case the reverse of Fig. 2.30(a) is observed, namely that the experimental $\varepsilon_2(E)$ values at low energy are in fact larger than those in the model. In the region of the critical points, however, the eight-parameter, two-oscillator fit to the data for the microcrystalline silicon film is as good as the 21-parameter, four-oscillator fit to the data of single crystal silicon. Further efforts are needed to develop a single comprehensive analytical model suitable for simulating the dielectric functions for the entire range of silicon materials.

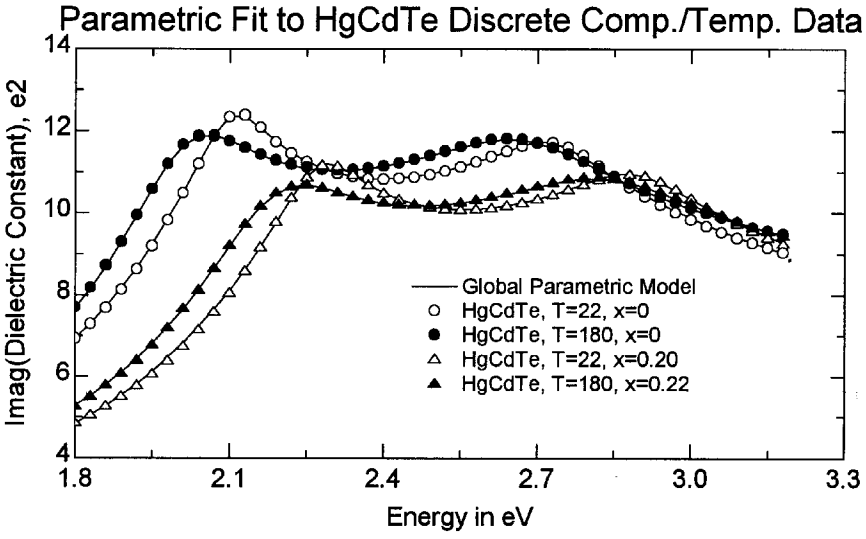


Figure 2.31 Spectra in the imaginary part of the dielectric function ε_2 for $Hg_{1-x}Cd_xTe$ ternary alloys of selected composition x prepared by molecular beam epitaxy and measured in situ at selected temperatures T by spectroscopic ellipsometry [(x, T) = (0, 22° C); (0, 180° C); (0.20, 22° C); (0.22, 180° C)] (points). The solid lines represent results from a global fit of these spectra using the generalized critical point model of Eqs. (2.91). In addition to the dielectric function spectra for the four (x, T) conditions depicted here, three other pairs of spectra [(x, T) = (0.17, 180° C); (0.27, 180° C); (0.32, 180° C)] were included in a global fitting procedure. (After Ref. 55.)

(b) Generalized Critical Point Model

Real time SE studies of epitaxial layers of mercury-cadmium telluride ($Hg_{1-x}Cd_xTe$) by Johs et al. provide an important application example of the generalized critical point model for the dielectric functions of crystalline semiconductor materials.[55] Figure 2.31 (points) provide selected ε_2 spectra of $Hg_{1-x}Cd_xTe$ measured in situ under different conditions of composition x and temperature. The solid lines are the fits obtained using the generalized critical point model of Eqs. (2.90) and (2.91). Similar such high-quality fits to both parts of the dielectric functions were obtained for five alloy compositions (x = 0, 0.17, 0.22, 0.27, and 0.32) measured at the growth temperature of T = 180° C and for two alloy compositions (x = 0 and 0.20) measured at room temperature (T = 22° C). These data sets were fit simultaneously using a global parametric analysis in which the dielectric function parameters (energies, amplitudes, and broadening parameters of the critical points within the measurement energy range and those of the oscillators outside the energy range) were modeled as quadratic functions

of composition and linear functions of temperature. Using the best fit coefficients that describe these dielectric function parameters vs. composition and temperature allows one to predict the dielectric function for any composition and temperature over the ranges spanned by the measurement. Examples of such predictions are given in Figs. 2.32(a) and 2.32(b). With such a predictive capability, both composition and temperature can be used in least squares regression analysis of $Hg_{1-x}Cd_xTe$ epitaxial growth.

Finally, an assessment will be made regarding the multitude of parameters associated with each critical point.[55] Considering a critical point structure composed of four polynomials (see Fig. 2.28), two of fourth order and two of second order, there will be 16 coefficients p_{jk} and 4 broadening parameters σ_j. However, six coefficients will be eliminated by the constraints that the polynomials, as well as their first and second derivatives, be continuous at the two control points. Thus, the two critical point structures apparent in Figs. 2.31 and 2.32 will have at most 28 parameters associated with them. Additional parameter triplets are added for every oscillator that is required outside the energy range of measurement. As will be shown in the next section, a large number of free parameters is characteristic of any model that uses polynomial segments in the dielectric function to portray accurately the physical processes of the electronic transitions. In the generalized critical point model, the polynomials are centered around each critical point whereas for the specific study of the dielectric function of GaAs to be described in the next section, the polynomials span the regions between the critical points.[54]

It should be pointed out that the large number of free parameters is sure to lead to parameter correlations and large confidence limits. This problem, compounded by the use of a limited number of dielectric function data sets (versus temperature and composition for the example of Fig. 2.32) in the fitting, may lead to erroneous results in interpolation. Additional efforts are needed in the future to assess this problem for the generalized oscillator and critical point models in which case the number of free parameters routinely exceeds 20.

2.5.4 Specialized Quantum Mechanical Models

The development of specialized quantum mechanical models for the dielectric functions of crystalline solids requires a basic understanding of their electronic band structures, i.e., the energy of the electronic states as a function of electron wavevector $\vec{k}$ for all the valence and conduction bands that can serve as initial and final states for the electronic transitions. Such an understanding can come from theoretical calculations supplemented by optical

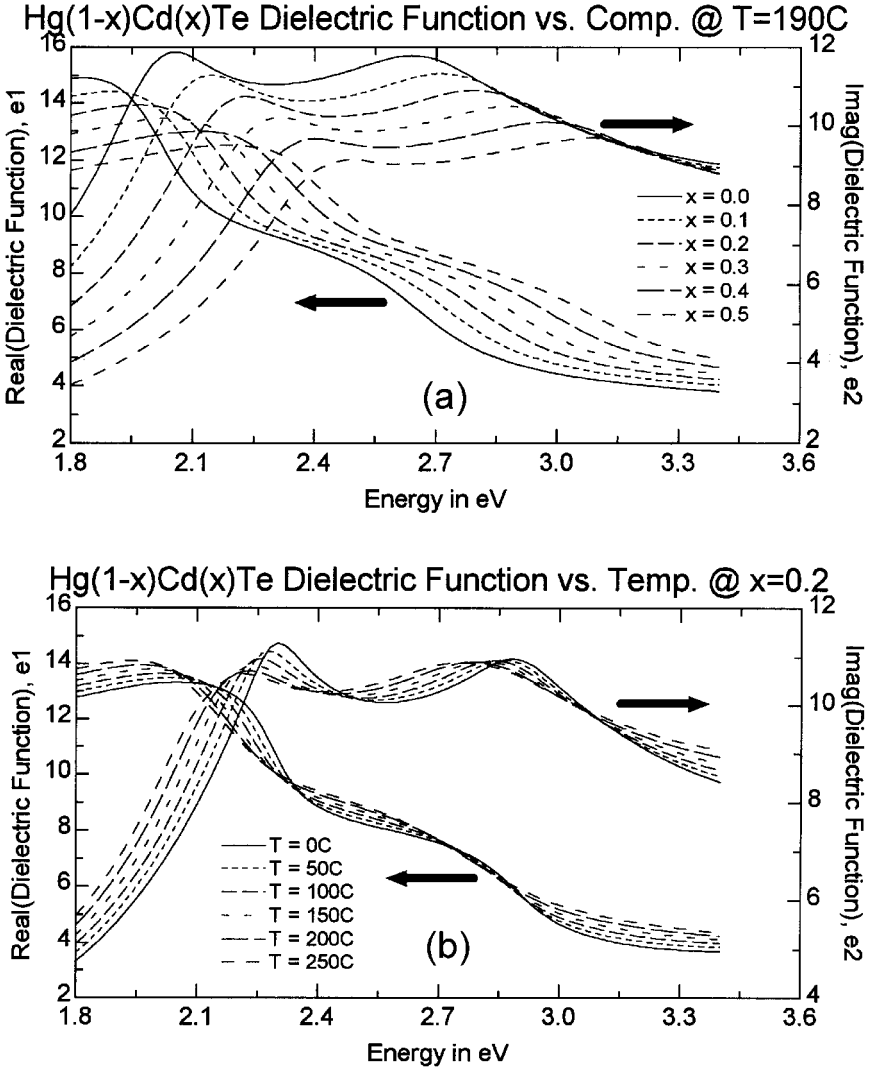


Figure 2.32 Dielectric functions (ε_1, ε_2) for $Hg_{1-x}Cd_xTe$ ternary alloys calculated from the generalized critical point model of Eqs. (2.91) assuming (a) different alloy compositions x at a fixed measurement temperature of $T = 190°$ C and (b) different measurement temperatures T at a fixed alloy composition of $x = 0.2$. The database used to establish the model was developed from the global fits of the dielectric functions in Fig. 2.31. (After Ref. 55.)

measurements. In particular, the analysis of SE data applying Eq. (2.69) provides information on the energy position of the critical points and the characteristics of the bands in the vicinity of the critical points. In this section, band structure-based models of the dielectric functions will be presented for a well-studied semiconductor and metal, gallium arsenide and aluminum.

2.5.4.1 Case Study: Gallium Arsenide

As an example of the identification of band structure critical points, Fig. 2.33 highlights the second derivative of the experimentally determined

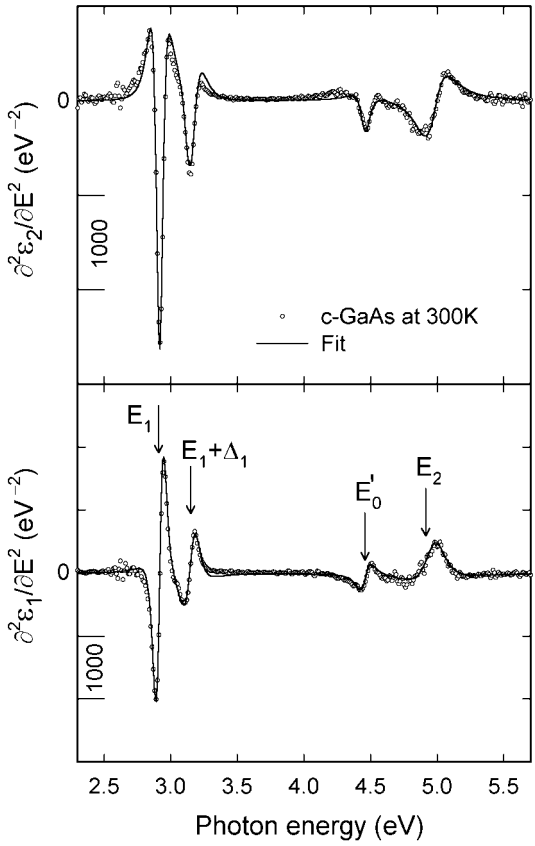


Figure 2.33 Second derivative spectra versus photon energy E in the real (bottom) and imaginary (top) parts of the complex dielectric function (ε_1, ε_2) for single crystal GaAs, measured at room temperature by spectroscopic ellipsometry (points). Also shown are fits to the experimental results (solid lines) using Eq. (2.70). The best fit critical point energies obtained from these fits are included in Fig. 2.13. (Data from Ref. 49).

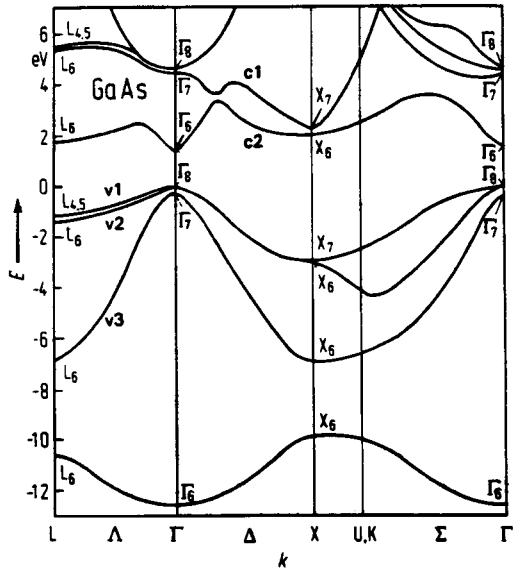


Figure 2.34 Band diagram of electron energy versus wavevector for GaAs, depicting the band designations used to identify the transitions listed in Table 2-4. (Band structure after Ref. 82; band designations after Ref. 54.)

real and imaginary parts of the dielectric function of GaAs at room temperature (points),[49] along with a fit using Eq. (2.70) at each of the critical points (lines). Four prominent critical points are evident in the data. Kim et al.[54] have identified the associated transitions on the band structure diagram[82] of Fig. 2.34 in the following way (i) E_1 at 2.92 eV occurring from $v1$ to $c2$ near Λ; (ii) $E_1 + \Delta_1$ at 3.15 eV occurring from $v2$ to $c2$ near Λ; (iii) E_0' at 4.46 eV occurring from $v1$ to $c1$ near Δ; and (iv) E_2 at 4.97 eV occurring from $v1$ to $c2$ near Σ. The band designations and $\vec{k}$-space regions (Λ, Δ, Σ) for these critical points can be located in Fig. 2.34; however, some of the assignment details are controversial.[21] Figure 2.13 includes the best fit energies of these four dominant critical points for comparison with the best-fit resonance energies in the multiple Lorentz oscillator model. From this comparison, one can determine which oscillators are actually fitting critical point features and which are simulating the broad background in the dielectric function. Because three of the oscillators do not correspond to critical points, one finds that the Lorentz oscillator model simply provides a good parameterization of the optical properties; however, it lacks a strong physical basis. This can be clarified by performing successive derivatives on the spectra from the experiment

and from the 7-oscillator model and observing the increasing discrepancies between the two results.[54]

A proposal for the electronic transitions in terms of the band structure has been developed in order to fit the dielectric function of GaAs and related III–V semiconductors, applying an improved model based on Eq. (2.67b).[54] Because this model accounts for the band structure of GaAs, it is physically realistic and, unlike the oscillator model, can simultaneously fit ε and its first, second, and third derivatives. The model begins with the following expression for the dielectric function:

$$\varepsilon(E) = 1 + i\frac{2e^2\hbar^2}{\varepsilon_0 m^2}\sum_{j=1}^{m}\int_{E_0}^{E_{\max}} \frac{w_j(E')}{E'^2}\,\Phi_j(E, E', \sigma_j)dE', \qquad (2.93)$$

where $w_j(E')$ is defined as a continuous function over the j^{th} energy range between two successive critical points $E_{j-1} \leq E' < E_j$ and is zero elsewhere [see, e.g., Fig. 2.35(e)]. Here, it is assumed that there are m critical points at energies: E_0 (fundamental band gap), … E_{m-1}, and that $E_m = E_{\max}$, the maximum energy of the measurement. Thus, Eq. (2.93) is also closely related to the first term of Eq. (2.90a) where $W_j(E') = [2e^2\hbar^2 w_j(E')]/[\varepsilon_0 m^2 E'^2]$. The difference arises from the fact that the $w_j(E')$ functions in Eq. (2.93) span the regions between successive critical points, whereas the $W_j(E')$ polynomials in Eq. (2.90a) are placed on both sides of a single critical point. In fact, Eq. (2.93) derives from Eq. (2.67b), and a rearrangement and comparison shows that

$$\sum_j w_j(E') = \sum_{c,v} \left|\vec{e}\cdot\vec{p}_{c,v}(E')\right|^2 J_{cv}(E').$$

In Eq. (2.93), however, Eq. (2.92a) has been employed to replace the Lorentzian broadening function with the more general function $\Phi_j(E, E', \sigma_j)$ of Eq. (2.90b). With this general function, either Lorentzian broadening [Eq. (2.92a)] or Gaussian broadening [Eq. (2.90d)] can be simulated.

The key to evaluating Eq. (2.93) for GaAs is to determine the functions $\{w_j(E'); j = 1, \ldots m\}$ for a specific model of the band structure as depicted in Figs. 2.34 and 2.35 and outlined in Table 2-4.[54] It is proposed that $\{w_j(E'); j = 1, \ldots m\}$ derive from a sum of the following four discontinuous functions with a total of eight critical points (one of which is suppressed because it is too weak to be observed experimentally). In the energy regions between the critical points where $J_{cv}(E')$ is smoothly varying, the $w_j(E')$ functions are described either by constants, specifically those in Table 2-2, or by linear or quadratic polynomial functions. In other words, the constants employed in Table 2-2 to describe $J_{cv}(E')$ on both

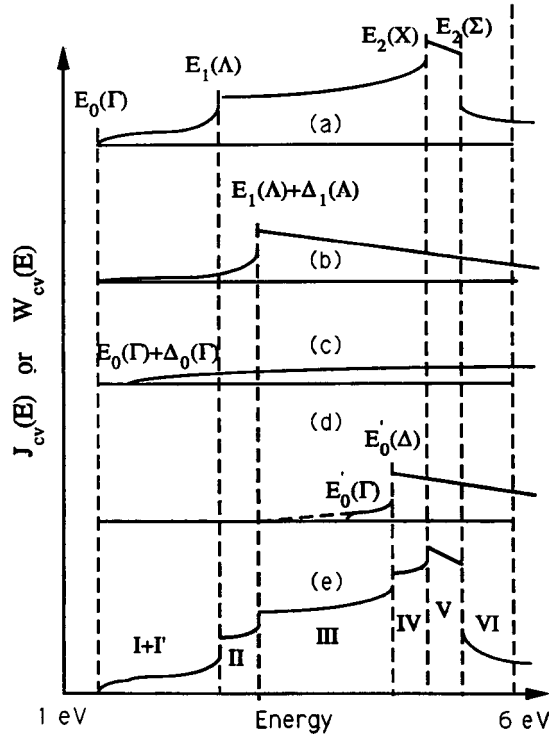


Figure 2.35 Schematic joint density of states functions $w_j(E')$ in Eq. (2.93) for crystalline GaAs using the scheme for electronic transitions given in Table 2-4. This plot depicts the functions associated with the transitions between the (initial, final) bands (a) ($v1$, $c2$); (b) ($v2$, $c2$); (c) ($v3$, $c2$); and (d) ($v1$, $c1$), where the band designations are provided in Fig. 2.34. In part (e), the summation of the five functions $\{w_j(E'); j = 1, \ldots 5\}$ is shown. This summation is divided into six continuous regions denoted (I + I', II, III, IV, V, VI), and the mathematical forms of the functions in these six regions are given in Eqs. (2.94a)–(2.94e). (After Ref. 54.)

sides of the critical points may be replaced by linear or quadratic functions of E'.

(i) The first set of contributions to $w_j(E')$ is shown in Fig. 2.35(a) and describes the electronic transitions from $v1$ to $c2$ in Fig. 2.34. This set includes the largest number of critical points over the measured energy range ($E_0 = 1.4$ eV $< E < E_{\max} = 6$ eV) since this pair of valence and conduction bands has the minimum energy separation for a given $\vec{k}$. The critical points include the fundamental bandgap $E_0(\Gamma)$ at 1.41 eV which is of the 3D M_0 type (minimum) with $J_{cv}(E) \propto \Theta(E - E_0)\ (E - E_0)^{1/2}$ (see

Table 2-4. Band structure transitions associated with the critical points observed below 6 eV in the dielectric function of GaAs. Also included are the critical point designation, type, and best-fit photon energy (data from Kim et al.[54]).

Band pair	Critical point	Type	Photon energy (eV)
$v1 \rightarrow c2$	$E_0(\Gamma)$	3D-M_0	1.411
$v1 \rightarrow c2$	$E_1(\Lambda)$	2D-M_0 + 3D-M_1	2.930
$v1 \rightarrow c2$	$E_2(X)$	2D-M_0 + 3D-M_1	4.811
$v1 \rightarrow c2$	$E_2(\Sigma)$	2D-M_2 + 3D-M_2	5.003
$v2 \rightarrow c2$	$E_0(\Gamma)$	3D-M_0	1.411
$v2 \rightarrow c2$	$E_1(\Lambda) + \Delta_1(\Lambda)$	2D-M_0 + 3D-M_1	3.171
$v3 \rightarrow c2$	$E_0(\Gamma) + \Delta_0(\Gamma)$	3D-M_0	1.748
$v1 \rightarrow c1$	$E_0'(\Delta)$	2D-M_0 + 3D-M_1	4.478

Table 2-2 and Fig. 2.4); the $E_1(\Lambda)$ and $E_2(X)$ at 2.93 and 4.81 eV, respectively, each of which is described as a superposition of a 2D M_0 transition with an upward step in $J_{cv}(E)$, i.e., $J_{cv}(E) \propto \text{const} + \Theta(E - E_j)$, and a 3D M_1 transition with $J_{cv}(E) \propto \text{const} - \Theta(E_j - E)\ (E_j - E)^{1/2}$; and finally the $E_2(\Sigma)$ at 5.00 eV which is described as a superposition of a 2D M_2 transition with a downward step in $J_{cv}(E)$, i.e., $J_{cv}(E) \propto \text{const} - \Theta(E - E_j)$, and a 3D M_2 transition with $J_{cv}(E) \propto \text{const} - \Theta(E - E_j)(E - E_j)^{1/2}$. The energy values quoted here were obtained in simultaneous fits to ε_2 and its first three derivatives, applying the model developed in this part; thus, they differ somewhat from those obtained in Fig. 2.33 and enumerated in Fig. 2.13. In addition, in the proportionalities for J_{cv}, $\Theta(t)$ is the unit step function such that $\Theta(t) = 0$ for $t < 0$ and $\Theta(t) = 1$ for $t > 0$. The superposition of 2D and 3D critical points is based on the concept that if two bands are exactly parallel along a symmetry direction yielding a 2D critical point (as occurs, e.g., for bands $v1$ and $c2$ near Λ), then there must also exist a 3D critical point of the same energy occurring in $\vec{k}$-space at the point along the same direction where the bands diverge from one another (e.g., the maximum in energy in the band $c2$ between Λ and Γ).

(ii) The second set of contributions is shown in Fig. 2.35(b) and describes the transitions from $v2$ to $c2$, where $v2$ is the valence band split off from $v1$ due to the electron spin-orbit magnetic interaction. In this case, only one additional critical point needs to be included, the $E_1(\Lambda) + \Delta_1(\Lambda)$

at 3.17 eV which is described as a superposition of a 2D M_0 transition and a 3D M_1 transition.

(iii) The third set is shown in Fig. 2.35(c) and describes the transitions from $v3$ to $c2$. Here another single weak critical point $E_0(\Gamma) + \Delta_0(\Gamma)$ is present at 1.75 eV which is of the 3D M_0-type. Because these two bands are not parallel for any direction of $\vec{k}$-space, the joint density of states is the lowest of any of the four functions, and as a result no other critical points need to be included.

(iv) The fourth set of contributions is shown in Fig. 2.35(d) and describes the transitions from $v1$ to $c1$. This set is expected to exhibit two critical points, $E_0'(\Gamma)$ and $E_0'(\Lambda)$; however, only the latter is observed experimentally—at 4.48 eV. Thus, the $E_0'(\Gamma)$ critical point is omitted, and the $E_0'(\Lambda)$ critical point is described as a superposition of a 2D M_0 transition and a 3D M_1 transition.

It should be emphasized that some details of these assignments remain controversial; however, the model leads to a physically realistic description of the dielectric function.[54] Using this model, along with the critical point functional dependences in Table 2-2 (with the constants replaced by polynomial functions in general), the individual dependences of $w_j(E')$ are obtained by summing the four contributions described above. This sum is shown in Fig. 2.35(e) and the functional forms are as follows:

Regime v = I:

$$w_{\mathrm{I}}(E') = (E' - E_0)^{1/2}\,[p_{\mathrm{I}}(E') - q_{\mathrm{I}}(E')\,(E_{\mathrm{I}} - E')^{1/2}]; \quad E_0 \le E' < E_{\mathrm{I}}, \tag{2.94a}$$

Regime v = I′:

$$w_{\mathrm{I}'}(E') = [E' - (E_0 + \Delta_0)]^{1/2}\,p_{\mathrm{I}'}(E'); \quad E_0 + \Delta_0 \le E' < E_{\mathrm{I}}, \tag{2.94b}$$

Regime v = II, III, IV:

$$w_v(E') = p_v(E') - q_v(E')\,(E_v - E')^{1/2}; \quad E_{v-\mathrm{I}} \le E' < E_v, \tag{2.94c}$$

Regime v = V:

$$w_{\mathrm{V}}(E') = p_{\mathrm{V}}(E'); \quad E_{\mathrm{IV}} \le E' < E_{\mathrm{V}}, \tag{2.94d}$$

Regime v = VI:

$$w_{\mathrm{VI}}(E') = p_{\mathrm{VI}}(E') - q_{\mathrm{VI}}(E')\,(E' - E_{\mathrm{V}})^{1/2}; \quad E_{\mathrm{V}} \le E' < E_{\mathrm{VI}}. \tag{2.94e}$$

In these expressions, $w_\nu(E)$ vanishes outside the designated ranges and $p_\nu(E')$ and $q_\nu(E')$ are either constants, linear functions, or quadratic functions of E'. In Eqs. (2.94), the notation $w_j(E)$ has been converted to $w_\nu(E)$ in order to be consistent with Fig. 2.35(e) in the original work of Kim et al.[54] The Roman numeral index ν identifies the regimes I, I′, II, … VI in Fig. 2.35(e), and ($E_{\nu-\mathrm{I}}$, E_ν) are the (lower, upper) energy bounds for the segment $w_\nu(E)$ (unless noted otherwise). From Fig. 2.35, the following assignments for the upper energy bounds are made:

$$E_{\mathrm{I}} = E_1(\Lambda), \tag{2.94f}$$

$$E_{\mathrm{II}} = E_1(\Lambda) + \Delta_1(\Lambda), \tag{2.94g}$$

$$E_{\mathrm{III}} = E_0'(\Delta), \tag{2.94h}$$

$$E_{\mathrm{IV}} = E_2(\mathrm{X}), \tag{2.94i}$$

$$E_{\mathrm{V}} = E_2(\Sigma), \tag{2.94j}$$

$$E_{\mathrm{VI}} = E_{\max}. \tag{2.94k}$$

Upon substitution of Eqs. (2.94a–e) into Eq. (2.93), one obtains:

$$\begin{aligned}\varepsilon(E) = 1 + i\frac{2e^2\hbar^2}{\varepsilon_0 m^2}\sum_{n=0}^{2}&[(p_{\mathrm{I}n}H_{\mathrm{I}n} - q_{\mathrm{I}n}F_{\mathrm{I}n}) + (p_{\mathrm{I}'n}H_{\mathrm{I}'n})\\ &+ (p_{\mathrm{II}n}G_{\mathrm{II}n} - q_{\mathrm{II}n}K_{\mathrm{II}n}) + (p_{\mathrm{III}n}G_{\mathrm{III}n} - q_{\mathrm{III}n}K_{\mathrm{III}n})\\ &+ (p_{\mathrm{IV}n}G_{\mathrm{IV}n} - q_{\mathrm{IV}n}K_{\mathrm{IV}n}) + (p_{\mathrm{V}n}G_{\mathrm{V}n}) + (p_{\mathrm{VI}n}G_{\mathrm{VI}n} - q_{\mathrm{VI}n}H_{\mathrm{VI}n})]\\ &+ \sum_n B_n E^n,\end{aligned} \tag{2.95}$$

where the final summation gives the contribution to $\varepsilon_1(E)$ due to critical points above $E_{\max}$.[54] In this equation, $p_{\nu n}$ and $q_{\nu n}$ denote the nth polynomial coefficient ($n = 0, 1, 2$) of p_ν and q_ν and $F_{\nu n}$, $G_{\nu n}$, $H_{\nu n}$, and $K_{\nu n}$ are the following integrals:

$$F_{\mathrm{I}n}(E) = \int_{E_0}^{E_{\mathrm{I}}} E'^{n-2}\sqrt{(E' - E_0)(E_{\mathrm{I}} - E')}\,\Phi_{\mathrm{I}}[E, E', \sigma_{\mathrm{I}}(E')]dE'; \tag{2.96a}$$

$$G_{\nu n}(E) = \int_{E_{\nu-1}}^{E_\nu} E'^{n-2}\Phi_\nu[E, E', \sigma_\nu(E')]dE'; \quad \nu = \mathrm{II},\ \mathrm{III},\ \mathrm{IV},\ \mathrm{V},\ \mathrm{VI}; \tag{2.96b}$$

$$H_{\mathrm{I}n}(E) = \int_{E_0}^{E_{\mathrm{I}}} E'^{n-2} \sqrt{(E' - E_0)}\ \Phi_{\mathrm{I}}[E, E', \sigma_{\mathrm{I}}(E')]dE'; \tag{2.96c}$$

$$H_{\mathrm{I}'n}(E) = \int_{E_0+\Delta_0}^{E_{\mathrm{I}}} E'^{n-2} \sqrt{E' - (E_0 + \Delta_0)}\ \Phi_{\mathrm{I}}[E, E', \sigma_{\mathrm{I}}(E')]dE'; \tag{2.96d}$$

$$H_{\mathrm{VI}n}(E) = \int_{\mathrm{E_V}}^{E_{\mathrm{VI}}} E'^{n-2} \sqrt{(E' - E_{\mathrm{V}})}\ \Phi_{\mathrm{VI}}[E, E', \sigma_{\mathrm{VI}}(E')]dE'; \tag{2.96e}$$

$$K_{\nu n}(E) = \int_{E_{\nu-1}}^{E_\nu} E'^{n-2} \sqrt{(E_\nu - E')}\ \Phi_\nu[E, E', \sigma_\nu(E')]dE'; \quad \nu = \mathrm{II},\ \mathrm{III},\ \mathrm{IV}. \tag{2.96f}$$

These integrals can be solved analytically in the case where $\Phi_\nu[E, E', \sigma_\nu(E)]$ is a Lorentzian broadening function as described in the next paragraph.

Because the segments in Fig. 2.35(e) span the ranges between successive critical points, the Lorentzian broadening parameter $\sigma_\nu(E') = \Gamma'_\nu(E')$ for the segment $w_\nu(E')$ is assumed to be a linear function of energy E' with the value $\Gamma_{\nu-\mathrm{I}}$ for the lower critical point at $E_{\nu-\mathrm{I}}$ and Γ_ν for the upper critical point at E_ν. Thus,

$$\Gamma'_\nu(E') = \frac{\Gamma_\nu - \Gamma_{\nu-\mathrm{I}}}{E_\nu - E_{\nu-\mathrm{I}}} E' + \frac{E_\nu \Gamma_{\nu-\mathrm{I}} - E_{\nu-\mathrm{I}} \Gamma_\nu}{E_\nu - E_{\nu-\mathrm{I}}}. \tag{2.97a}$$

In fact, it has been proposed that a mixture of Gaussian and Lorentzian broadening can be simulated by replacing Γ_ν in Eq. (2.97a) with

$$\Gamma_\nu(L\text{-}G) = \Gamma_\nu \exp\left[-\alpha_\nu \frac{(E' - E_\nu)^2}{\Gamma_\nu^2}\right], \tag{2.97b}$$

and $\Gamma_{\nu-\mathrm{I}}$ by the corresponding expression with the subscript ν replaced by $\nu - \mathrm{I}$. In Eq. (2.97b), α_ν is a Lorentzian-Gaussian mixing parameter for the upper critical point of the segment $w_\nu(E')$. Purely Lorentzian broadening holds for $\alpha_\nu = 0$, and Gaussian broadening is simulated using $\alpha_\nu = 0.2$.

Equations (2.94) through (2.97) completely specify the dielectric function of GaAs in terms of energy-independent free parameters. The maximum number of such parameters required to specify $\varepsilon_2(E)$ is 57. Such a model for $\varepsilon_2(E)$ includes 36 polynomial coefficients if all 12 polynomials (p_ν, q_ν) are of second order, and 21 critical point parameters if all seven critical points require $(E_\nu, \Gamma_\nu, \alpha_\nu)$, i.e., a critical point energy, a broadening

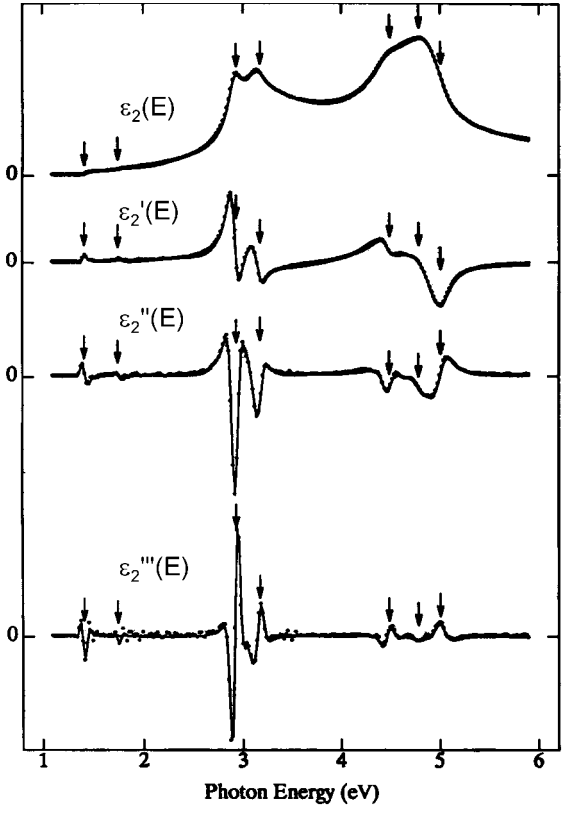


Figure 2.36 Simultaneous fits (lines) to the experimentally determined imaginary part of the dielectric function ε_2 for GaAs and to the first three derivatives of ε_2 (points). The simulation used in the fit applies a specific band structure model [see Table 2-4, Figs. 2.34–2.35, and Eqs. (2.94)–(2.97)] that incorporates 37 energy independent free parameters. These parameters describe the energy positions of the seven critical points in the model (arrrows), their broadening parameters, and the associated polynomial functions. (After Ref. 54.)

parameter, and a mixing parameter. Additional parameters associated with the last term in Eq. (2.95) are required to specify $\varepsilon_1(E)$. In the best-fit analysis of ε_2 and its first three derivatives shown in Fig. 2.36, however, only 37 of the possible 57 free parameters are used owing to the following simplifications.[54] First, all polynomials q_v are assumed to be constants except for q_{I} which is assumed to be quadratic (yielding an overall reduction by 8 parameters). Second, constants are assumed for p_v in regions IV, V, and VI; linear functions are assumed for p_v in regions I′, II, and III; and a quadratic function is retained for p_{I} (yielding an overall

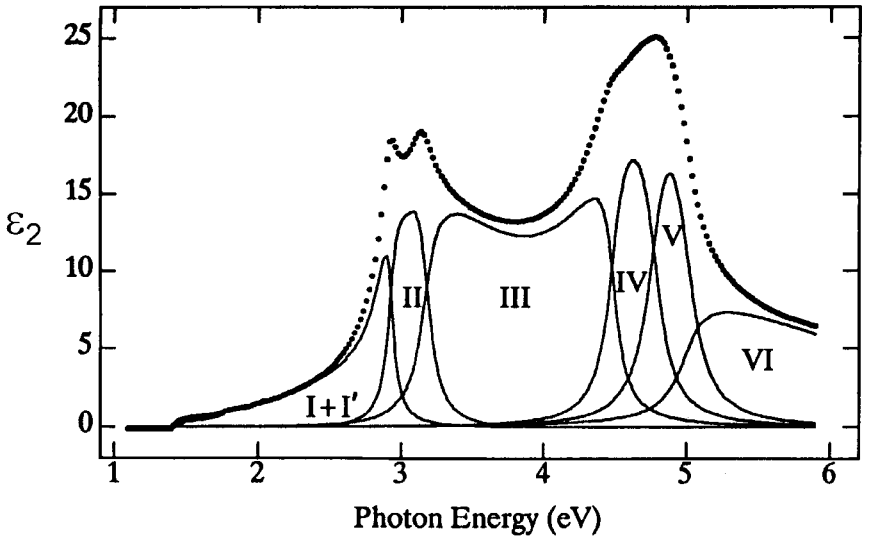


Figure 2.37 Best fit contributions to the imaginary part of the dielectric function ε_2 for GaAs (lines) enumerated according to the scheme of Fig. 2.35 and plotted along with the experimental ε_2 spectrum. The best fit contributions were obtained from a decomposition of the simulation of Fig. 2.36. (After Ref. 54.)

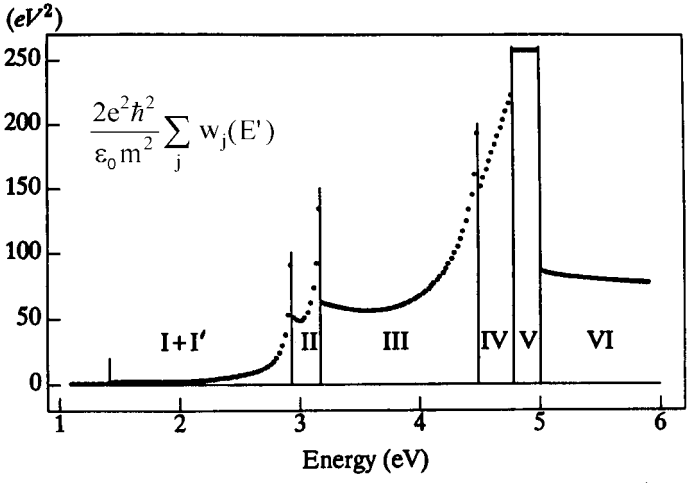


Figure 2.38 The joint density of states function of Eq. (2.93), $(2e^2\hbar^2/\varepsilon_0 m^2)[\Sigma_j w_j(E')]$, obtained from the best fit to the imaginary part of the dielectric function ε_2 of GaAs depicted in Fig. 2.36. (After Ref. 54.)

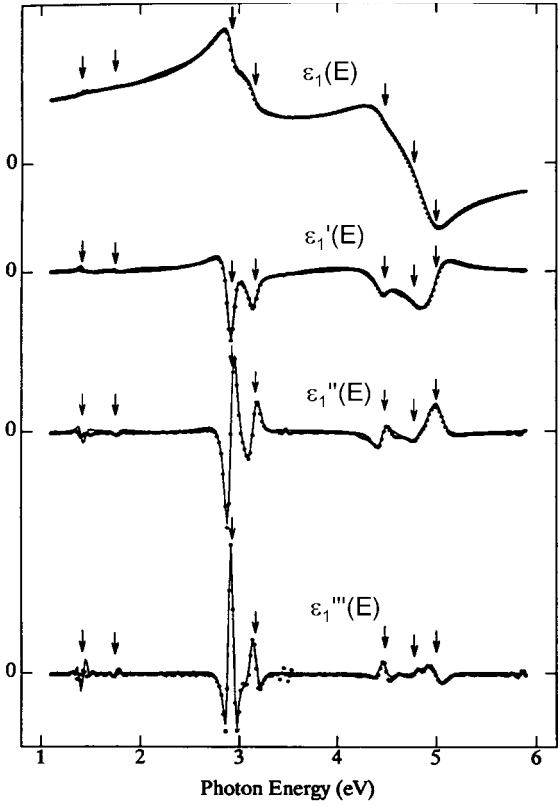


Figure 2.39 Simultaneous fits (lines) to the experimentally determined real part of the dielectric function ε_1 for GaAs and to the first three derivatives of ε_1 (points). The simulation used in the fit applies the same fixed values for the 37 parameters as were determined in the fit of Fig. 2.36. The only free parameters in the fit arise from the last term in Eq. (2.95), which is assumed to be a quadratic function. (After Ref. 54.)

reduction by 9 parameters). Finally, only the lower four critical points are assumed to be a mixture of Gaussian and Lorentzian lineshapes (yielding an overall reduction by 3 parameters). Figure 2.37 depicts the six separate contributions to ε_2 that describe the 37-parameter best fit, and Fig. 2.38 depicts the resulting spectrum in the sum $(2e^2\hbar^2/\varepsilon_0 m^2)\sum_v w_v(E)$.[54] Finally in Fig. 2.39, the spectra in $\varepsilon_1(E)$ and its first three derivatives are shown along with the best fits, obtained using Eqs. (2.94) through (2.97) and the same set of 37 free parameters as in the best fit of $\varepsilon_2(E)$ in Fig. 2.36. In addition, the last term in Eq. (2.95) is assumed to be a quadratic polynomial. As a result, 40 free parameters are used to fit both ε_1 and ε_2.

An interesting aspect about Fig. 2.38 is that downward steps occur at the $E_I = E_1(\Lambda)$, $E_{II} = E_1(\Lambda) + \Delta_1(\Lambda)$, $E_{III} = E_0'(\Delta)$ critical point transitions. From the analytical properties of J_{cv}, these steps should be upward since they arise from 2D M_0-type critical points. This discrepancy suggested to the authors that although the dielectric function model has been designed in an attempt to include all the essential physical processes, some ingredient(s) may be missing.[54] It was pointed out that the direction of the steps can be modified by incorporating an adjustable phase projection factor as in Eq. (2.70), rather than the fixed phases of Eqs. (2.69). An adjustable phase has been incorporated in previous critical point analyses of optical spectra, and best fit values different from theoretically predicted ones have been attributed to excitonic effects as described earlier.[21,22] Other effects such as built-in electric fields at the surface or modifications of the electronic wavefunctions due to the presence of the surface have been proposed as well.[54] Before new physical processes (and possibly additional free parameters) are incorporated, however, it is important as always to consider the confidence limits in the 37 parameters of the best fit. In view of this large number of parameters, it seems plausible that a physically correct result could be obtained while remaining within the confidence limits of the free parameters.

In spite of these uncertainties in the statistical analysis of the best fit (and possibly in the underlying physical processes), the dielectric function model described here provides several advantages over all previous models. First, the model is applicable over the full photon energy range, both above and below the lowest direct gap. Second, the theoretically predicted structure of $J_{cv}(E')$ can be incorporated at each of the critical points (notwithstanding possible limitations of existing theory), in contrast to the generalized critical point model described previously. As a result, the model allows one to fit not only $\varepsilon(E)$ but also its derivatives. Third, the model exactly satisfies the Kramers-Kronig relationships, and also allows one to incorporate Gaussian broadening mechanisms that provide a better description of data for elemental and compound semiconductors at room temperature.

2.5.4.2 Case Study: Aluminum

As described in the previous case study of GaAs, a physics-based approach for modeling dielectric functions of materials can be undertaken when sufficient band structure information is available. The dielectric function of aluminum will be discussed briefly as a second example. Because the final analytical expressions for the real and imaginary parts of the dielectric function of aluminum are quite lengthy, they will not be reproduced here. Instead, references will be made to the appropriate equations in the article by Ashcroft and Sturm where the

derivations appear.[83] In spite of the cumbersome expressions, several different groups have successfully applied them in fits to the optical spectra of aluminum.[47,83–85]

In the near-infrared and visible photon energy range, the imaginary part of the dielectric function of aluminum is dominated by the transitions

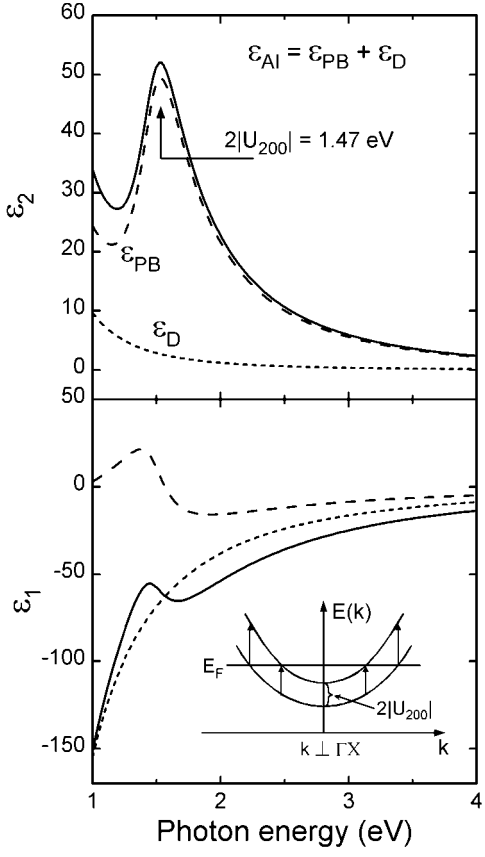


Figure 2.40 Calculated dielectric function (ε_1, ε_2) for Al at room temperature (solid lines), including a decomposition into intraband (dotted lines) and interband (dashed lines) components. For the intraband component, the Drude expression was used with an amplitude and relaxation time deduced from best fits to experimental data (as in Fig. 2.41). For the interband component two parallel-band transitions associated with the (200) and (111) reciprocal space planes were used with energy gaps and broadening parameters also deduced from best fits to data. The interband component is described by a nearly free electron model and is dominated by transitions between parallel bands. For example for the (200) component, the bands are parallel in planes perpendicular to the ΓX line [and thus parallel to the square (200) Brillouin zone face (see inset)]. (After Ref. 76.)

between nearly parallel bands that occur in $\vec{k}$-space planes parallel to the (200) and (111) faces of the Brillouin zone.[45] The (200) transitions (see inset of Fig. 2.40) are especially important in describing the optical properties of Al as they give rise to a maximum in $\varepsilon_2(E)$ and a minimum in the reflectance spectrum of R~0.86 at ~1.5 eV due to the associated absorbance. The (111) transitions are less important as they give rise to a weaker feature near ~0.5 eV, below the photon energy range of conventional SE instruments. Assuming that all other interband transitions are negligible, one can deduce a theoretical expression for the interband dielectric function using a nearly free electron model in which the first-order perturbation due to the periodic potential accounts for the splitting between the two bands.[83] Each of the two so-called "parallel-band" transitions exhibits the following form for the imaginary part of the dielectric function:

$$\varepsilon_{2,PB}(E,\vec{K}) = \frac{\dfrac{D_{\vec{K}}e^2K}{24\pi\varepsilon_0 E}\left|\dfrac{2U_{\vec{K}}}{E}\right|^2}{\sqrt[4]{\left[1-\left(\dfrac{2U_{\vec{K}}}{E}\right)^2-\left(\dfrac{2\Gamma_{m\vec{K}}}{E}\right)^2\right]^2+\left(\dfrac{4\Gamma_{m\vec{K}}}{E}\right)^2}}\left\{\frac{E^2J_{\vec{K}}(E)}{E^2+4\Gamma_{m\vec{K}}^2}\right\}. \tag{2.98}$$

In this equation, (i) $D_{\vec{K}}$ is the degeneracy of the reciprocal space planes, either six-fold for (200) or eight-fold for (111); (ii) K is the length of the reciprocal lattice vectors, either $\vec{K} = (2\pi/a)(2, 0, 0)$ or $(2\pi/a)(1,1,1)$, that define the normals to the reciprocal space planes containing the parallel bands ($a = 0.405$ nm); (iii) $2U_{\vec{K}}$ is the spacing between the parallel bands, and (iv) $2\Gamma_{m\vec{K}} \equiv \hbar/\tau_{\vec{K}}$ where $\tau_{\vec{K}}$ is the relaxation time responsible for the transition broadening. The associated expressions for $J_{\vec{K}}(E)$ and for the real part $\varepsilon_{1,PB}(\vec{K}, E)$ are given by Ashcroft and Sturm as Eqs. (19) and (22) in Ref. 83.

The analytical expression for the complete dielectric function is written as follows

$$\varepsilon(E) = \varepsilon_{0s} + [\varepsilon_D(E) - 1] + \sum_{\vec{K}}[\varepsilon_{PB}(E,\vec{K}) - 1]. \tag{2.99}$$

In this expression, $\varepsilon_D(E)$ represents the Drude or intraband contribution to the dielectric function, and $\varepsilon_{PB}(E, \vec{K})$ represents the two interband or parallel-band transitions identified by the reciprocal lattice vector. Because the three contributions can be considered as

independent responses, the unity contributions to ε_1 from each are subtracted and combined together in a single term given by ε_{0s}, which includes possible interband contributions above the upper spectral limit. The same approach has been applied in Section 2.5.2.2 in an analysis of the intraband and interband transitions in silver (see Fig. 2.18).[36]

As a result, seven free parameters are required in the final analytical expression for the dielectric function of Al. These include

(i) the amplitude of the free-electron component, $A_D^2 \equiv (e^2\hbar^2 N_{ec}/\varepsilon_0 m_o)$, where A_D can be considered a phenomenological plasma energy [see Eq. (2.75a)];
(ii) the relaxation time of the free electrons $\tau_D = \hbar/2\Gamma_m$ [see also Eq. (2.75a)];
(iii) the parallel-band transition energy gap and relaxation time for each of the two components, $(2|U_{200}|, \tau_{200})$ and $(2|U_{111}|, \tau_{111})$; and
(iv) the constant contribution to the dielectric function, ε_{0s}.

Figure 2.40 shows the calculated dielectric function for Al at room temperature (solid lines), using the analytical model of Eq. (2.98) along with the expressions from Ashcroft and Sturm.[83] Also included here is the decomposition of $(\varepsilon_1, \varepsilon_2)$ into intraband (dotted lines) and interband (dashed lines) components. For the intraband component, the following free parameters were employed: $A_D = 12.50$ eV and $\tau_D = 10.6 \times 10^{-15}$ s, determined in best fits to experimental data. For the interband component, the following parameters were employed $2|U_{200}| = 1.47$ eV, $2|U_{111}| = 0.49$ eV, and $\tau_{200} = \tau_{111} = 4.1 \times 10^{-15}$ s, also deduced from best fits to data.[47]

Figure 2.41 shows experimental results obtained by real time SE during the growth of an Al film by thermal evaporation from Nguyen et al. (points),[47] along with data on a static Al film reported by Shiles et al. (lines).[86] The advantage of the real time SE measurements is that they avoid distortion of the optical properties by oxide overlayers that form rapidly on static surfaces, especially upon removing the sample from the vacuum chamber. The agreement between the data sets is quite good in spite of the differences in measurement procedure. The sharply increasing ε_2 spectrum with decreasing E reflects the (200) parallel-band contribution centered near 1.5 eV, as in Fig. 2.40. The sharply decreasing ε_1 spectrum with decreasing E reflects the intraband contribution, with the superposition of the (200) parallel-band contribution leading to the weaker feature. Small differences between the data sets in Fig. 2.41 appear to be due to a sharper (200) parallel-band transition, i.e., a longer electron scattering time τ_{200}, for

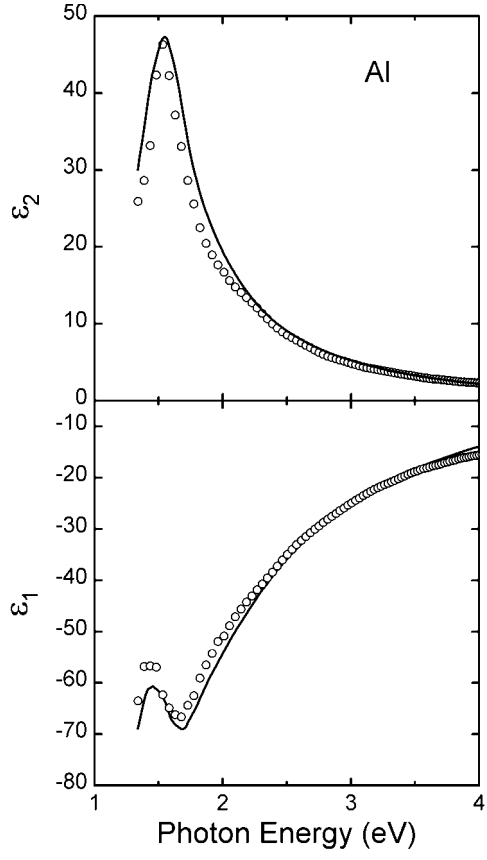


Figure 2.41 Room temperature dielectric functions of thin film Al as reported in Ref. 86 (solid lines) and as obtained by real time spectroscopic ellipsometry during evaporation on a silicon wafer substrate (points). (After Ref. 47.)

the data of Nguyen et al.[47] possibly due to a larger grain size in the Al film.

Similar such differences appear more clearly when characterizing films by real time SE as a function of accumulated film thickness. Figure 2.42 shows the measured ε_2 spectra for aluminum films with thicknesses of 10.6 nm and 75.0 nm, along with fits to a theoretical expression that includes, not only the (200) parallel-band contribution of Eq. (2.98), but also the much weaker intraband contribution (see Fig. 2.40).[47] Good agreement is obtained with the interband transition energies, $E_{PB} \equiv 2|U_{200}|$, and broadening parameters, $\Gamma_{PB} \equiv \Gamma_{m200} = \hbar/(2\tau_{200})$,

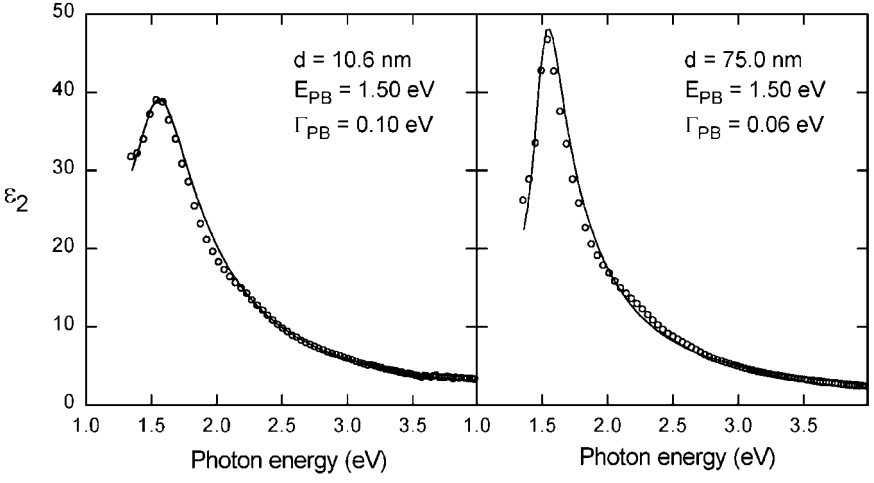


Figure 2.42 Imaginary parts of the dielectric function for aluminum films 10.6 nm (left) and 75.0 nm (right) thick along with the best fit to an analytical formula including both interband and intraband contributions. Free parameters in the fit include the interband (or parallel-band) relaxation time $\Gamma_{PB} = \hbar/2\tau_{PB}$ and the interband energy gap $E_{PB} = |2U_{200}|$, where U_{200} is the Fourier component of the crystal potential for the (200) reciprocal space direction. (Data from Ref. 47).

provided in the figure. The 70% larger broadening parameter for the thinner film in Fig. 2.42 is attributed a smaller grain size. This leads to a reduction in the free carrier relaxation time due to scattering at grain boundaries.

In fact, when grain boundary scattering dominates, the relaxation time for direct interband transitions in metallic films can provide information on the mean free path $L_{\vec{K}}$ of excited electrons through a relationship of the form: $\tau_{\vec{K}}^{-1} = \tau_{b\vec{K}}^{-1} + (v_{\vec{K}}/L_{\vec{K}})$.[47,87] Here $\tau_{\vec{K}}$ is the relaxation time as modified by scattering, $\tau_{b\vec{K}}$ is the relaxation time in the limit of large grain size, and $v_{\vec{K}}$ is the excited electron velocity. Figure 2.43 (solid symbols) shows the thickness dependence of $\tau_{PB} \equiv \tau_{200}$ deduced in fits of ε from real time SE measurements.[47] In these fits, the analytical expression for ε outlined above has been applied, not only in the coalesced film regime, but also in the nucleation regime (i.e., for thicknesses $d < 5$ nm). In the latter regime, a generalized Maxwell-Garnett effective medium theory is applied as well. The mean free path is also shown in Fig. 2.43 (open symbols) as calculated from τ_{PB} assuming $v_{200} = v_F = 2.02 \times 10^8$ cm/s, the Fermi velocity of Al. The solid line is a fit to the thickness dependence of the electron

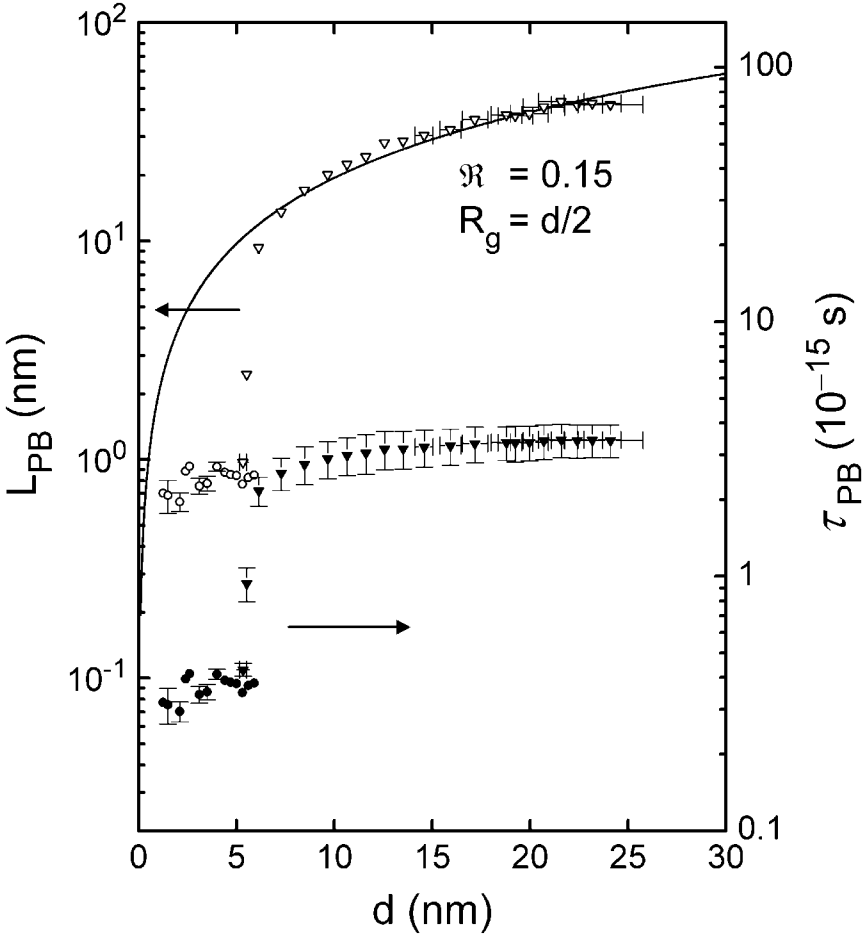


Figure 2.43 Electron mean free path (open points; left scale) and (200) parallel-band relaxation time (filled symbols; right scale) as a function of film thickness d deduced from real time spectroscopic ellipsometry measurements during the growth of aluminum films at low and high rates (circles: 4.3 nm/min; triangles: 28.0 nm/min). The triangles for $d > 5.0$ nm were determined by fitting ellipsometric spectra with a continuous film model. All other results were determined by fitting dielectric functions using a particle film model and the generalized Maxwell-Garnett effective medium theory. The solid line is a fit to L for $d > 6.0$ nm using a grain growth model. For a grain radius of $d/2$, the solid line corresponds to a grain boundary reflection coefficient of 0.15. Error bars are the 90% confidence limits. (After Ref. 47.)

mean free path using a scattering model in which (i) boundaries between spherical grains of radius $R_g = d/2$ limit the mean free path and (ii) these boundaries exhibit a reflection coefficient of 0.15.[88] The former choice is based on the approximate equality between grain size and thickness due to the coarsening of microstructure during growth; the latter result is borrowed from a description of free electrons. The good agreement between the data and theory for the mean free path supports the grain boundary scattering model for thicknesses $d > 5$ nm. For thinner films, isolated particles are present and in this regime scattering appears to be controlled by defects within the Al particles. The thin film coalescence process is clearly evident by the abrupt increase in τ_{PB} and L_{PB}.

2.6 Overview and Concluding Remarks

Sections 2.2 through 2.4 of this chapter have provided a broad overview of the physical principles underlying the optical properties of solid-state materials. In general, such principles were established more than 30 years ago, and so the development in these sections has paralleled the classic text of Wooten.[20] The approach used here differs in significant ways, however, as the emphasis has been placed on derivations that simultaneously provide both the real and imaginary parts of the dielectric function, owing to the interest in applying the outcome in modeling both real and imaginary parts of ellipsometric spectra. In Wooten's text, in contrast, the emphasis was placed on derivations of the imaginary part of the dielectric function, e.g., via Fermi's golden rule for the transition rate, owing to the interest in applying the outcome in modeling reflectance and photoemission spectra. Thus, the central theme throughout this chapter has been the complex dielectric response that fully characterizes the macroscopic, linear response of a solid to a monochromatic light wave, and in particular, the inherent wavelength dependence of this response (i.e., both the absorption and dispersion) for solids over the wavelength range from the near-infrared to the ultraviolet. Over this wavelength range, the interaction of the light wave with the valence and conduction electrons is primarily responsible for the observed wavelength-dependent characteristics.

In Section 2.2, the theoretical foundations began with a detailed presentation of the propagation of light in solids. This presentation progressed in a series of steps, starting with the complex dielectric function that describes wave propagation in an optically isotropic solid, and ending with the complex dielectric and gyration tensors that describe wave propagation in an optically anisotropic, optically active crystal. In Sections 2.3 and 2.4, the presentation also progressed in steps, starting with a simple classical mechanical model for the absorption and dispersion

of weakly interacting atoms, and ending with more sophisticated quantum mechanical models based on the electronic band structure of solids. The former model provides a rudimentary understanding of the differences in the dielectric functions among metals, semiconductors, and insulators; the latter models provide an in-depth understanding of the differences among different materials within the same group. With the more sophisticated models, the wavelength dependence of the dielectric function yields insights into the underlying electronic transitions and the electronic joint density of states in the solid. In fact, comparisons of experimental data for the dielectric function with results calculated from theoretically derived band structures apply the theoretical background presented here and remain a subject of current interest for newly developed material systems.

In Section 2.5 these basic principles have been applied in the development and demonstration of so-called "parameterizations" of the experimentally determined dielectric functions of solids. In a dielectric function parameterization, the experimental results are fit to an analytical expression for the dielectric function yielding best-fit wavelength-independent parameters in the expression that provide information on the electronic processes. Furthermore, the best fit parameters for a series of systematically prepared materials can be expressed as polynomial functions of the material characteristics such as alloy composition, temperature, crystalline content, grain size, doping level, etc. Such polynomial functions can then be used in analyses of raw ellipsometric spectra directly in terms of material characteristics. When analytical expressions are derived from oversimplified physical principles, however, the deduced free parameters tend to carry less meaning or even may be misleading (e.g., an oscillator simulating a smooth background in the dielectric function centered in a region where no critical points exist). As long as such fits reproduce the data within experimental error, such expressions can still be very useful in a materials characterization mode of analysis. Although analytical expressions derived from rigorous physical principles are more likely to provide useful information, in some cases they may uncover deficiencies in existing theories (e.g., incorrect critical point attributions). Finally, throughout Section 2.5, the need for a proper statistical assessment of the deduced best-fit parameters has been emphasized. Correlation coefficients and confidence limits should be calculated to ensure that the significance of the parameter values is not overestimated.

First, in Section 2.5, the Lorentz oscillator and Drude free electron expressions for the optical properties of solids have been applied to model dielectric functions of materials. Such models are found to be most successful in characterizing amorphous, nanocrystalline, and microcrystalline materials, including discontinuous particulate films, owing to the

very broad optical features observed in such materials. However, for an amorphous or nano/microcrystalline semiconductor, the Lorentz oscillator model does not provide a sufficiently fast fall-off in ε_2 with decreasing photon energy in the neighborhood of the band gap and below. Such a problem has been solved using the so called Tauc-Lorentz model and similarly modified Lorentz oscillator models. The first development of such a model by Jellison and Modine,[51] which is Kramers-Kronig consistent by its nature, has been viewed as one of the most important developments over the last decade in modeling the optical responses of disordered solids. A similar problem, i.e., an overestimation of ε_2 below the band gap, must also occur when the Lorentz oscillator model is used to simulate the dielectric function of a single-crystal semiconductor. Furthermore, the Lorentz oscillator model fails to adequately fit the sharp features in the dielectric function due to critical points in the band structure, and it becomes necessary to add unphysical oscillators in the regions of the spectra between these features in order to simulate the broad background in the dielectric response. The overall situation for crystalline materials can be improved by various modifications and quantum mechanical generalizations of the Lorentz oscillator model, e.g., by using generalized oscillators with variable phases and exponents, and/or with suppressed ε_2 at low photon energies. The best solution, however, is to apply the physically correct functional dependences for the joint density of states (JDOS) at the critical points, along with polynomial approximations of the (JDOS)−(matrix-element) product between the critical points. This approach has been applied first by Kim and coworkers[54] and can lead to analytical expressions that can provide fits not only to the experimental dielectric functions, but also to their first, second, and third derivatives. In fact, this accomplishment should be viewed as one of the most important developments over the last decade in modeling the optical properties of crystalline solids.

In addition to the fundamental physical information that the dielectric function can provide on electronic states in solids, it provides indispensible information in technological applications, as well. In general, the dielectric functions allow one to predict the performance of solids in a wide range of technologically important optical components and devices. These include passive components such as mirrors, lenses, windows, retarders, and associated thin film coatings, as well as active components such as photodetectors, photovoltaic cells, solid state light-emitting diodes and lasers, liquid crystal displays, electrochromic materials, etc. Spectroscopic ellipsometry (SE) is the most powerful single technique for measuring the dielectric functions of strongly absorbing solids such as metals and semiconductors, as well as all kinds of thin films. The key to this technique is its ability to obtain both amplitude and phase

information, i.e., the real and imaginary parts of the optical response. As a result, SE avoids the need to perform Kramers-Kronig analysis as in the case of conventional reflectance data, which must be collected well into the vacuum ultraviolet and extrapolated beyond the boundaries of the measurement. In fact, because of the ability of SE to provide better data than was possible with earlier techniques, renewed interest has been stimulated in the optical properties of solids over the last decade.[11,12] New applications have included, for example, the determination of the electronic structure of novel materials such as high temperature superconductors through ex situ spectroscopic ellipsometry studies, and the correlation of microstructural nucleation and coalescence in thin films with their electronic performance through real time spectroscopic ellipsometry studies. More advanced experimental techniques and applications along these lines are expected to be the driving force for future progress on the optical properties of solids.

Acknowledgments

The preparation of this chapter was supported in part by the National Science Foundation under Grant Nos. DMR-9217169, DMR-9820170, and DMR-0137240, and by the National Renewable Energy Laboratory under Subcontract Nos. XAF-8-17619-22 and AAD-9-18668-9. The authors express their appreciation to Dr. Hien V. Nguyen for his contributions to the chapter and to G.E. Jellison, Jr. for fruitful discussions and a careful reading of the manuscript.

2.7 References and Bibliography

2.7.1 Numbered References

1. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, Amsterdam (1977).
2. N. M. Bashara, A. B. Buckman, and A. C. Hall, editors, *Proceedings of the Symposium on Recent Developments in Ellipsometry*, North-Holland, Amsterdam (1969); also published as *Surf. Sci.*, **16**, (1969).
3. J. L. Ord, *Surf. Sci.*, **16**, 155 (1969).
4. B. D. Cahan and R. F. Spanier, *Surf. Sci.*, **16**, 166 (1969).
5. S. N. Jasperson and S. E. Schnatterly, *Rev. Sci. Instrum.*, **40**, 761 (1969).
6. D. E. Aspnes in: *Optical Properties of Solids: New Developments*, edited by B. O. Seraphin, North-Holland, Amsterdam (1976), p. 799.

7. J. A. Woollam, B. Johs, C. M. Herzinger, J. Hilfiker, R. Synowicki, and C. L. Bungay, *Proc. Soc. Photo-Opt. Instrum. Eng.*, **CR 72**, 3 (1999).
8. F. Hottier and J. B. Theeten, *J. Cryst. Growth*, **48**, 644 (1980).
9. M. Schubert, *Thin Solid Films*, **313–314**, 323 (1998).
10. V. A. Yakovlev and E. A. Irene, *J. Electrochem. Soc.*, **139**, 1450 (1992).
11. A. C. Boccara, C. Pickering, and J. Rivory, editors, *Proceedings of the First International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam (1993).
12. R. W. Collins, D. E. Aspnes, and E. A. Irene, editors, *Proceedings of the Second International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam (1998).
13. E. D. Palik, editor, *Handbook of Optical Constants of Solids*, Academic, New York (1985).
14. E. D. Palik, editor, *Handbook of Optical Constants of Solids II*, Academic, New York (1991).
15. D. E. Aspnes, *Proc. Soc. Photo-Opt. Instrum. Eng.*, **276**, 188 (1981).
16. K. Vedam, P. J. McMarr, and J. Narayan, *Appl. Phys. Lett.*, **47**, 339 (1985).
17. P. G. Snyder, M. C. Rost, G. H. Bu-Abbud, J. A. Woollam, and S. A. Alterovitz, *J. Appl. Phys.*, **60**, 3293 (1986).
18. R. W. Collins, B. G. Yacobi, K. M. Jones, and Y. S. Tsuo, *J. Vac. Sci. Technol. A*, **4**, 153 (1986).
19. E. A. Irene, *Thin Solid Films*, **233**, 96 (1993).
20. F. Wooten, *Optical Properties of Solids*, Academic, New York (1972).
21. P. Lautenschlager, M. Garriga, S. Logothetidis, and M. Cardona, *Phys. Rev. B*, **35**, 9174 (1987).
22. P. Lautenschlager, M. Garriga, L. Viña, and M. Cardona, *Phys. Rev. B*, **36**, 4821 (1987).
23. J. I. Pankove, *Optical Processes in Semiconductors*, Dover, New York (1971).
24. D. E. Aspnes and A. A. Studna, *Appl. Phys. Lett.*, **39**, 316 (1981).
25. W. G. Oldham, *Surf. Sci.*, **16**, 97 (1969).
26. J. C. Comfort and F. K. Urban, III, *Thin Solid Films*, **270**, 78 (1995).
27. D. E. Aspnes, A. A. Studna, and E. Kinsbron, *Phys. Rev. B*, **29**, 768 (1984).
28. H. Arwin and D. E. Aspnes, *Thin Solid Films*, **113**, 101 (1984).
29. G. E. Jellison, Jr., *Thin Solid Films*, **234**, 416 (1993).
30. G. E. Jellison, Jr., *Thin Solid Films*, **313–314**, 33 (1998).
31. G. E. Jellison, Jr., *Appl. Opt.*, **30**, 3354 (1991).
32. M. Cardona, *Modulation Spectroscopy*, Academic, New York (1969).
33. M. Erman, J. B. Theeten, N. Vodjdani, and Y. Demay, *J. Vac. Sci. Technol. B*, **1**, 328 (1983).
34. M. Erman and P. M. Frijlink, *Appl. Phys. Lett.*, **43**, 285 (1983).
35. A. R. Heyd, R. W. Collins, K. Vedam, S. S. Bose, and D. L. Miller, *Appl. Phys. Lett.*, **60**, 2776 (1992).
36. R. W. Collins and K. Vedam in: *Encyclopedia of Applied Physics*, Vol. 12, edited by G. L. Trigg, VCH, New York (1995), p. 285.
37. L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Addison-Wesley, Reading MA (1965).

38. M. Born and E. Wolf, *Principles of Optics*, Pergamon, Oxford (1970).
39. J. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, New York (1975).
40. G. N. Ramachandran and S. Ramaseshan, in: *Handbuch der Physik*, Vol. 25/1, edited by S. Flugge, Springer, Berlin (1961), p. 1.
41. A. Yariv and P. Yeh, *Optical Waves in Crystals*, John Wiley & Sons, New York (1984).
42. P. K. L. Drude, *Theory of Optics*, Dover, New York (1965).
43. H. A. Lorentz, *The Theory of Electrons*, Dover, New York (1952).
44. J. M. Ziman, *Principles of the Theory of Solids*, Cambridge University, Cambridge (1972).
45. N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Holt, Rinehart, & Winston, New York (1976).
46. R. M. Dawson, Y. M. Li, M. Gunes, D. Heller, S. Nag, R. W. Collins, C. R. Wronski, M. Bennett, and Y.-M. Li, *Mater. Res. Soc. Symp. Proc.*, **258**, 595 (1992).
47. H. V. Nguyen, I. An, and R. W. Collins, *Phys. Rev. B*, **47**, 394 (1993).
48. I. An, J. Lee, B. Hong, and R. W. Collins, *Thin Solid Films*, **313–314**, 79 (1998).
49. D. E. Aspnes, S. M. Kelso, R. A. Logan, and R. Bhat, *J. Appl. Phys.*, **60**, 754 (1986).
50. A. R. Heyd, I. An, R. W. Collins, Y. Cong, K. Vedam, S. S. Bose, and D. L. Miller, *J. Vac. Sci. Technol. A*, **9**, 810 (1991).
51. G. E. Jellison, Jr., and F. A. Modine, *Appl. Phys. Lett.*, **69**, 371 (1996); **69**, 2137 (1996).
52. A. S. Ferlauto, J. Koh, P. I. Rovira, C. R. Wronski, R. W. Collins, and G. Ganguly, *J. Non-Cryst. Solids*, **266**, 269 (2000).
53. J. Leng, J. Opsal, H. Chu, M. Senko, and D. E. Aspnes, *Thin Solid Films*, **313–314**, 132 (1998).
54. C. C. Kim, J. W. Garland, and P. M. Raccah, *Phys. Rev. B*, **45**, 11749 (1992).
55. B. Johs, C. M. Herzinger, J. H. Dinan, A. Cornfeld, and J. D. Benson, *Thin Solid Films*, **313–314**, 137 (1998).
56. E. Merzbacher, *Quantum Mechanics*, John Wiley & Sons, New York (1970).
57. G. Baym, *Lectures on Quantum Mechanics*, W. A. Benjamin, Reading MA (1973).
58. O. S. Heavens, *Optical Properties of Thin Solid Films*, Dover, New York (1965).
59. P. Yeh, *Optical Waves in Layered Media*, John Wiley & Sons, New York (1988).
60. V. M. Agranovich and V. L. Ginzburg, *Spatial Dispersion in Crystal Optics and the Theory of Excitons*, John Wiley & Sons, London (1966).
61. R. Guenther, *Modern Optics*, John Wiley & Sons, New York (1990).
62. D. Pines, *Elementary Excitations in Solids*, W. A. Benjamin, New York (1963).
63. A. Otto, *Festkoerperprobleme*, **XIV**, 1 (1974).
64. A. Otto, in: *Optical Properties of Solids: New Developments*, edited by B. O. Seraphin, North-Holland, Amsterdam (1976), p. 677.

65. H. Raether, in: *Physics of Thin Films*, Vol. 9, edited by G. Hass, Academic, New York (1977), p. 145.
66. H. Ehrenreich, in: *The Optical Properties of Solids*, edited by J. Tauc, Academic, New York (1966), p. 106.
67. G. D. Cody, in: *Semiconductors and Semimetals*, Vol. 21B, edited by J. I. Pankove, Academic, New York (1984), p. 11.
68. J. Tauc, R. Grigorovici, and A. Vancu, *Phys. Stat. Solidi*, **15**, 627 (1966).
69. D. E. Aspnes, in: *Handbook on Semiconductors*, Vol. 2, edited by M. Balkanski, North-Holland, Amsterdam (1980), p. 109.
70. R. Zallen, *The Physics of Amorphous Solids*, John Wiley & Sons, New York (1983).
71. D. N. Jarrett and L. Ward, *J. Phys. D*, **9**, 1515 (1976).
72. G. A. Niklasson, C. G. Granqvist, and O. Hunderi, *Appl. Opt.*, **20**, 26 (1981).
73. H. Fujiwara, J. Koh, P. I. Rovira, and R. W. Collins, *Phys. Rev. B*, **61**, 10832 (2000).
74. D. E. Aspnes, *J. Opt. Soc. Am.*, **70**, 1275 (1980).
75. P. Drude, *Ann. Phys. Chem. N.F. (Leipzig)*, **36**, 865 (1889).
76. A. S. Ferlauto, G. M. Ferreira, C. Chen, P. I. Rovira, C. R. Wronski, R. W. Collins, X. Deng, and G. Ganguly, in: *Photovoltaics for the 21st Century II*, edited by R. D. McConnell and V. K. Kapur, ECS, Pennington NJ (2001), p. 199.
77. H. V. Nguyen, Y. Lu, S. Kim, M. Wakagi, and R. W. Collins, *Phys. Rev. Lett.*, **74**, 3880 (1995).
78. J. Spanier and K. B. Oldham, *An Atlas of Functions*, Hemisphere, New York (1987), p. 233.
79. A. S. Ferlauto, G. M. Ferreira, C. R. Wronski, and R. W. Collins, in: *Proceedings of the 29th IEEE Photovoltaics Specialists Conference*, May 20–24, 2002, New Orleans LA, IEEE, New York (2002), p. 1076.
80. B. G. Bagley, D. E. Aspnes, A. C. Adams, and C. J. Mogab, *Appl. Phys. Lett.*, **38**, 56 (1981).
81. H. V. Nguyen and R. W. Collins, *Phys. Rev. B*, **47**, 1911 (1993).
82. M. L. Cohen and J. R. Chelikowsky, in: *Handbook on Semiconductors*, Vol. 1, edited by W. Paul, North-Holland, Amsterdam (1982), p. 219.
83. N. W. Ashcroft and K. Sturm, *Phys. Rev. B*, **3**, 1898 (1971).
84. A. G. Matthewson and H. P. Myers, *J. Phys. F*, **2**, 403 (1972).
85. G. A. Niklasson, D. E. Aspnes, and H. G. Craighead, *Phys. Rev. B*, **33**, 5363 (1986).
86. E. Shiles, T. Sasaki, M. Inokuti, and D. Y. Smith, *Phys. Rev. B*, **22**, 1612 (1980).
87. U. Kreibig and C. von Fragstein, *Z. Phys.*, **224**, 307 (1969).
88. A. F. Mayadas and M. Schatzkes, *Phys. Rev. B*, **1**, 1382 (1970).

2.7.2 Bibliography

Classical texts on the optical behavior of electrons in solids:

P. K. L. Drude, *Theory of Optics*, Dover, New York (1965).
H. A. Lorentz, *The Theory of Electrons*, Dover, New York (1952).

Selected background texts on electromagnetic waves and optics:

M. Born and E. Wolf, *Principles of Optics*, Pergamon, Oxford (1970).

R. Guenther, *Modern Optics*, John Wiley & Sons, New York (1990).

J. D. Jackson, *Classical Electrodynamics*, John Wiley & Sons, New York (1975).

L. D. Landau and E. M. Lifshitz, *Electrodynamics of Continuous Media*, Addison-Wesley, Reading MA (1965).

Selected background texts on quantum mechanical principles:

G. Baym, *Lectures on Quantum Mechanics*, W. A. Benjamin, Reading MA (1973).

E. Merzbacher, *Quantum Mechanics*, John Wiley & Sons, New York (1970).

Selected background texts on the theory of solids:

N. W. Ashcroft and N. D. Mermin, *Solid State Physics*, Holt, Rinehart and Winston, New York (1976).

J. C. Phillips, *Bonds and Bands in Semiconductors*, Academic, New York (1973).

J. M. Ziman, *Principles of the Theory of Solids*, Cambridge University, Cambridge (1972).

Selected background texts on crystal physics:

H. J. Juretschke, *Crystal Physics*, W. A. Benjamin, Reading MA (1974).

W. P. Mason, *Crystal Physics of Interaction Processes*, Academic, New York (1966).

J. F. Nye, *Physical Properties of Crystals*, Clarendon, Oxford (1976).

Readings in crystal optics:

G. N. Ramachandran and S. Ramaseshan, in: *Handbuch der Physik*, Vol. 25/1, edited by S. Flugge, Springer, Berlin (1961), p. 1.

G. Szivessy, in: *Handbuch der Physik*, Vol. 20, Springer, Berlin (1928), p. 635.

A. Yariv and P. Yeh, *Optical Waves in Crystals*, John Wiley & Sons, New York (1984).

Texts including modern theories of the optical properties of solids:

F. Bassani and G. Pastori Parravicini, *Electronic States and Optical Transitions in Solids*, Pergamon, Oxford (1975).

D. L. Greenaway and G. Harbeke, *Optical Properties and Band Structure of* Semiconductors, Pergamon, Oxford (1968).

J. N. Hodgson, *Optical Absorption and Dispersion in Solids*, Chapman and Hall, London (1970).

T. S. Moss, *Optical Properties of Semiconductors*, Butterworth, London (1959).

A. V. Sokolov, *Optical Properties of Metals*, American Elsevier, New York (1967).

F. Wooten, *Optical Properties of Solids*, Academic, New York (1972).

Collections of articles devoted to modern topics in the optical properties of solids:

F. Abeles, editor, *Optical Properties and Electronic Structure of Metals and Alloys*, Academic, New York (1966).

F. Abeles, editor, *Optical Properties of Solids*, North-Holland, Amsterdam (1972).
S. Nudelman and S. S. Mitra, editors, *Optical Properties of Solids*, Plenum, New York (1969).
B. O. Seraphin, editor, *Optical Properties of Solids: New Developments*, North-Holland, Amsterdam (1976).
J. Tauc, editor, *The Optical Properties of Solids*, Academic, New York (1966).

Some sources of data on the optical properties of solids:
D. E. Aspnes and A. A Studna, *Phys. Rev. B*, **27**, 985 (1983).
E. D. Palik, editor, *Handbook of Optical Constants of Solids*, Academic, New York (1985).
E. D. Palik, editor, *Handbook of Optical Constants of Solids II*, Academic, New York (1991).

Texts devoted to thin film optics:
O. S. Heavens, *Optical Properties of Thin Solid Films*, Dover, New York (1965).
H. A. Macleod, *Thin Film Optical Filters*, Adam Hilger, Bristol (1986).
A. Vasicek, *Optics of Thin Films*, North-Holland, Amsterdam (1960).
P. Yeh, *Optical Waves in Layered Media*, John Wiley & Sons, New York (1988).

3 Data Analysis for Spectroscopic Ellipsometry

Gerald E. Jellison, Jr.

Solid State Division, Oak Ridge National Laboratory

3.1 Introduction

Spectroscopic ellipsometry (SE) is a very powerful tool for characterization of thin film systems. Under appropriate circumstances, SE determines film thicknesses more accurately than any other known technique. In addition, SE measurements can provide information concerning the optical functions, surface roughness, and interface layers of films. SE is also a very useful tool in measuring optical functions of bulk materials, and it is often the best available technique for the measurement of bulk optical functions for photon energies greater than the band edge of a semiconductor or insulator. However, the measured data from an SE experiment are usually not very interesting by themselves. The useful information, such as film thicknesses and optical functions, can only be determined by modeling the near-surface region of the sample and then fitting the SE data to the model using the desired parameters as variables in the numerical analysis. Obviously, the way in which the data analysis is performed is critical, and inappropriate modeling of the SE results can often lead to worthless results.

One of the first issues that must be addressed in SE data analysis is the data itself. There are several different kinds of spectroscopic ellipsometers, and many of these instruments measure quite different quantities. Moreover, it is often common practice to transform the measured data into another representation before performing the data analysis, where the fitted data representation is nonlinearly related to the measured data. Of course, SE data, like all experimental results, have errors. It is usually quite easy to deal with the errors associated with random processes, which can be treated using common error analysis techniques. Ideally, all the errors in an SE measurement will be random. However, some of the error associated with SE data is due to alignment errors of the various optical elements, finite band widths associated with the wavelength discriminating device, angle of incidence, gains in the amplifiers, etc. These errors are systematic, and cannot be rigorously treated in the same way as random

errors. Unfortunately, systematic errors can be the most important errors in some experiments, and therefore must be estimated and incorporated into the data analysis in some way. Of course, after the random and systematic error limits of the data have been estimated, it is necessary to propagate these errors into the actual data being fit.

The proper way to analyze complicated experimental spectra, such as results from SE measurements, has been understood for several decades by workers in many fields of science and engineering. The major element of the proper statistical analysis is the incorporation of reasonable error limits into the figure of merit function used in the optimization algorithms. In many cases, ignoring the errors may not result in a serious degradation of the results, but in other cases, ignoring errors can lead to totally wrong interpretations of SE data. Since SE measurements do not measure the quantities of interest, it is extremely important to use every available tool in the analysis of the data to obtain these quantities. Moreover, proper statistical analysis can result in "goodness of fit" information, where a statistical figure of merit is obtained that states whether or not a particular model actually fits the experimental data. If errors are not included in the data analysis or if improper data analysis is used, one never knows whether or not the calculated parameters actually fit the data!

In this chapter, the process of the analysis of spectroscopic ellipsometry data will be described. Because of the nature of this subject matter, there will be some overlap with other chapters in this book. However, the approach taken here will be to present a complete picture of the data analysis, briefly covering all aspects of the process, referring the reader to the other chapters for more details. This approach also gives a consistency to the notation that is only possible by covering the entire subject matter in one chapter, even at the expense of some redundancy. Much of the material presented here is well known, and can be found in standard texts, many of which are listed in the Further Reading section (Section 3.8).

Section 3.2 will cover the different ways in which ellipsometry parameters are represented and the theoretical differences between the parameters that are measured and the parameters that are calculated. Much of this material will also be in Chapter 1, and has been discussed in the standard texts by Azzam and Bashara and Tompkins. Section 3.3 will cover the methods used to calculate the complex reflection coefficients for an arbitrary layer structure. This is well known, and the reader is referred to standard textbooks on optics such as Born and Wolf. The calculation of the complex reflection coefficients from anisotropic materials is also discussed in Section 3.3. This is also the subject of Chapter 9 (Schubert), so the reader is referred to this chapter for more details. Section 3.4 deals with some of the ways in which the optical functions of materials are parameterized in order to use fitting algorithms on SE data. Not all models or

parameterizations are included in this section, but rather representative models are discussed. Again, much of this material will also be discussed in more detail in Chapter 2. Section 3.5 discusses the actual fitting procedure, including the selection of a figure of merit function and the Levenberg-Marquardt algorithm, for situations where the problem is over-determined. These numerical analysis techniques are several decades old, but many people in the spectroscopic ellipsometry field have been slow to adopt these proven mathematical techniques. The interested reader is referred to several well-known texts such as chapter 15 of *Numerical Recipes* for a well-written and sometimes humorous discussion, and *Non-Linear Parameter Estimation* by Bard for a more scholarly discussion. Section 3.6 will discuss data analysis techniques that are commonly used to determine optical functions from SE data where the problem is not over-determined. At first appearance, there might be significant overlap between the subject matter of this section and Section 3.5, but the numerical analysis is fundamentally different. In many spectroscopic ellipsometry experiments, the sample can depolarize the light beam; these effects will be discussed in Section 3.7. Suggested texts for further reading and the numbered references are listed in Section 3.8.

3.2 Ellipsometry Parameters

There are several different ways of performing SE measurements, and several of the techniques measure quite different parameters. For example, rotating analyzer or rotating polarizer ellipsometers measure quite different parameters than do ellipsometers based on photoelastic modulators. Standard nulling ellipsometers measure yet another set of parameters. All of these measured quantities are related to one another and to standard data representations through nonlinear relationships.

Ellipsometric calculations are usually based on models of ideal (that is, non-depolarizing) surfaces with planar interfaces using Maxwell's equations. In general, light reflecting from a sample surface will result in a loss of intensity and will undergo a phase change. Therefore, the normal parameters calculated for reflection ellipsometry are complex reflection coefficients $r = |r|e^{i\delta}$, where δ is the phase shift upon reflection. Since most materials examined in an ellipsometric experiment are nonmagnetic, it is common practice to suppress the magnetic field component of the electromagnetic wave, focusing on the electric field vector.

It is important to recognize, at least on a theoretical basis, that the parameters that result from ellipsometry measurements are different from the parameters that are calculated. Ellipsometry measurements are performed on real samples, which can depolarize the incident light beam,

while most ellipsometry calculations assume that the sample does not depolarize the incident light beam. If sample depolarization is significant, then it is not possible to relate the measured and calculated quantities in a rigorous way. Fortunately, most ellipsometry samples do not significantly depolarize the incident light beam. However, there are many real situations where sample depolarization is an issue and this depolarization must be taken into account, even if only approximately.

3.2.1 Calculated Parameters: Jones Matrices

Figure 3.1 shows a schematic diagram of the reflection of light from an ideal sample surface. The incident light beam is at an angle ϕ with respect to the normal of the sample surface. The plane of incidence is then defined as the plane that contains the incident light beam and is normal to the sample surface. Most ellipsometry experiments are carried out with the specular reflected beam, which comes off the sample surface at the same angle ϕ with respect to the sample surface.

The incident light beam at a point in space is conveniently represented by the electric field vector given by

$$E^i(t) = E^i_p \exp[i(\omega t + \varphi^i_p)]\,\mathbf{p} + E^i_s \exp[i(\omega t + \varphi^i_s)]\,\mathbf{s} = \exp(i\omega t)[\hat{E}^i_p \mathbf{p} + \hat{E}^i_s \mathbf{s}]. \tag{3.1}$$

It is assumed that the light beam is monochromatic, and the **p** and **s** directions are defined to be parallel and perpendicular to the plane of incidence,

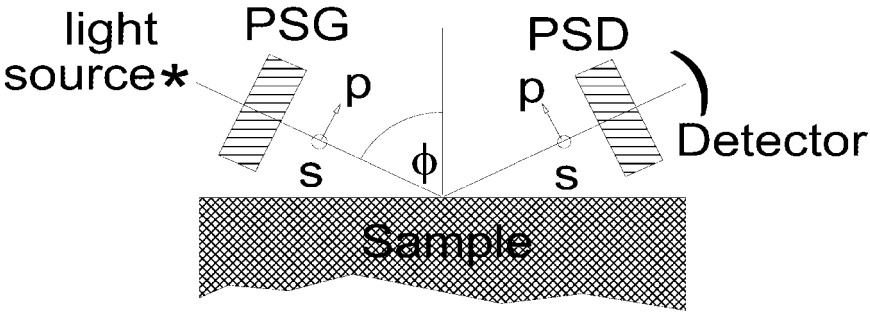


Figure 3.1 Schematic diagram of an ellipsometry experiment. The light beam passes through the polarization state generator (PSG) before being reflected from the sample surface. After reflection, the light beam is repolarized by the polarization state detector (PSD) before being detected. Light polarized in the plane of incidence is p-polarized light, while light polarized perpendicular to the plane of incidence is s-polarized light (out of the page). The angle ϕ is the angle of incidence.

respectively. The frequency of the light beam is $\omega/2\pi$. Similarly, the reflected beam is described by the complex quantities $\hat{E}_p^o$ and $\hat{E}_s^o$.

In general, four complex reflection coefficients are required to describe reflection from an ideal surface. This is easily expressed by the Jones matrix

$$\mathbf{J} = \begin{pmatrix} r_{pp} & r_{ps} \\ r_{sp} & r_{ss} \end{pmatrix} = \begin{pmatrix} \hat{E}_p^o/\hat{E}_p^i & \hat{E}_s^o/\hat{E}_p^i \\ \hat{E}_p^o/\hat{E}_s^i & \hat{E}_s^o/\hat{E}_s^i \end{pmatrix} \tag{3.2}$$

$$= r_{ss} \begin{pmatrix} \rho & \rho_{ps} \\ \rho_{sp} & 1 \end{pmatrix} = \begin{pmatrix} \gamma e^{i\Delta} & \gamma_{ps} e^{i\Delta_{ps}} \\ \gamma_{sp} e^{i\Delta_{sp}} & 1 \end{pmatrix} = \begin{pmatrix} \tan(\psi) e^{i\Delta} & \tan(\psi_{ps}) e^{i\Delta_{ps}} \\ \tan(\psi_{sp}) e^{i\Delta_{sp}} & 1 \end{pmatrix}.$$

The elements of the normalized Jones matrix will be called the ρ-representation.

Not all eight parameters of **J** are of interest in an ellipsometric measurement. The absolute phase is usually not measurable, so only three of the phases are independent. Most ellipsometry measurements do not measure the actual intensity of the light beam, so it is reasonable to ratio the calculations, where r_{ss} is normally chosen as the denominator. Therefore, six parameters are required to define the ellipsometric response from a sample. The sample reflectivity can be a seventh parameter and it is given by

$$R = (r_{pp} r_{pp}^* + r_{ps} r_{ps}^* + r_{sp} r_{sp}^* + r_{ss} r_{ss}^*)/2 \tag{3.3}$$

The expressions in Eqs. (3.2) and (3.3) include the cross-polarization reflection coefficients r_{ps} and r_{sp}. As can be seen from Eq. (3.2), they correspond to $\mathbf{s} \rightarrow \mathbf{p}$ and to $\mathbf{p} \rightarrow \mathbf{s}$ transformations. These parameters are normally equal to zero for isotropic samples, but can be significantly different from zero for anisotropic samples.

3.2.2 Measured Parameters: Mueller Matrices

As mentioned above, real samples can depolarize the incident light beam, so a more general way is needed to treat ellipsometry measurements. The traditional way is to use the Mueller-Stokes formalism, which includes light depolarization. From Fig. 3.1, the source light first passes through a polarization state generator (PSG), which sets the polarization state for the light beam incident upon the sample. In reflecting off the sample, the polarization state of the light beam is changed. The reflected light is then re-polarized and detected by the polarization state detector (PSD).

The polarization state for any light beam in the Mueller-Stokes formalism is given by its Stokes vector

$$\mathbf{S} = \begin{pmatrix} I_o \\ Q \\ U \\ V \end{pmatrix} = \begin{pmatrix} I_o \\ I_0 - I_{90} \\ I_{45} - I_{-45} \\ I_{rc} - I_{lc} \end{pmatrix}, \tag{3.4a}$$

where I_o is the intensity of the light beam, and I_0, I_{45}, I_{90}, and I_{-45} are the light intensities for linearly polarized light at 0°, 45°, 90°, and –45° respectively with respect to the plane of incidence. I_{rc} and I_{lc} are the intensities of right- and left-circularly polarized light, respectively. All elements of the Stokes vector are intensities and therefore are real. The total light intensity

$$I_0 \geq \sqrt{Q^2 + U^2 + V^2}, \tag{3.4b}$$

where the equality holds only if the light beam is totally polarized.

The intensity of the light beam incident upon the detector in any ellipsometry experiment is given by

$$\mathrm{I} = \mathbf{S_{PSD}}^{\mathrm{T}} \mathbf{M}\, \mathbf{S_{PSG}}, \tag{3.5}$$

where $\mathbf{S_{PSG}}$ is the column Stokes vector representing the polarization state of the incident light beam and $\mathbf{S_{PSD}}^{\mathrm{T}}$ is the transposed Stokes vector representing the effects of the optics in the polarization state detector. The 4×4 matrix $\mathbf{M}$ (called the sample Mueller matrix) describes the polarization effects of light reflecting off the sample surface. Of course, not all of the elements of $\mathbf{M}$ are independent since the sample cannot "overpolarize" the light beam, nor can the light energy of the reflected beam be more than the energy of the incident beam. These conditions lead to several restrictions on the individual elements of $\mathbf{M}$, which are described in detail in Brousseau's book *Polarized Light*. The sample Mueller matrix can be normalized with respect to the m_{11} component resulting in

$$\mathbf{M} = \begin{pmatrix} 1 & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{pmatrix}. \tag{3.6}$$

3.2.3 Mueller-Jones Matrices

In order to correlate reflection coefficients with ellipsometry measurements, it is usually assumed that the sample does not depolarize the

incident light beam. If this is the case, then one can define a Mueller-Jones matrix

$$\mathbf{M} = \mathbf{A}\,(\mathbf{J} \otimes \mathbf{J}^*)\mathbf{A}^{-1}, \tag{3.7a}$$

where the matrix **A** is given by

$$A = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & i & -i & 0 \end{pmatrix}. \tag{3.7b}$$

The symbol $\otimes$ indicates the Kroneker product of the sample Jones matrix **J** and its complex conjugate **J***. Since the normalized Jones matrix has six independent parameters, then the normalized sample Mueller-Jones matrix must also have only six independent parameters; this leads to nine bilinear relationships between the various elements of the sample Mueller[1] matrix:

$$m_{12} - m_{21}m_{22} - m_{31}m_{32} - m_{41}m_{42} = 0 \tag{3.8a}$$

$$m_{13} - m_{21}m_{23} - m_{31}m_{33} - m_{41}m_{43} = 0 \tag{3.8b}$$

$$m_{14} - m_{21}m_{24} - m_{31}m_{34} - m_{41}m_{44} = 0 \tag{3.8c}$$

$$m_{12}m_{13} - m_{22}m_{23} - m_{32}m_{33} - m_{42}m_{43} = 0 \tag{3.8d}$$

$$m_{12}m_{14} - m_{22}m_{24} - m_{32}m_{34} - m_{42}m_{44} = 0 \tag{3.8e}$$

$$m_{13}m_{14} - m_{23}m_{24} - m_{33}m_{34} - m_{43}m_{44} = 0 \tag{3.8f}$$

$$1 + m_{12}^2 - m_{22}^2 - m_{32}^2 - m_{42}^2 - m_{21}^2 - m_{31}^2 - m_{41}^2 = 0 \tag{3.8g}$$

$$1 + m_{13}^2 - m_{23}^2 - m_{33}^2 - m_{43}^2 - m_{21}^2 - m_{31}^2 - m_{41}^2 = 0 \tag{3.8h}$$

$$1 + m_{14}^2 - m_{24}^2 - m_{34}^2 - m_{44}^2 - m_{21}^2 - m_{31}^2 - m_{41}^2 = 0 \tag{3.8i}$$

If the sample *does* depolarize the incident light beam, then there is no rigorous connection between the elements of the sample Jones matrix **J** determined from calculations and the experimentally determined sample Mueller matrix **M**.

If the sample is isotropic, then its Mueller matrix is particularly simple, and is given by

$$M = \begin{pmatrix} 1 & -N & 0 & 0 \\ -N & 1 & 0 & 0 \\ 0 & 0 & C & S \\ 0 & 0 & -S & C \end{pmatrix}, \tag{3.9a}$$

where the quantities N, S, and C are given by

$$N = \cos 2\psi, \tag{3.9b}$$

$$S = \sin 2\psi \sin \Delta, \tag{3.9c}$$

$$C = \sin 2\psi \cos \Delta. \tag{3.9d}$$

N, S, and C are not independent and are constrained by the relation

$$N^2 + S^2 + C^2 = 1. \tag{3.9e}$$

The diagonal element of the Jones matrix is then given by

$$\rho = \frac{r_p}{r_s} = \frac{C + iS}{1 + N} = \tan(\psi) e^{i\Delta}, \tag{3.10}$$

and the cross-polarization ratios ρ_{ps} and ρ_{sp} are 0.

If the sample is anisotropic, then the sample Mueller matrix in the NSC representation[2] is given by

$$\mathbf{M} = \begin{pmatrix} 1 & -N - \alpha_{ps} & C_{sp} + \zeta_1 & S_{sp} + \zeta_2 \\ -N - \alpha_{sp} & 1 - \alpha_{sp} - \alpha_{ps} & -C_{sp} + \zeta_1 & -S_{sp} + \zeta_2 \\ C_{ps} + \xi_1 & -C_{ps} + \xi_1 & C + \beta_1 & S + \beta_2 \\ -S_{ps} + \xi_2 & S_{ps} + \xi_2 & -S + \beta_2 & C - \beta_1 \end{pmatrix}, \tag{3.11}$$

where

$$N = (1 - \gamma^2 - \gamma_{sp}{}^2 - \gamma_{ps}{}^2)/D \tag{3.12a}$$

$$D = (1 + \gamma^2 + \gamma_{sp}{}^2 + \gamma_{ps}{}^2) = 2/(1 + N) \tag{3.12b}$$

$$S = 2\,\gamma \sin(\Delta)/D \qquad C = 2\,\gamma \cos(\Delta)/D \tag{3.12c, d}$$

$$S_{sp} = 2\,\gamma_{sp} \sin(\Delta_{sp})/D \qquad C_{sp} = 2\,\gamma_{sp} \cos(\Delta_{sp})/D \tag{3.12e, f}$$

$$S_{ps} = 2\,\gamma_{ps} \sin(\Delta_{ps})/D \qquad C_{ps} = 2\,\gamma_{ps} \cos(\Delta_{ps})/D \tag{3.12g, h}$$

$$\alpha_{sp} = 2\,\gamma_{sp}{}^2/D \qquad \alpha_{ps} = 2\,\gamma_{ps}{}^2/D \tag{3.12i, j}$$

$$\beta_1 = (D/2)(C_{sp}\,C_{ps} + S_{sp}\,S_{ps}) \qquad \beta_2 = -(D/2)(C_{sp}\,S_{ps} - S_{sp}\,C_{ps}) \tag{3.12k, l}$$

$$\zeta_1 = (D/2)(C\,C_{ps} + S\,S_{ps}) \qquad \zeta_2 = -(D/2)(C\,S_{ps} - S\,C_{ps}) \tag{3.12m, n}$$

$$\xi_1 = (D/2)(C\,C_{sp} + S\,S_{sp}) \qquad \xi_2 = (D/2)(C\,S_{sp} - S\,C_{sp}). \tag{3.12o, p}$$

The equivalent cross-polarization coefficients in the ρ representation are given by

$$\rho_{ps} = \frac{C_{ps} + iS_{ps}}{1+N} = \tan(\psi_{ps})e^{i\Delta_{ps}}, \quad \rho_{sp} = \frac{C_{sp} + iS_{sp}}{1+N} = \tan(\psi_{sp})e^{i\Delta}. \quad (3.13a,b)$$

Three representations of the ellipsometry data have been presented. The traditional representation is the $\psi\Delta$ representation, which uses the ellipsometric angles ψ and Δ. Only two angles are required for isotropic samples, but the extension to anisotropic samples is straightforward. The ρ representation for isotropic samples is defined in Eq. (3.10) with the anisotropic extensions given above, and is closely related to the $\psi\Delta$ representation [see Eq. (3.2)]. One complex number is sufficient for isotropic samples, but two additional complex numbers are required to completely describe light reflection from an anisotropic sample. The ρ representation uses complex numbers instead of angles, so there is no artificial discontinuity in the spectra when Δ goes through 0° or 180°. However, discontinuities in the ρ representation of the data can occur if ψ approaches 90°. As already discussed, the NSC representation for isotropic samples requires three parameters, constrained by Eq. (3.9e). The NSC representation for anisotropic samples requires seven parameters, constrained by the relation:

$$N^2 + S^2 + C^2 + S_{ps}^2 + C_{ps}^2 + S_{sp}^2 + C_{sp}^2 = 1. \quad (3.14)$$

All parameters in the NSC representation are bounded ±1. One advantage of the NSC representation is that it can represent simple depolarization, where the = sign in Eqs. (3.9e) and (3.14) is replaced by a < sign. This will be discussed in Section 3.7.

For the purposes of this chapter, the ρ representation will be used unless otherwise stated. It will be assumed that experimental error limits have been determined (both random and systematic), and have been propagated into the ρ representation. Obviously, other representations can be used in any fitting procedure, provided error limits are properly propagated into the new representation.

Often data taken with rotating polarizer or rotating analyzer ellipsometers express the data in terms of $\tan\psi$ and $\cos\Delta$. These parameters are actually derived from the actual experimental quantities, which are Fourier coefficients. Clearly, this representation can also be transformed to the $\psi\Delta$, ρ or the NSC representations. The propagation of errors is particularly important in this case, since $\cos\Delta$ is very insensitive to Δ when Δ is close to 0° or 180°.

3.3 Calculation of Complex Reflection Coefficients

3.3.1 Isotropic, Homogeneous Systems

The complex reflection coefficients shown in Eq. (3.2) can be calculated for many very complicated layer structures. The simplest case is that for a single interface where both media are isotropic. The coefficients are the Fresnel reflection coefficients and are given by

$$r_{pp} = \frac{\tilde{n}_1\cos(\phi_0) - \tilde{n}_0\cos(\phi_1)}{\tilde{n}_1\cos(\phi_0) + \tilde{n}_0\cos(\phi_1)}, \quad r_{ss} = \frac{\tilde{n}_0\cos(\phi_0) - \tilde{n}_1\cos(\phi_1)}{\tilde{n}_0\cos(\phi_0) + \tilde{n}_1\cos(\phi_1)}. \qquad (3.15a, b)$$

The quantities $\tilde{n}_0$ and $\tilde{n}_1$ are the complex refractive indices for the upper region and the substrate respectively. The complex refractive index is defined as $\tilde{n} = n - ik$. The angles ϕ_0 and ϕ_1 are the angles of the light propagation within the specific medium as defined by Snell's law:

$$\xi = \tilde{n}_0\sin\phi_0 = \tilde{n}_1\sin\phi_1. \qquad (3.16)$$

If the sample consists of a single film, then the complex reflection coefficients are given by the Airy formula:

$$r_{pp,ss} = \frac{r_{1,pp,ss} + r_{2,pp,ss}\exp(-2ib)}{1 + r_{1,pp,ss}r_{2,pp,ss}\exp(-2ib)}, \qquad (3.17a)$$

$$b = \frac{2\pi d_f}{\lambda}\tilde{n}_f\cos(\phi_f) \qquad (3.17b)$$

where d_f is the film thickness, λ is the wavelength of light, $\tilde{n}_f$ is the complex refractive index of the film and ϕ_f is the angle of light propagation within the film as determined by Snell's law. The quantities r_1 and r_2 are the Fresnel reflection coefficients for interfaces 1 and 2, respectively. As indicated, Eqs. (3.17a) and (b) are valid for both the pp and ss reflection coefficients.

For multiple isotropic layers, the calculation of the complex reflection coefficients is more complicated. Most of these calculations are performed using a matrix representation, where each layer is represented by two 2×2 complex matrices, one for the pp polarization and the other for the ss polarization. The Abelés method[3] is used extensively in the analysis of ellipsometry data, where the j^{th} layer is

represented by two transfer matrices

$$\mathbf{P}_{j,pp} = \begin{pmatrix} \cos(b_j) & -i\dfrac{\cos(\phi_j)}{\tilde{n}_j}\sin(b_j) \\ i\dfrac{\tilde{n}_j}{\cos(\phi_j)}\sin(b_j) & \cos(b_j) \end{pmatrix}, \tag{3.18a}$$

$$\mathbf{P}_{j,ss} = \begin{pmatrix} \cos(b_j) & i\dfrac{\sin(b_j)}{\tilde{n}_j\cos(\phi_j)} \\ i\tilde{n}_j\cos(\phi_j)\sin(b_j) & \cos(b_j) \end{pmatrix}, \tag{3.18b}$$

where b_j is the phase factor for the j^{th} layer given in Eq. (3.17b), and ϕ_j is the complex angle in the j^{th} layer as given by Snell's law [Eq. (3.16)]. If a film is extremely thin (where $b_j \ll 1$), then its Abelés matrices simplify to

$$\mathbf{P}_{j,pp} = \begin{pmatrix} 1 & -i\dfrac{\cos(\phi_j)}{\tilde{n}_j}b_j \\ i\dfrac{\tilde{n}_j}{\cos(\phi_j)}b_j & 1 \end{pmatrix}, \tag{3.18c}$$

$$\mathbf{P}_{j,ss} = \begin{pmatrix} 1 & i\dfrac{b_j}{\tilde{n}_j\cos(\phi_j)} \\ i\tilde{n}_j\cos(\phi_j)b_j & 1 \end{pmatrix}. \tag{3.18d}$$

The characteristic matrix for the layer stack consisting of N films is determined by matrix multiplication:

$$\mathbf{M}_{pp} = \chi_{0,pp}\left(\prod_{j=1}^{N}\mathbf{P}_{j,pp}\right)\chi_{sub,pp} \quad \text{and} \quad \mathbf{M}_{ss} = \chi_{0,ss}\left(\prod_{j=1}^{N}\mathbf{P}_{j,ss}\right)\chi_{sub,ss}. \tag{3.19a, b}$$

The χ_0 and χ_{sub} matrices are the characteristic matrices for the ambient and the substrate, respectively, and are given by

$$\chi_{0,pp} = \frac{1}{2}\begin{pmatrix} 1 & \dfrac{\cos(\phi)}{\tilde{n}_0} \\ -1 & \dfrac{\cos(\phi)}{\tilde{n}_0} \end{pmatrix}, \qquad \chi_{0,ss} = \frac{1}{2}\begin{pmatrix} 1 & \dfrac{1}{\tilde{n}_0\cos(\phi)} \\ 1 & \dfrac{-1}{\tilde{n}_0\cos(\phi)} \end{pmatrix} \tag{3.20a, b}$$

$$\chi_{sub,pp} = \begin{pmatrix} \dfrac{\cos(\phi_{sub})}{\tilde{n}_{sub}} & 0 \\ 1 & 0 \end{pmatrix} \qquad \chi_{sub,ss} = \begin{pmatrix} \dfrac{1}{\tilde{n}_{sub}\cos(\phi_{sub})} & 0 \\ 1 & 0 \end{pmatrix} \tag{3.20c, d}$$

The complex reflection coefficients are then calculated from the elements of the characteristic matrices using the following relations:

$$r_{pp} = \frac{M_{21,pp}}{M_{11,pp}} \quad \text{and} \quad r_{ss} = \frac{M_{21,ss}}{M_{11,ss}} \tag{3.21a, b}$$

If the film stack consists of two very thin films [see Eqs. (3.18c) and (d)], then their representative Abelés matrices commute to first order in the parameter b_j. Practically, it is not possible to use optical techniques to distinguish between stacks of very thin films, where film combination AB looks just like film combination BA. Even though SE is very sensitive to the thicknesses of very thin films, it is extremely difficult for ellipsometry (even spectroscopic ellipsometry) to measure thicknesses and optical functions of films in the limit of small thicknesses. To increase the sensitivity, one must use shorter wavelengths so that the b parameter is not small and second order effects become important.

3.3.2 Anisotropic Systems

If one or more of the media in the film stack is anisotropic, then the analysis given above in Section 3.3.1 is not sufficiently general. For isotropic systems, the s- and p-polarizations are eigenmodes of the reflection. For systems that contain anisotropic media, they are not. A similar matrix-based approach discussed above can still be used, but the 2×2 complex layer matrices $\mathbf{P}_{j,pp}$ and $\mathbf{P}_{j,ss}$ must now be combined into a single 4×4 complex matrix for each layer. Since the s- and p-polarization states are no longer eigenmodes, then the cross-polarization reflection coefficients r_{sp} and r_{ps} will no longer be zero.

Chapter 9 (Schubert) will deal with this subject in detail, but here an outline of the calculations involved in determining the complex reflection coefficients for anisotropic systems will be given. It is based on the paper by Berreman[4] and the modifications of Lin-Chung and Teitler.[5] Other formalisms (such as that due to Yeh in *Optical Waves in Layered Media*) can also be used.

The Berreman equation is just a re-formulation of Maxwell's equations and is given by

$$\frac{d\Psi}{dz} = -i\Delta\Psi \tag{3.22}$$

where $\Psi^{T} = (E_x, H_y, E_y, -H_x)$, E and H being the magnitudes of the electric and magnetic fields, respectively and the z-direction is perpendicular

to the sample surface. The 4×4 complex Berreman matrix Δ (not to be confused with the ellipsometric angle Δ) contains all the information about the anisotropic dielectric functions of the medium, including the wavelength-dependent dielectric functions along the principal directions and the orientation angles of the principal axis system with respect to the plane of incidence. For a uniaxial material, the Berreman matrix Δ[6] has only seven independent elements and is given by

$$\Delta = \begin{pmatrix} \Delta_{11} & \Delta_{12} & \Delta_{13} & 0 \\ \Delta_{21} & \Delta_{11} & \Delta_{23} & 0 \\ 0 & 0 & 0 & 1 \\ \Delta_{23} & \Delta_{13} & \Delta_{43} & 0 \end{pmatrix}, \tag{3.23}$$

where

$$\Delta\varepsilon = \varepsilon_e - \varepsilon_o \qquad \eta = 1/(\varepsilon_o + \Delta\varepsilon \cos^2\theta) \tag{3.24a, b}$$

$$\Delta_{11} = -\eta\Delta\varepsilon \sin\theta \cos\theta \cos\varphi\, \xi \qquad \Delta_{12} = 1 - \xi\eta \tag{3.24c, d}$$

$$\Delta_{13} = -\eta\Delta\varepsilon \sin\theta \cos\theta \sin\varphi\, \xi \qquad \Delta_{21} = \varepsilon_o \eta(\varepsilon_e - \Delta\varepsilon \sin^2\theta \sin^2\varphi) \tag{3.24e, f}$$

$$\Delta_{23} = \eta\, \Delta\varepsilon\, \varepsilon_o \sin^2\theta \sin\varphi \cos\varphi \qquad \Delta_{43} = \eta\varepsilon_o(\varepsilon_e - \Delta\varepsilon \sin^2\theta \cos^2\varphi) - \xi^2 \tag{3.24g,h}$$

The dielectric functions of the uniaxial medium are ε_o and ε_e for light polarized perpendicular and parallel to the plane of incidence, respectively. The angles of the optic axis are given by θ and φ, where θ is the angle of the optic axis with respect to the normal to the sample surface ($\theta = 0°$ when the optic axis is perpendicular to the surface, and $\theta = 90°$ when the optic axis is parallel to the surface) and φ is the angle of the optic axis with respect to the plane of incidence ($\varphi = 0°$ when the optic axis is in the plane of incidence). (Note that the angle $\varphi = \phi - 90°$ for the angle ϕ used in ref. 6). The quantity ξ is given by Snell's law in Eq. (3.16).

The layer transfer matrix is then given by

$$\mathbf{P}_j = \chi_j\, \mathbf{K}_j\, \chi_j^{-1}, \tag{3.25}$$

where the $\mathbf{K}_j$ matrix is given by

$$K_j = \begin{pmatrix} e^{-iq_{1,j}d_j} & 0 & 0 & 0 \\ 0 & e^{-iq_{2,j}d_j} & 0 & 0 \\ 0 & 0 & e^{-iq_{3,j}d_j} & 0 \\ 0 & 0 & 0 & e^{-iq_{4,j}d_j} \end{pmatrix}. \tag{3.26}$$

The thickness of the j^{th} layer is given by d_j, and the $q_{1,j}$, $q_{2,j}$, $q_{3,j}$, and $q_{4,j}$ are the eigenvalues of the Δ matrix. For uniaxial materials, the four eigenvalues are reasonably simple[6] and are given by

$$q_{1,2} = \pm(\varepsilon_o - \xi^2); \tag{3.27a}$$

$$q_{3,4} = -\eta\,\Delta\varepsilon\,\xi\,\sin\theta\,\cos\theta\,\cos\varphi \pm \eta(\varepsilon_o\,\varepsilon_e)^{1/2}\,[(1/\eta) - (1 - \Delta\varepsilon\,\sin^2\theta\,\sin^2\varphi/\varepsilon_e)\,\xi^2]^{1/2} \tag{3.27b}$$

The χ matrix consists of a column-wise set of eigenvectors of the Δ matrix, where the i^{th} column corresponds to the eigenvalue $q_{i,j}$. It is assumed that the Δ matrix is independent of z in the j^{th} layer.

For layers which are isotropic, or for anisotropic layers with the optical axes in certain symmetry orientations, then the layer matrix $\mathbf{P}_j$ is block diagonal and can be expressed as

$$\mathbf{P}_j = \begin{pmatrix} \mathbf{P}_{j,1} & 0 \\ 0 & \mathbf{P}_{j,2} \end{pmatrix}. \tag{3.28}$$

The matrices $\mathbf{P}_{j,1}$ and $\mathbf{P}_{j,2}$ are 2×2 submatrices. If the material is isotropic, then $\mathbf{P}_{j,1} = \mathbf{P}_{j,pp}$ and $\mathbf{P}_{j,2} = \mathbf{P}_{j,ss}$. If the medium is uniaxial, then the layer matrix $\mathbf{P}_j$ is block diagonal if the optical axis is either in or perpendicular to the plane of incidence.

As with the Abelés matrices for the isotropic case, the final characteristic matrix for a layer stack and substrate containing at least one anisotropic material is given by

$$\mathbf{M} = \chi_0\,(\Pi\,\mathbf{P}_j)\,\chi_{sub}. \tag{3.29}$$

If it can be assumed that the ambient is isotropic, then the χ_0 matrix is block-diagonal of the form shown in Eq. (3.28), where $\mathbf{P}_{0,1} = \chi_{0,pp}$ and $\mathbf{P}_{0,2} = \chi_{0,ss}$ [see Eqs, (3.20a) and (b)] and χ_{sub} is the column-wise set of eigenvectors for the Δ matrix associated with the substrate. The ordering of the eigenvectors in χ_{sub} is important, and the 1st and 3rd columns must contain the eigenvectors that correspond to the solutions that are propagating back to the layer stack. If the substrate is isotropic, then it reduces to a block-diagonal matrix of the form of Eq. (3.28), where $\mathbf{P}_{sub,1} = \chi_{sub,pp}$ and $\mathbf{P}_{sub,2} = \chi_{sub,ss}$ as shown in Eqs. (3.20c) and (d). The complex reflection coefficients are then determined from the elements of the characteristic matrix:

$$r_{pp} = \frac{M_{23}M_{31} - M_{21}M_{33}}{M_{31}M_{31} - M_{11}M_{33}}, \qquad r_{ps} = \frac{M_{43}M_{31} - M_{41}M_{33}}{M_{31}M_{31} - M_{11}M_{33}}, \tag{3.30a, b}$$

$$r_{sp} = \frac{M_{21}M_{13} - M_{23}M_{11}}{M_{31}M_{31} - M_{11}M_{33}}, \qquad r_{ss} = \frac{M_{41}M_{13} - M_{43}M_{11}}{M_{31}M_{31} - M_{11}M_{33}}. \qquad (3.30c, d)$$

3.3.3 Inhomogeneous Layers

All of the layers considered to this point are assumed to be homogeneous in the sense that the optical functions of the materials making up the individual layers do not vary with the z-position. In many cases, this is not an appropriate assumption since the optical functions of many films will also be a function of the depth from the sample surface. The standard way of treating this type of film structure is to divide the inhomogeneous layer into several thinner lamellae, where it is assumed that the optical functions of each lamella are constant. This process always works if the number of lamellae is large enough, but often this results in many layers and therefore a very long calculation. Furthermore, there is no way to determine if a sufficient number of layers have been chosen with a single calculation, so at least two calculations must be performed to determine the accuracy of the approximation.

For some cases, the lamella approximation is the only realistic way of performing the calculations. However, several situations lend themselves to other approximations that can simplify the calculations significantly. If it can be assumed that the dielectric function profile varies smoothly with z, then the methods of Jacobsson[7] can be used. For example, if the dielectric function of the j^{th} inhomogeneous layer varies linearly with depth

$$\varepsilon_{j+1} = \varepsilon_j + az;\ 0 < z < d_j, \qquad (3.31)$$

then the Abelés matrices for the j^{th} layer become

$$\mathbf{P}_{j,pp} = \begin{pmatrix} 1 - \frac{v_j^2}{2}\left[\eta_j + \frac{ad_j}{3}(\varepsilon_j + \xi)\right] & iv_j\left(\varepsilon_j + \frac{ad_j}{2}\right) \\ i\frac{v_j}{\varepsilon_j}\left(\eta_j + \frac{ad_j\xi}{2\varepsilon_j}\right) & 1 - \frac{v_j^2}{2}\left[\eta_j + \frac{ad_j}{3\varepsilon_j}(\varepsilon_j + \eta_j)\right] \end{pmatrix}, \qquad (3.32a)$$

$$\mathbf{P}_{j,ss} = \begin{pmatrix} 1 - \frac{v_j^2}{2}\left(\eta_j + \frac{2ad_j}{3}\right) & iv_j \\ iv_j\left(\eta_j + \frac{ad_j}{2}\right) & 1 - \frac{v_j^2}{2}\left(\eta_j + \frac{ad_j}{3}\right) \end{pmatrix}. \qquad (3.32b)$$

In Eqs. (3.32), $v_j = 2\pi d_j/\lambda$, $\eta_j = \varepsilon_j - \xi$ (not to be confused with the η used for anisotropic materials), and ξ is defined from Snell's law [(see

Eq. (3.16)]. This calculation is performed to second order in v_j and the first order in ad_j, so the thickness of the layer must be thin enough that v_j^3 and $(ad_j)^2$ are very small.

Even more complicated variations in the dielectric functions for a layer can be approximated using the linear approximation given in Eq. (3.31). The variation is broken into several piece-wise linear regions, where Eqs. (3.32) are applied for each region. The convergence criteria also indicate the maximum thickness that can be used for each layer in that the layer thickness d_j can be chosen such that both v_j^3 and $(ad_j)^2$ are small.

If the change in the complex dielectric function with thickness is small, then the light will be minimally reflected from the dielectric function gradient. This type of profile can be treated with an analog of the WKB approximation[7], where the dielectric function of the entire layer is taken to be the dielectric function of the surface. In order for the WKB approximation to be valid, two conditions must be valid:

$$\frac{1}{\beta}\frac{d^2\beta}{d\theta^2} \ll 1, \qquad \frac{1}{\beta}\frac{d\beta}{d\theta} \ll 1, \tag{3.33a, b}$$

where

$$\beta = \cos(\phi)/\tilde{n}(0) \text{ (p-polarization or TM)}, \tag{3.33c}$$

$$\beta = \tilde{n}(0)\cos(\phi) \text{ (s-polarization or TE)}, \tag{3.33d}$$

$$\theta(d) = \frac{2\pi}{\lambda}\int_0^d \tilde{n}(x)\cos(\phi(x))dx. \tag{3.33e}$$

where $\tilde{n}(0)$ is the complex refractive index at the surface, and ϕ is the angle of incidence. In Eqs. (3.33), the integration is performed over the film thickness d. Obviously, these criteria are satisfied when the dielectric function gradient is small with respect to the wavelength of light. If it is assumed that the dielectric function varies linearly with thickness near the surface of the layer, then the conditions stated in Eqs. (3.33) reduce to

$$\frac{\left(\lambda\dfrac{\partial\varepsilon}{\partial x}\right)^2}{\varepsilon^3 4\pi^2} = \frac{(\varepsilon_{top}-\varepsilon_{bottom})^2}{\varepsilon_{ave}^3}\left(\frac{\lambda}{2\pi d}\right)^2 \ll 1. \tag{3.34}$$

3.4 Models for Dielectric Functions

In order to perform the calculations of the complex reflection ratios described in Section 3.3, it is necessary to use realistic values of the dielectric functions as a function of wavelength. For insulators and

semiconductors, the variation of the optical functions for wavelengths greater than the band gap, the extinction coefficient k is very close to zero, and the refractive index n varies very slowly with wavelength. Above the band gap, critical point phenomena for crystalline materials create structure in the dielectric functions. Amorphous materials have less structure, but they do have band gaps and dielectric functions that vary significantly with wavelength. Metals can also have critical point features in their dielectric functions, but they have no band gap, so they have no region where k is very close to 0; in fact, k tends to diverge at large wavelengths due to free carrier effects. Free carrier effects also play a part in the optical functions of very heavily doped semiconductors. See Chapter 2 for a complete discussion of the optical functions of materials.

3.4.1 Tabulated Data Sets

Certainly one approach for incorporating dielectric function data into ellipsometry calculations is to use tabulated data for the materials in each layer that is given in several standard references. Most of the optical data in these references has been obtained from measurements on bulk materials, so the use of this data for substrates is often a very good approach. The reader is referred to Chapter 2 for a discussion of the ellipsometric measurements of standard optical function spectra.

There are, however, several cautions that must be taken when using tabulated data sets:

1) Thin film materials usually do not have the same dielectric functions as the same material in bulk form. For example, fused silica has a lower refractive index in its bulk state than in its thin-film state, and the refractive index of thin-film SiO_2 is can be somewhat higher for very thin films than for thicker films.
2) Many thin films are polycrystalline or amorphous. Experience has shown that the optical functions of these types of materials are often dependent upon the fabrication procedure, so it is inappropriate to use a single data set for a material such as polycrystalline or amorphous silicon.
3) Films grown on different substrates will have different interfaces and therefore different amounts of stress in the films. The stress in the films can have an effect on the optical properties of the film, which will depend on the wavelength-dependent stress-optic coefficients of the material.

4) Most of the tabulated data sets for semiconductors above the band edge are taken from spectroscopic ellipsometry measurements. As will be seen in Section 3.6, the accuracy of these data sets depends significantly on the corrections applied for surface roughness and/or surface oxides. Moreover, optical functions of nominally cubic crystalline materials (such as silicon) will depend upon which face [(001) versus (111)] is being measured with ellipsometry. This effect can be as large as 1 to 2% in some regions of the spectrum, and is due in part to differences in surface reconstruction and oxidation characteristics. Moreover, very few of the tabulated data sets include error limits, so it is not possible to accurately compare the different data sets.

3.4.2 Lorentz Oscillator Model

One of the oldest models for the dielectric functions of materials is due to Lorentz. The classical calculation of the complex dielectric function assumes that the solid is a collection of non-interacting oscillators. If the calculation is performed using quantum mechanics, a similar result is obtained, but the meaning of the individual parameters is quite different (see Chapter 2). This model is only valid for photon energies considerably less than the band gap of a material. The Lorentz model equation for the complex dielectric function is given by

$$\varepsilon(\lambda) = \tilde{n}(\lambda)^2 = 1 + \sum_j \frac{A_j \lambda^2}{\lambda^2 - \lambda_{o,j}^2 + i\zeta_j \lambda} \tag{3.35a}$$

for the dielectric function expressed as a function of the wavelength of light or

$$\varepsilon(E) = \tilde{n}(E)^2 = 1 + \sum_j \frac{B_j}{E_{o,j}^2 - E^2 + i\Gamma_j E} \tag{3.35b}$$

for the dielectric function expressed in terms of the photon energy.

Several other models used in spectroscopic ellipsometry calculations are based on the Lorentz oscillator model. The most common is the Sellmeier approximation, where it is assumed that the extinction coefficient is 0 (this is equivalent to assuming $\zeta_j = \Gamma_j = 0$):

$$\varepsilon = n^2 = 1 + \sum_j \frac{A_j \lambda^2}{\lambda^2 - \lambda_{o,j}^2}. \tag{3.36}$$

Another parameterization based loosely on the Lorentz oscillator model is the Cauchy expansion:

$$n = B_0 + \sum_j \frac{B_j}{\lambda^{2j}}. \tag{3.37}$$

The Drude expression, which is used primarily to express the optical functions of metals and the free carrier effects in semiconductors, is given by setting $E_{o_j} = 0$ in Eq. (3.35b):

$$\varepsilon(E) = 1 - \sum_j \frac{B_j}{E}\left(\frac{1}{E - i\Gamma_j}\right). \tag{3.38}$$

Note that the dielectric functions expressed in Eqs. (3.35) through (3.38) include a summation over an index j. Most of the time, a single term is sufficient to give the accuracy required for spectroscopic ellipsometry measurements. If one is interested in parameterizing refractive index data from minimum deviation experiments (where the accuracy in n can be in the 6th or 7th place) additional terms may be needed. However, for the purposes of spectroscopic ellipsometry data analysis, n need only be determined to three or four places, so a single term is usually all that is required. Similarly, more complicated formulae which are often encountered in the parameterization of the refractive index of glasses are usually an unnecessary complication for spectroscopic ellipsometry data analysis.

3.4.3 Optical Functions of Amorphous Materials

Many thin film samples encountered in spectroscopic ellipsometry are amorphous or nearly amorphous. It is well known that optical transitions in amorphous materials lack **k**-vector conservation, so one expects that the optical functions of amorphous materials will not have very sharp features that are characteristic of the optical function spectra of crystalline materials. Moreover, the optical functions of amorphous materials can vary considerably with growth conditions. In a very real sense, all amorphous materials, and particularly thin film amorphous materials, are different. Therefore, it is extremely important in interpreting spectroscopic ellipsometry data taken on amorphous materials to have models that are capable of dealing with this diversity of optical properties.

There have been several attempts to parameterize the optical functions of amorphous materials. One such parameterization is due to Forouhi and

Bloomer,[8] where the extinction coefficient is given by

$$k(E) = \frac{A(E - E_g)^2}{E^2 - BE + C} = \frac{A(E - E_g)^2}{(E - E_0)^2 + \Gamma^2}. \tag{3.39}$$

The parameters A, E_g, B, and C (or A, E_g, E_o, and Γ) in Eq. (3.39) are treated as fitting parameters. The refractive index is obtained by performing a Kramers-Kronig like integration, which introduced a fifth fitting parameter $n(\infty)$. The last expression in Eq. (3.39) is obtained by substituting $B = 2E_o$ and $C = E_o^2 + \Gamma^2$, and useful when the denominator of the second expression is close to zero.

Although this parameterization has received much attention, there are at least two fundamental problems with the physics of Eq. (3.39).

1) Equation (3.39) was derived assuming that only interband transitions contributed to optical absorption. Yet, $k(E) > 0$ for $E < E_g$. This is clearly unphysical. A glass that is transparent in the visible part of the spectrum has $k \leq 10^{-5}$, while the Forouhi and Bloomer approximation can give $k \geq 0.003$ in the same region. Optical fiber materials have $k \leq 10^{-10}$ near 1 μm, nearly seven orders of magnitude less than that predicted by Eq. (3.39)!
2) Equation (3.39) predicts that $k(E)$ approaches a constant as E gets large. This also is unphysical, since it is known that $k(E)$ approaches 0 at least as fast as $1/E^3$ for all materials. This problem is particularly important in performing the Kramers-Kronig integration, since a proper Kramers-Kronig integration of Eq. (3.39) diverges. This has the consequence that some rules commonly applied in optical analysis cannot be satisfied.

Another parameterization of the optical functions of amorphous materials[9] has been developed that also has only five parameters and corrects the objections given above. This empirical parameterization is based on the Tauc expression for the imaginary part of the dielectric function near the band edge,[10] and the Lorentz oscillator model, given in Eqs. (3.35). If only a single transition is considered, then

$$\varepsilon_2(E) = 2n(E)k(E) = \frac{A(E - E_g)^2}{(E^2 - E_o^2)^2 + \Gamma^2} \frac{\Theta(E - E_g)}{E}, \tag{3.40}$$

where $\Theta(E)$ is the Heaviside function [$\Theta(E) = 1$ for $E \geq 0$ and $\Theta(E) = 0$ for $E < 0$]. The real part of the dielectric function is obtained by Kramers-Kronig integration

$$\varepsilon_1(E) = \varepsilon_1(\infty) + \frac{2}{\pi} P \int_{E_g}^{\infty} \frac{\xi \varepsilon_2(\xi)}{\xi^2 - E^2} d\xi. \tag{3.41}$$

This integral can be performed analytically and is given by

$$\varepsilon_1(E) = \varepsilon_1(\infty) + \left[\frac{A}{\pi\zeta^4}\right]\left\{\frac{a_{\ln}}{2\alpha E_o^2}\ln\left(\frac{E_o^2 + E_g^2 + \alpha E_g}{E_o^2 + E_g^2 - \alpha E_g}\right) - \frac{E^2 + E_g^2}{E}\ln\left(\frac{|E - E_g|}{E + E_g}\right)\right.$$
$$- \frac{a_{a\tan}}{E_o^2\Gamma}\left[\pi - \operatorname{atan}\left(\frac{\alpha + 2E_g}{\Gamma}\right) + \operatorname{atan}\left(\frac{\alpha - 2E_g}{\Gamma}\right)\right]$$
$$\left. + 2E_g \ln\left[\frac{|E - E_g|(E + E_g)}{\sqrt{(E_o^2 - E_g^2)^2 + E_g^2\Gamma^2}}\right] + 2E_g\frac{E^2 - \gamma^2}{\alpha\Gamma}\left[\pi + 2\operatorname{atan}\left(2\frac{\gamma^2 - E_g^2}{\alpha\Gamma}\right)\right]\right\} \tag{3.42a}$$

where

$$a_{\ln} = E^2(E_g^2 - E_o^2) + E_g^2\Gamma^2 - E_o^2(E_o^2 + 3E_g^2) \tag{3.42b}$$
$$a_{a\tan} = (E_o^2 + E_g^2)(E^2 - E_o^2) + E_g^2\Gamma^2 \tag{3.42c}$$
$$\zeta^4 = (E^2 - \gamma^2)^2 + \frac{\alpha^2\Gamma^2}{4} \tag{3.42d}$$
$$\alpha = \sqrt{4E_o^2 - \Gamma^2} \tag{3.42e}$$
$$\gamma = \sqrt{E_0^2 - \Gamma^2/2}. \tag{3.42f}$$

Other models for the optical functions of amorphous materials can also be found in the literature. Adachi[11] also has proposed a four-parameter model based on the Tauc formulation of the imaginary part of the dielectric function near the band edge, given by

$$\varepsilon_2(E) = \frac{D}{E^2}(E - E_g)^2\Theta(E - E_g)\Theta(E - E_c). \tag{3.43}$$

Kramers-Kronig integration [Eq. (3.43)] is used to obtain the real part of the dielectric function, which can be obtained in closed form. Although this model approximately fits the optical functions of several amorphous materials, the abrupt cutoff of the Heaviside functions causes the fits to be less accurate than the Tauc-Lorentz formulation.

Similarly, Yamaguchi, et. al.[12] have proposed an empirical dielectric function based on the Lorentz oscillator model, which is given by

$$\varepsilon(E) = \varepsilon(\infty) + \int \frac{f(E_o)}{E_o^2 - E^2 - iE\Gamma}\, dE_o \tag{3.44}$$

The function $[f(E_o)]^{1/2}$ consists of a hyperbola and an exponential function, smoothly connected, and is completely defined using seven parameters. Again, this function approximately fits the optical functions of several materials, but because the Tauc energy gap is not correctly modeled, the fits are generally less accurate than fits from the Tauc-Lorentz model.

3.4.4 Models for Crystalline Materials

The optical functions of crystalline materials are much more complicated than the optical functions of amorphous materials. This is because the long-range order of crystalline materials leads to critical points in the band structure, which in turn leads to sharp features in the optical functions. However, many applications of spectroscopic ellipsometry involve the characterization of epitaxial thin films of crystalline materials, which may be alloys. The optical functions of epitaxial alloy films, for example, will have well-defined critical points, but the energy, shape, and width will all be different from equivalent bulk crystalline materials due to strain and alloy composition. Therefore, there is a real need for parameterizations for crystalline materials similar to those that have proven to be successful for amorphous materials.

There have been several attempts at parameterizing the optical functions of crystalline materials. These parameterizations are often very good for photon energies very near the critical point, but are not as accurate for photon energies further away. One solution to this problem is to take first or second derivatives of the dielectric function spectra determined from the spectroscopic ellipsometry data and match the experimental derivatives to calculated derivatives. In regions where the dielectric function is slowly varying, the derivatives are very close to zero. Moreover, the sensitivity to thin overlayers is reduced over the entire spectrum. If the critical point is Lorentz-like, then the second derivative of the dielectric function has the form[13]:

$$\frac{d^2\varepsilon(E)}{dE^2} = \frac{Ae^{i\phi}}{(E - E_o + i\Gamma)^n}. \tag{3.45a}$$

where E_o is the central energy value of the critical point, Γ is the broadening parameter, A is the magnitude, which is proportional to the matrix element of the transition, and ϕ is the phase factor. The exponent n in Eq. (3.45a) is dependent on the dimensionality of the critical point, and the phase is dependent upon the nature of the critical point. For a 3-dimensional critical point, n = 3/2, where $\phi = 0°$ for a minimum, $\phi = 90°$ or $180°$ for a saddle point, and $\phi = 270°$ for a maximum. For a 2-dimensional critical

point (one where the effective mass is very large in one dimension), $n = 2$, where $\phi = 0°$ for a minimum, $\phi = 90°$ for a saddle point, and $\phi = 180°$ for a maximum. For a 1-dimensional critical point, $n = 5/2$, where $\phi = 270°$ for a minimum, and $\phi = 0°$ for a maximum. Discrete excitons are represented in this model by $n = 3$, $\phi = 0°$. The actual dielectric function has the same form as Eq. (3.45a), but $n = 1/2$ for a 3-dimensional critical point, $n = -1/2$ for a 1-dimensional critical point, and $n = -1$ for a discrete exciton. The dielectric function for a 2-dimensional critical point is logarithmic, and is given by

$$\varepsilon(E) = Ae^{i\phi}\log(E - E_o + i\Gamma) \quad \text{(2-dimensional)} \tag{3.45b}$$

Many other workers[14] have attempted to fit the dielectric functions for semiconductors to a series of Lorentz oscillators:

$$\varepsilon(E) = \varepsilon_o + \sum_j \frac{A_j e^{i\phi_j}}{E - E_j + i\Gamma_j}, \tag{3.45c}$$

where four parameters are required to for each oscillator included in the sum. By including enough oscillators, it is possible to fit a given experimental dielectric function to reasonable accuracy where the imaginary part is reasonably large. This type of parameterization has been very useful for characterizing the optical functions of some semiconductor alloys, and the fitted parameters can then be in turn fitted to the composition of the alloy. However, this parameterization is not so useful for a material such as silicon, which has an indirect band gap. See Chapter 2 for an extensive discussion of this.

Although Eqs. (3.45a) and (3.45b) have been extended to discrete excitons, most excitons are not discrete, so these expressions are not valid. In particular, it is necessary to consider the type of critical point that is associated with the exciton. For a 3-dimensional, minimum critical point, the dielectric function, is given by[15]

$$\varepsilon(E) = \frac{A\sqrt{R}}{2E^2}\left(\sum_j [g_{b,j}(E + i\Gamma_j) - g_{b,j}(i\Gamma_j)] + g_u(E + i\Gamma_0) - g_u(i\Gamma_0)\right), \tag{3.46a}$$

where the $g_{b,j}$ is the factor for the j^{th} bound orbital of the exciton, and is given by,

$$g_{b,j}(\xi) = \frac{8R}{j^3}\left[\frac{E_0 - R/j^2}{(E_0 - R/j^2)^2 - \xi^2}\right], \tag{3.46b}$$

and the g_u is the factor for the unbound exciton, and is given by

$$g_u(\xi) = -\ln\left(\frac{E_0^2 - \xi^2}{R^2}\right) - \frac{1}{2}\sum_j g_{b_j}(\xi) - \pi\left[\cot\left(\frac{\pi\sqrt{R}}{\sqrt{E_0 - \xi}}\right) + \cot\left(\frac{\pi\sqrt{R}}{\sqrt{E_0 + \xi}}\right)\right]. \tag{3.46c}$$

This model requires at least five parameters. Four of the parameters are the amplitude A, the direct band-gap energy E_0, the excitonic binding energy R, and the broadening energy of the direct band edge Γ_0. Each of the j levels of the excitonic orbital could conceivably require its own broadening term, but it is found empirically that

$$\Gamma_j = \Gamma_0 - \frac{\Gamma_0 - \Gamma_0^{ex}}{j^2}, \tag{3.46d}$$

resulting in only one additional term to describe the broadening of the excitonic structure in the dielectric function. This formulation assumes that the broadening mechanism is Lorentzian, which gives too much absorption below direct band edges.

One of the main problems with all the approaches listed above is that they are designed to model critical points well, but do not model the spectral region between the critical points well, nor do they model indirect gaps well. One approach is to introduce several additional broad oscillators [14] to "fill in" the optical absorption. Another approach, due to Johs et al.,[16] uses a series of piece-wise functions to "fill in" the dielectric function between the critical points, and has been successfully applied to several semiconductor alloys. In all of these cases, the number of parameters required to fit the dielectric function can be very large, which introduces inevitable correlations between the fitted parameters. (See Chapter 2).

3.4.5 Effective Medium Theories

Often one would like to model the optical functions of thin films using an average of two or more other sets of optical functions. The issue then is how best to perform the average. In order to perform this average, one tries to find a composite or effective medium dielectric function for the whole film based on the dielectric functions of two or more other materials.

Three common effective medium approximation (EMA) theories can be jointly expressed[17] by

$$\frac{<\varepsilon> - \varepsilon_h}{<\varepsilon> + \gamma\varepsilon_h} = \sum_j f_j \frac{\varepsilon_j - \varepsilon_h}{\varepsilon_j + \gamma\varepsilon_h}, \tag{3.47}$$

where $<\varepsilon>$ is the dielectric function of the effective medium, ε_h is the dielectric function of the host, f_j is the fraction of the j^{th} constituent, and γ is a factor related to the screening and the shape of the inclusions (for example, $\gamma = 2$ for 3-dimensional spheres).

Within the structure of Eq. (3.47) the only difference between the three EMA models is the choice of the host material.

1) Lorentz-Lorentz: $\varepsilon_h = 1$, where the host is chosen as air. This is the earliest EMA theory, and is based on the Clausius-Mossotti equation. It assumes that the individual constituents are mixed on the atomic scale, and is therefore of limited usefulness in describing real materials, which tend to be mixed on a much larger scale.
2) Maxwell-Garnett: $\varepsilon_h = \varepsilon_1$, where the host material is the material that has the largest constituent fraction. This is the most realistic EMA theory when the fraction of inclusions is significantly less than the fraction of host material. This EMA is very useful for cermats or for certain types of nanocrystals embedded in hosts well below the perculation threshold.
3) Bruggeman: $\varepsilon_h = <\varepsilon>$, where the host material is just the EMA dielectric function. The Bruggeman EMA makes no assumption concerning the material that has the highest constituent fraction, and is therefore self-consistent. It is most useful when no constituent forms a clear majority of the material. Surface roughness has been successfully modeled using the Bruggeman EMA consisting of ~50% voids and ~50% material underneath. If the surface layer is very thick, it is sometimes necessary to incorporate several layers, each with a different Bruggeman EMA fraction of voids and underlying material.

Each of these EMAs can be calculated for any material, including those with complex dielectric functions. However, if either of the materials has a complex dielectric function, then care must be taken in the Bruggeman calculation of Eq. (3.47), since it is easy to use the wrong branch of a multivalued inverse of a complex function. For two constituents, this can be avoided by using a re-parameterization of Eq. (3.47) due to Roussel et al.[18] If

$$p = \sqrt{\frac{\varepsilon_1}{\varepsilon_2}}, \quad b = \frac{1}{4}\left[(3f_2 - 1)\left(\frac{1}{p} - p\right) + p\right], \tag{3.48a, b}$$

$$z = b + \sqrt{b^2 + 0.5}, \tag{3.48c}$$

then the EMA value of the dielectric function is given by

$$<\varepsilon> = z\sqrt{\varepsilon_1 \varepsilon_2}. \tag{3.48d}$$

3.5 Fitting Models to Data

Spectroscopic ellipsometry data analysis consists, for the most part, of fitting experimental data to a model, where certain parameters are allowed to "float" to obtain the best fit. Clearly, this process is critical to getting useful results from spectroscopic ellipsometry experiments. Even the most accurate and reproducible experiments can be rendered useless by inappropriate modeling.

The fitting process consists of three parts:

1) Select the model. In this step, the number of layers is fixed, and basic structure concerning the contents of each layer is set. One must decide if any layer is anisotropic at this stage, and whether or not interface layers are to be modeled as a single effective medium approximation, or is a more complicated graded interface to be used.
2) Assign optical functions to each layer. Here, the decision must be made to use existing data sets or some kind of parameterization. For example, if the sample consisted of thin-film SiO_2 grown on silicon, it is very likely that the optical functions of the substrate layer will be very close to the reference values of crystalline silicon. The film might be modeled with three layers and a substrate: surface roughness / amorphous SiO_2 / interface layer / crystalline silicon. The surface roughness (the top layer) might be best modeled as a Bruggeman effective medium consisting of 50% air and 50% SiO_2, while the middle layer might be best modeled using a single Lorentz oscillator. The interface layer might again be modeled using a 2-medium Bruggeman effective medium, consisting of 50% layer 2 and 50% silicon. The initial values for all parameters and film thicknesses must be assigned, and the "floating" parameters selected.

3) Fit the data. The last step consists of the fitting process itself. Here, the metric must be chosen that will decide between good and bad fits to the data. Moreover, this metric must tell us, at least in a qualitative sense, whether or not the model actually fits the data, given the error limits of the data. If the metric that is chosen is inappropriate, it may be that we accept wrong models, or unnecessarily complicate a reasonably valid model. Here, the choice of the numerical analysis optimization algorithm must be defined.

The mathematics of this data analysis has been understood for a number of years, and is called nonlinear parameter estimation. Depending upon how the fitting process is performed, very defined, mathematically proven, statements can be made concerning the validity of the fitting results. All fitting procedures are not equal! In the fitting of spectroscopic ellipsometry data, as in all modeling of data, it is very important to keep this in mind. As stated in *Numerical Recipes, 2nd edition*, pg. 650:

> The important message we want to deliver is that fitting of parameters is not the end-all of parameter estimation. To be genuinely useful, a fitting procedure should provide, (i) parameters, (ii) error estimates on the parameters, and (iii) a statistical measure of goodness of fit. When the third item suggests that the model is an unlikely match to the data, then items (i) and (ii) are probably worthless. Unfortunately, many practitioners of parameter estimation never proceed beyond item (i). They deem a fit acceptable if a graph of the data and model "looks good." This approach is known as *chi-by-eye*. Luckily, its practitioners get what they deserve.

3.5.1 Figures of Merit

In any numerical fitting procedure, the most critical choice is that of the figure of merit, for this tells the computer and the user how well the model, with the present set of parameters, actually fits the data. This has been a significant controversial issue in the analysis of spectroscopic ellipsometry data for the last several years, though it need not be: the mathematics of nonlinear parameter estimation have been well-known for decades (see *Nonlinear Parameter Estimation* by Bard, or *Numerical* Recipes, Chapter 15 by Press, Flannery, Teukolsky, and Vetterling).

There are two common choices used as a figure of merit for spectroscopic ellipsometry data analysis. The oldest, called either the unbiased

estimator or the mean squared error (MSE), is given by

$$MSE = \frac{1}{N - m - 1} \sum_{j=1}^{N} [\rho_{exp}(\lambda_j) - \rho_{calc}(\lambda_j, \mathbf{z})]^2. \tag{3.49}$$

The other is the reduced χ^2, which is given by

$$\chi^2 = \frac{1}{N - m - 1} \sum_{j=1}^{N} \frac{[\rho_{\exp}(\lambda_j) - \rho_{calc}(\lambda_j, \mathbf{z})]^2}{\sigma(\lambda_j)^2}. \tag{3.50}$$

In both Eqs. (3.49) and (3.50), N is the total number of data points taken, $\rho_{exp}(\lambda_j)$ is the experimental data points taken at different values of the wavelength of light λ_j, $\rho_{calc}(\lambda_j,\mathbf{z})$ is the calculated value of that experimental parameter at λ_j and for a particular set of the variable parameters, indicated by the vector $\mathbf{z}$. The only difference between the two expressions is the inclusion of the point-wise experimental error $\sigma(\lambda_j)$ in the denominator of Eq. (3.50). Yet, from a mathematical point of view, this is a big difference and has far-reaching consequences concerning the conclusions that can be drawn from spectroscopic ellipsometry data fits.

When Eq. (3.49) is used as a figure of merit function, it is assumed that all the experimental errors are unmeasureable and that they are the same for each data point. The *MSE* is in the square of the units of the measured quantities, therefore the value of the *MSE* is dependent upon the representation of the data; change the data representation, and you change the value of the *MSE*. This presents a significant problem with the analysis of spectroscopic ellipsometry results, because often the experimental data is transformed into a different representation before fitting the data.

One can easily get quite different fit results from the *MSE* figure of merit, depending upon which representation is chosen for the data analysis. For example, if one were using data from a standard rotating analyzer ellipsometer, there are several choices of data sets that can be fit to a model, where common choices of fitted "experimental" data are (ψ, Δ), (tan(ψ), cos(Δ)), or the actual experimental data (a_1, a_2). Given that the standard rotating analyzer ellipsometer measures Δ very inaccurately when Δ is close to 0° or 180°, it is highly inappropriate to weight data points where Δ is close to 0° or 180° as much as data points where Δ is measured more accurately. From this point of view, it is much better to use the actual data (the Fourier coefficients a_1 and a_2) as suggested by Chindaum and Vedam.[19]

Although Eq. (3.49) is routinely used to analyze spectroscopic ellipsometry data, and error limits on the fitted parameters quoted, the error limits so obtained ASSUME that the selected model is the right model. *This is often the very thing that we wish to determine from spectroscopic*

ellipsometry experiments! Moreover, the actual value of the *MSE* has the dimensions of the chosen parameterization of the experimental data, and is therefore meaningless; one never knows when a good fit is actually obtained! The only mathematically justifiable reason for using the *MSE* as a figure of merit is if there is no possible way of obtaining an estimate of the errors in the experimental data. As will be discussed below, it is quite possible to obtain reasonable error limits of spectroscopic ellipsometry experiments.

These objections are removed if the reduced χ^2 is chosen as the figure of merit (there are other figure of merit functions which would also be valid in special circumstances). The reduced χ^2 is unitless, and is in fact a measure of the "goodness of fit:" if $\chi^2 \approx 1$, then the model fits the data, but if $\chi^2 \gg 1$, then the model does not fit the data. Furthermore, the results of the fitting procedure are independent of the representation of the data, IF the errors are propagated into the data representation used in the fitting procedure. For example, if the errors in (a_1, a_2) from a rotating analyzer ellipsometry experiment were propagated into (ψ, Δ), then the errors in Δ near 0° and 180° would naturally be very large, and would therefore not be heavily weighted in the fitting procedure.

Another very important issue in fitting experimental data is the determination of proper error limits from the results of the fit. If the *MSE* is used as the figure of merit, then any error limits obtained are only relative and again assume that the model selected is the correct model. However, error limits obtained from a fitting procedure using χ^2 as the figure of merit can be statistically related (if only approximately) to the actual error limits of the fitted parameters, and therefore have considerably more validity.

3.5.2 Errors in Spectroscopic Ellipsometry

Since error limits are such an important characteristic of any spectroscopic ellipsometry data set, they must be discussed first. As in any experiment, errors come in two different forms: random and systematic.

Random errors are those errors resulting from random or statistical processes. In spectroscopic ellipsometry experiments, the two major sources of random errors are due to photon counting or shot noise, and the normal thermally generated noise in electronic circuits. This noise is reduced by signal averaging, and can be measured by performing several identical experiments and calculating the mean and the standard deviation. Once the error is determined, it can be propagated from the actual experimental data into other representations of the data using standard data analysis techniques.

In any ideal experiment, systematic errors would be totally eliminated by alignment, calibrations, and by performing multiple experiments. However, spectroscopic ellipsometry measurements are sufficiently accurate that systematic errors can be as large as random errors in some circumstances. These errors come from such diverse sources as misalignment of optical elements, calibration drifts, and depolarization effects. Because systematic errors do not include a random component, signal averaging cannot be used to reduce the error. Random errors can be rigorously propagated into different representations of spectroscopic ellipsometry data, but systematic errors cannot be rigorously propagated into other representations.

Some sources of systematic errors are given below:

1) Azimuthal alignment of optical elements: The azimuthal alignment errors of the polarizers, compensators, photoelastic modulators, etc. with respect to the plane of incidence in ellipsometry experiments contribute an error to the measured components of the elements of the sample Mueller matrix. This error is $2|q|\sigma_\theta$, where σ_θ is the error in alignment of a particular optical element and is measured in radians and q is a factor containing Mueller matrix elements and sines and cosines of optical element azimuthal angles, where $0 \leq |q| \leq 1$. For example, if each optical element is aligned with an error of 0.02°, then the error in the measured quantities will be < 0.0007 for each optical element. Since there are several optical elements in an SE experiment, each of which has approximately this error limit, this source of error can result in a significant systematic error in the experiment.
2) Wavelength errors: Spectroscopic ellipsometers use a monochromator either before the light goes through the polarization state generator or after the polarization state detector. In either case, only quasi-monochromatic light is produced or detected, due to the finite spread in wavelengths that passes through the monochromator. For ellipsometers that use spectrographs and photodiode arrays, this error can be quite large since a wide wavelength spread is usually integrated into a single pixel. In addition, the calibration of the wavelength may be off or may drift between calibrations. These two sources of error are actually quite different and lead to quite different effects on the experimental spectra. The finite spread in wavelengths leads to a quasi-depolarization of the light beam, while the

error in wavelength calibration leads to shifts in the energy or wavelength scale of the ellipsometry spectrum. The latter error is particularly important in regions of the spectrum where the measured quantities vary rapidly with wavelength or photon energy.

3) Angle of Incidence: The angle of incidence is a particularly difficult parameter to measure, and its error is hard to quantify. Generally, it is very hard to measure this quantity to better than ~0.02°. In addition, all ellipsometers use light sources that are not perfectly collimated, so the sample is actually illuminated with a light beam with a distribution of angles of incidence.
4) Calibrations: There are many calibrations that must be performed on spectroscopic ellipsometers, some of which may drift during the running of actual experiments. For example, ellipsometers which contain photoelastic modulators (PEMs) or Faraday rotators must be calibrated for the amplitude of modulation and the static strain of the modulator as a function of wavelength. Spectroscopic ellipsometers that use fixed compensators must calibrate the amount of phase shift as a function of wavelengths. Of course, all ellipsometers use detectors and amplifier circuits that are not perfectly linear.
5) Assumed data sets: If the analysis of a spectroscopic ellipsometry measurement requires a standard data set of optical functions for a material, then errors in these optical functions also factor into the error in the fitting procedure. For example, there has been considerable work on the optical functions of silicon over the last few years, expressly for the purpose of improving spectroscopic ellipsometry analysis techniques of thin films grown on silicon. Above the direct band edge of silicon (~3.4 eV), there are small but statistically significant variances between the different data sets. Some of this variance comes about because the different data sets are taken on different faces of silicon (100), (110), and (111), while some of the variance arises because of differences in the way in which the surface roughness/oxide is taken into account. Even for this well-studied material, the errors in the optical functions of silicon above the band edge are in the 1 to 2% range. This will have a small but noticeable effect on the error limits of any fitting procedure using a standard silicon data set.

One way of dealing with systematic errors is to assume that they are random, propagating the error limits into the fitted data representation. This approach requires that an estimate of the random errors be made based on the knowledge of a particular ellipsometer. The total error of the ellipsometer is then a propagation of both random and systematic errors into the data representation. The resulting error estimates are not mathematically rigorous, but they do represent approximately the actual errors of the spectroscopic ellipsometer, which is far better than ignoring the systematic errors altogether.

Another way of dealing with systematic errors is to treat the particular parameters as a fittable parameter (see below). For example, if there is a significant uncertainty in the angle of incidence, it may be appropriate to treat the angle of incidence as a fittable parameter, thereby determining the angle of incidence in addition to the other parameters determined using the fitting procedure. If this approach is used, the same care must be taken to determine all errors and correlation parameters as one would use for the other fittable parameters in the problem. This will be discussed below.

3.5.3 Convergence Routines

Once the figure of merit is selected and the errors specified, the data can be fit to a selected model. This problem of fitting spectroscopic ellipsometry data to reflection calculations is mathematically nonlinear and many of the contributions to the errors are systematic rather than random. Therefore, many of the statements that can be made exactly for linear regression analysis with random errors can now only be made approximately. The standard numerical analysis routine used for non-linear curve fitting is the Levenberg-Marquardt algorithm. This algorithm is actually a combination of two different numerical analysis procedures: the inverse Hessian method and the steepest descent method.

Assume that the spectroscopic ellipsometry data is given by $\rho_{exp}(\lambda_i)$ and we want to fit this data to a model $\rho_{calc}(\lambda_i, \mathbf{z})$, where the quantities λ_i are the wavelengths for each spectrum for $i = 1$ to N and the vector $\mathbf{z}$ contains the parameters in the model that are to be varied in order to reduce χ^2 to a minimum. It has been assumed that there is a single independent variable λ, but there could be additional independent variables, such as the angle of incidence.

The object of the nonlinear fitting procedure is to minimize the figure of merit function, given in Eq. (3.50). Rewriting Eq. (3.50) to show the dependence on the parameter vector $\mathbf{z}$:

$$\chi^2(\mathbf{z}) = \frac{1}{N - m - 1} \sum_{j=1}^{N} \frac{[\rho_{exp}(\lambda_j) - \rho_{calc}(\lambda_j, \mathbf{z})]^2}{\sigma(\lambda_j)^2}. \tag{3.51}$$

At the minimum of $\chi^2(\mathbf{z})$ with respect to the parameters $\mathbf{z}$, the gradient will be zero. In the neighborhood of the minimum, the gradient is given by

$$\beta_k = \frac{1}{2}\frac{\partial \chi^2}{\partial z_k} = -\sum_{j=1}^{N} \frac{[\rho_{exp}(\lambda_j, \mathbf{z}) - \rho_{calc}(\lambda_j)]}{\sigma(\lambda_j)^2} \frac{\partial \rho_{calc}(\lambda_j, \mathbf{z})}{\partial z_k}. \quad (3.52a)$$

Taking another partial derivative and ignoring the 2nd derivative term gives:

$$\alpha_{kl} = \frac{1}{2}\frac{\partial^2 \chi^2}{\partial z_k \partial z_l} = \sum_{j=1}^{N} \frac{1}{\sigma(\lambda_j)^2} \frac{\partial \rho_{calc}(\lambda_j, \mathbf{z})}{\partial z_k} \frac{\partial \rho_{calc}(\lambda_j, \mathbf{z})}{\partial z_l}. \quad (3.52b)$$

The α matrix is usually called the curvature matrix, and is equal to 1/2 the equivalent Hessian matrix of least squares fitting procedures.

The Levenberg-Marquardt nonlinear fitting procedure requires that an additional parameter $\wp$ be defined (most textbooks refer to this as λ) such that a new matrix α' is defined:

$$\alpha_{jj}' \equiv \alpha_{jj}(1 + \wp), \qquad \alpha_{jk}' \equiv \alpha_{jk} \; (j \neq k). \quad (3.52c)$$

The Levenberg-Marquardt procedure follows the following recipe:

1) Make a reasonable estimation of the parameter vector $\mathbf{z}$. (This will be discussed below).
2) Compute $\chi^2(\mathbf{z})$.
3) Choose a reasonable initial value of $\wp$ (usually ~10^{-3}).
4) Calculate the α, β, and the α' matrices.
5) Calculate $\delta\mathbf{z} = [\alpha']^{-1}\beta$ and evaluate $\chi^2(\mathbf{z} + \delta\mathbf{z})$.
6) Check for stopping the iteration if $\chi^2(\mathbf{z} + \delta\mathbf{z}) < \chi^2(\mathbf{z})$ and if $\wp$ was not increased for the last iteration.
7) If $\chi^2(\mathbf{z} + \delta\mathbf{z}) \geq \chi^2(\mathbf{z})$, increase $\wp$ by a large factor ($\sqrt{10}$ or 10), and go back to 5).
8) If $\chi^2(\mathbf{z} + \delta\mathbf{z}) < \chi^2(\mathbf{z})$, decrease $\wp$ by a large factor ($\sqrt{10}$ or 10), update the solution for the $\mathbf{z}$-vector $\mathbf{z} + \delta\mathbf{z} \rightarrow \mathbf{z}$, and go back to 4)

Step 1) involves the initial guess for the parameter estimation. Given the nonlinearity of the data analysis problem, this is a critical step because it is very easy to have the routine converge to a local minimum and not the global minimum. For many problems, this is not an issue, since the user knows the approximate position in parameter space where the solution lies. However, if the film structure is very complicated, or the materials involved are unknown or complicated composites, one must first use

other numerical techniques to find the initial estimate. One way to find the initial estimate is to search over all possible parameter space using a grid search or a monte carlo search. Another way is to manually change the individual elements of the parameter vector **z** until the fit looks realistic. In both cases, it is important to use both the figure of merit function and the user's intuition as a guide to determining when a close solution is obtained.

Step 6) indicates that one must have one or more criteria for stopping the iteration. Generally, it is not a good idea to stop the iteration when $\wp$ has been increased. Nor is it a good idea to continue the iteration process if each iteration is not decreasing $\chi^2(\mathbf{z})$ very much (usually ~0.001).

Once the solution vector **z** has been found, the covariance matrix α^{-1} can be calculated, as can approximate errors in the solution vector **z** (see the discussion below). For the 95% confidence interval, the approximate correlated errors in the elements of the parameter vector are given by $2\sqrt{(\alpha_{jj}^{-1})}$, while the uncorrelated errors are given by $2/\sqrt{(\alpha_{jj})}$. The cross-correlation coefficients are given approximately by the elements of the symmetric matrix α^{-1}.

It must be emphasized that the relationships given above for the correlated and uncorrelated errors and for the cross correlation coefficients are only approximate. If the problem were a linear regression problem, with random errors, then the expressions for these quantities have a precise mathematical meaning. However, the nonlinearity and the systematic errors inherent in ellipsometry make the relations only approximate.

One way of looking at this problem from a statistical point of view is shown schematically in Fig. 3.2. Assume that one performs several spectroscopic ellipsometry experiments on the same sample of SiO_2 grown on silicon that is assumed to be stable over long periods of time. Furthermore, assume that when each of these experiments is performed, all the alignment procedures are performed from the beginning: one takes apart the ellipsometer and re-assembles it for each measurement (this is to randomize the systematic errors). Finally, assume that the refractive index of the SiO_2 film can be modeled by the Sellmeier approximation, Eq. (3.36), where $\lambda_{1o} = 93$. This model has only two parameters, the film thickness d and the Sellmeier parameter A. The results of each experiment will then be represented in Fig. 3.2 by a data point. If enough experiments are performed and their data points plotted, then an elliptical-like area will begin to be formed. That is, one could draw an ellipse around the set of data points that would enclose 95% of the points (for the 2σ confidence interval).

In this picture, the height and the width of the drawn ellipse would represent the correlated error for the values of the Sellmeier A and the film thickness d respectively. The uncorrelated error is not usually of much interest, but it is given by the height of the ellipse at $d = d_f$ and at

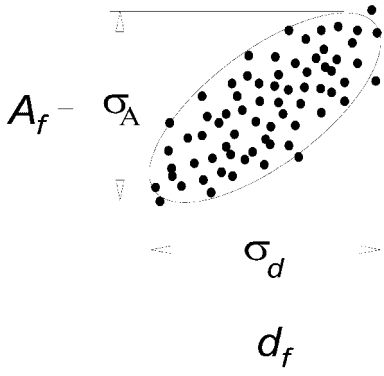


Figure 3.2 A Schematic of parameter space for a two-dimensional fit of spectroscopic ellipsometry data. Each point represents an independent experiment and fitting procedure, where all the errors are randomized. The drawn ellipse defines the 95% confidence limits for the parameters A_f and d_f, while the uncorrelated errors are given by σ_A and σ_d.

$A_1 = A_f$ respectively. If the nonlinearity of the problem near the solution point were severe, then the contour drawn for the 95% confidence limits would not be an ellipse at all, but might look like a banana! Two totally uncorrelated parameters would result in an ellipse that was symmetric about both the d and the A axes, while two very highly correlated parameters would result in an ellipse that was long and narrow, where the major axis followed the relation $d = \mathrm{t}\, A$, where t is a constant.

The example presented above has been simplified to two dimensions for ease in presentation. However, many fits to spectroscopic ellipsometry data use many more parameters, where the two-dimensional ellipse of Fig. 3.2 would become a multi-dimensional ellipsoid-like shape.

3.5.4 An Example: (a-Si_xN_y:H)

Amorphous silicon nitride is an extremely interesting material from the point of view of its optical properties. The material is grown as a thin film by many different techniques, and can be grown either as a silicon-rich or as a nitrogen-rich film, where the silicon-to-nitrogen ratio has a dramatic effect on its optical properties. The films are usually quite smooth, and are generally uniform throughout the thickness of the film. Being amorphous and nearly unstrained, the film is very nearly isotropic.

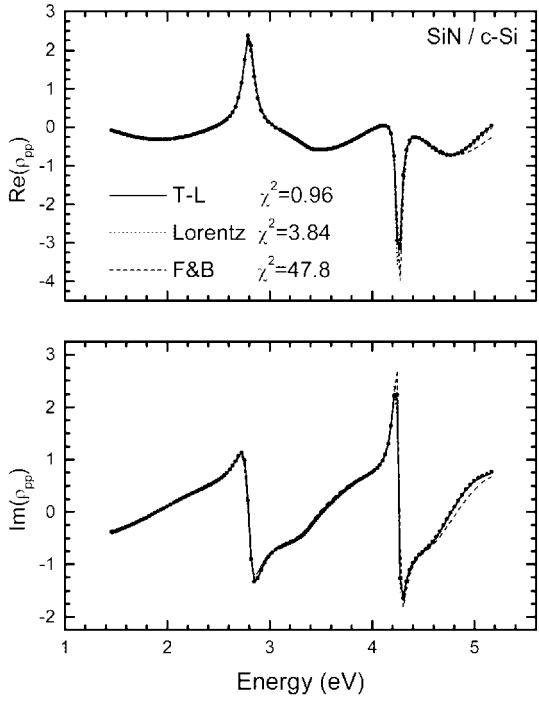


Figure 3.3 A Spectroscopic ellipsometry spectrum for amorphous silicon nitride grown on silicon. Three fits to the data are shown, based on the Lorentz oscillator model, the Forouhi and Bloomer amorphous model, and the Tauc-Lorentz model.

Figure 3.3 shows a typical set of data (presented as the real and imaginary parts of the complex reflection ratio, see Eq. (3.2), from a spectroscopic ellipsometry experiment for a silicon nitride film grown using PE-CVD. As can be seen, the data contains two resonance features in both Re(ρ) and Im(ρ) due to interference effects. Beyond this rather general statement, it is not possible to say much more about the film by a simple visual examination of the data.

To obtain more information about the film, such as the film thickness, surface roughness, and film optical properties, we must model the spectroscopic ellipsometry data. For this rather simple system, we will consider three possible models for the optical functions of the thin-film a-Si_xN_y:H:

1) a Lorentz oscillator with three parameters, B, E_o, and Γ, see Eq. (3.35b),

2) the Forouhi and Bloomer amorphous model with five parameters, see Eq. (3.39), and
3) the Tauc-Lorentz amorphous model with five parameters, see Eqs. (3.40) to (3.42).

Within each model, we can examine the effect of including a surface roughness layer and an interface layer. The surface roughness layer is modeled using a Bruggeman effective medium approximation with 50% voids and 50% film, while the interface layer is modeled using 50% film and 50% silicon. The optical functions of crystalline silicon are taken from ref. 20. Therefore, the Lorentz model will have a maximum of seven fitted parameters, and the Forouhi and Bloomer model will have a maximum of eight parameters, as will the Tauc-Lorentz model.

The details of the fitting results are shown in Table 3-1 for the Lorentz oscillator model, for the Forouhi and Bloomer model, and for the Tauc-Lorentz model. The first thing to note from these results is that there is a large discrepancy in the final, best values of the reduced χ^2. The Tauc-Lorentz model gives a very believable χ^2 value of 0.92, while the Lorentz model gives a higher value of 3.64 and the Forouhi and Bloomer model gives a value of 47.5. From this result alone, we can state that the Tauc-Lorentz model fits the data while the other two models do not fit the data.

Table 3-1 also lists the correlated error limits for the fitted parameters. Not shown, but just as important, are the cross-correlation coefficients, which indicate which parameters are inter-correlated. Several other things can be stated from the fit results presented in Table 3-1:

1) Film thickness. From the values and the correlated errors, we can see that all three models give the same values of film thickness, although the correlated error of the film thickness from the Tauc-Lorentz model is smaller. This emphasizes that a film thickness determination from spectroscopic ellipsometry measurements is often quite robust to differences in modeling. Even wrong models can give quite reasonable values of film thicknesses.
2) Roughness thickness. All three models predict a small but significant surface roughness. Since the surface roughness is so thin, it is not possible to be very precise concerning the optical functions of this layer.
3) Band Gap. Both the Tauc-Lorentz and the Forouhi and Bloomer models incorporate a band gap, but the uncorrelated error given for the Tauc-Lorentz band gap is small, as is the final value of the χ^2, making the band gap a reasonably

Table 3-1. The results of fitting the data shown in Fig. 3.3 to three different models for the optical functions of amorphous materials. The first part of the table shows the fitting results for a 7- or 8-parameter fit, while the second part shows the fitting results for a 5- or 6-parameter fit. The last part of the table shows representative values of the complex refractive index.

	Lorentz	**Forouhi and Bloomer**	**Tauc-Lorentz**
Roughness thickness (nm)	2.1 ± 0.3	4.9 ± 0.7	1.9 ± 0.3
Film thickness (nm)	197.8 ± 0.7	195.6 ± 1.1	198.2 ± 0.4
Interface thickness (nm)	0.6 ± 0.3	−0.6 ± 0.6	−0.1 ± 0.2
A	201.9 ± 4.6	4.56 ± 1.9	78.4 ± 12.7
E_o (eV)	9.26 ± 0.05	74.4 ± 30.5	8.93 ± 0.47
Γ (eV)	0.01 ± 0.02	0.74 ± 23.5	1.82 ± 0.81
E_g (eV)	—	2.85 ± 0.48	4.35 ± 0.09
ε(∞)	1.00 ± 0.02	0.93 ± 0.51	1.38 ± 0.26
χ^2	3.64	47.5	0.92
Roughness thickness (nm)	2.4 ± 0.3	4.7 ± 0.6	1.8 ± 0.2
Film thickness (nm)	198.6 ± 0.4	195.3 ± 0.9	198.1 ± 0.3
A	202.5 ± 1.4	5.03 ± 0.35	97.7 ± 3.2
E_o (eV)	9.26 ± 0.02	70.7 ± 2.2	9.61 ± 0.03
Γ (eV)	0.01 ± 0.01	40.1 ± 10.8	3.07 ± 0.33
E_g (eV)	—	2.97 ± 0.27	4.44 ± 0.04
χ^2	3.75	48.0	0.96
n(300 nm)	1.986 + i0.000	2.003 + i0.001	1.989 + i0.000
n(600 nm)	1.866 + i0.000	1.839 + i0.001	1.874 + i0.000

determined parameter from the fit of the spectroscopic ellipsometry data. On the other hand, the band gap determined from the Forouhi and Bloomer model is not believable since the final value of χ^2 is so high.

4) Interface thickness. All three models predict a very small, if negligible interface thickness. This result, coupled with the small χ^2 of the Tauc-Lorentz fit, indicates that the film is very uniform in its optical properties throughout its thickness.

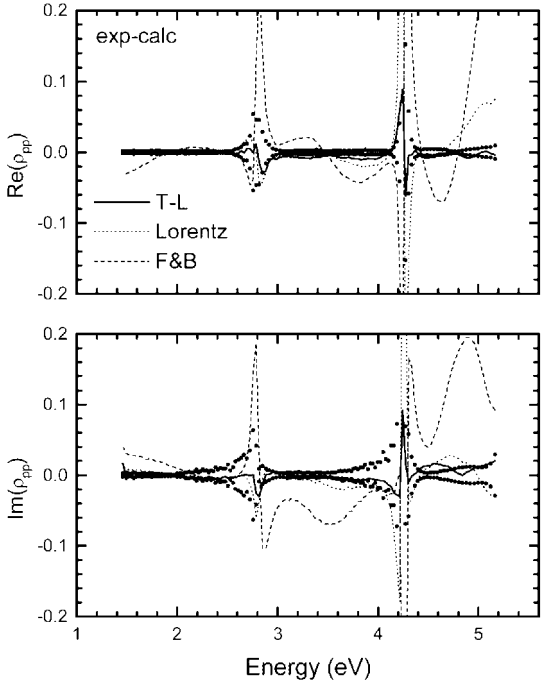


Figure 3.4 Difference spectra for the fits and data shown in Fig. 3.3. The dots indicate the error limits of the data.

5) $\varepsilon(\infty)$. All three models predict a value of $\varepsilon(\infty)$ very close to one. If there are no major transitions at higher values of energy, then this parameter should be one.
6) Correlated parameters. If any parameter has a relatively large value of correlated error, then that parameter can be very strongly correlated with at least one other parameter. The identity of the other parameter can be determined by an examination of the cross-correlation coefficients. For example, the fit to the Tauc-Lorentz model results in quite large correlated errors in the A, E_o, Γ, and $\varepsilon(\infty)$ parameters. An examination of the cross-correlation coefficients shows that all of these parameters are strongly inter-correlated.

Given that all fits give a value of $\varepsilon(\infty)$ close to one and a value of the interface thickness close to zero, it is reasonable to perform the fits again, this time eliminating the interface layer and setting $\varepsilon(\infty) = 1$. The results

to these fits are shown in the second part of Table 3-1. As can be seen, reducing the number of fitted parameters by two does not increase the reduced χ^2 significantly in any of the fits. However, it does have a profound effect on the correlated errors in all cases. Whereas the correlated errors for the A, E_o, and Γ parameters were quite large in the Tauc-Lorentz fit when the fit included eight parameters, the error limits in all the parameters were reduced significantly when the fit was restricted to six parameters.

In view of the large value of the reduced χ^2 for the Lorentz and Forouhi and Bloomer fits, it must be emphasized that the models do not fit the data and therefore discussions of fitted values and error limits are not appropriate from a rigorous point of view. In the third part of Table 3-1, the complex refractive index is shown at 300 nm and at 600 nm. The nonzero imaginary part of the Forouhi and Bloomer refractive index at 600 nm (below the band gap) is due to the $(E - E_g)^2$ factor in Eq. (3.51). Although the refractive indices quoted for all three models are not far from each other, the difference is significant in view of the χ^2 test.

There are situations where one might accept models with a χ^2 greater than one, particularly if the systematic errors in spectroscopic ellipsometry data are significant. Since systematic errors do not obey the same statistical laws that random errors do, it is quite possible for these errors to add together in such a way that the reduced χ^2 would be greater than one and the model would still fit the data. However, if another model does give a χ^2 of one when a favorite model does not to the same data set, one would be advised to reject the favorite model in favor of the model that does give $\chi^2 \sim 1$. One might also accept a model with a higher χ^2 if one were interested only in the film thickness, but this is can be quite risky.

In all this analysis, it must be remembered that spectroscopic ellipsometry, being model-dependent, can never state whether a model is the correct model. The only statement than can be made is whether a particular model fits the data or not.

3.6 Determination of Optical Functions from Spectroscopic Ellipsometry Data

A second major use of spectroscopic ellipsometry data is to obtain the optical functions of bulk and thin-film materials. For photon energies less than the band gap of the material, spectroscopic ellipsometry can be used to determine the real part of the dielectric function (or the refractive index), but the older technique of minimum angle deviation is more accurate. Typically, spectroscopic ellipsometry can measure a refractive index

to ±0.001 to 0.010, while minimum angle deviation methods can measure the refractive index to 10^{-4} to 10^{-7}. However, for some samples, spectroscopic ellipsometry is the preferred method, even with its decreased accuracy, since SE requires only one polished surface, while minimum deviation measurements require that a prism be made from the material.

For photon energies above the band edge, spectroscopic ellipsometry is far more accurate than any other competing technique. For decades, the standard technique for the determination of the complex refractive index has been normal-incidence reflectivity. However, the spectroscopic reflectivity spectrum is very difficult to determine with very high accuracy, due primarily to the loss of non-specular reflected light and the lack of well-characterized standards. In addition, the reflectivity measurement results in only a single number, the reflected light intensity. In order to get the complex refractive index, Kramers-Kronig integration is required, and this mathematical procedure requires an extrapolation of the measured spectrum to zero energy and to infinite energy. Of course, any surface overlayer or roughness will affect the reflectivity (particularly in the ultraviolet), which must be taken into account or it will increase the error of the measured value of the reflectivity.

Spectroscopic ellipsometry, on the other hand, measures at least two parameters (six parameters for generalized ellipsometry), thereby eliminating the need for the Kramers-Kronig integration. Because spectroscopic ellipsometry is normally performed at a large angle of incidence, the measurements are generally even more sensitive to surface overlayer or roughness effects, so it is essential to take surface effects into account when determining dielectric functions from the data. However, once this is done, the resulting dielectric functions are considerably more accurate than from reflectivity measurements. It is usually possible to determine the refractive index and extinction coefficient to better than 1% using spectroscopic ellipsometry measurements.

The data reduction techniques used to reduce spectroscopic ellipsometry data to complex refractive indices are fundamentally different from the data reduction discussed in Section 3.5. In Section 3.5, the data reduction was essentially over-determined; that is, there are many more data points then there are parameters that we wish to determine. In many cases where we are interested in determining optical functions, we measure at least two data points [such as $\mathrm{Re}(\rho)$ and $\mathrm{Im}(\rho)$] at each wavelength point, from which we wish to determine two values of the optical functions [ε_1 and ε_2, or n and k]. For uniaxial materials, the problem is again slightly over-determined, in that we measure six parameters [$\mathrm{Re}(\rho)$, $\mathrm{Im}(\rho)$, $\mathrm{Re}(\rho_{ps})$, $\mathrm{Im}(\rho_{ps})$, $\mathrm{Re}(\rho_{sp})$, and $\mathrm{Im}(\rho_{sp})$] and we want to determine four parameters [n_o, k_o, n_e, and k_e].

3.6.1 Optical Functions from Parameterization

In some cases, one is able to fit the spectroscopic ellipsometry data using a multilayer model, where the optical functions of the interesting layer can be parameterized. This often occurs when the material is either an insulator and/or an amorphous semiconductor where the dielectric functions can be modeled such as discussed in Section 3.4. In this case, the parameterization equations can be used directly to obtain the complex dielectric functions.

In the fitting process described in Section 3.5, film thicknesses, model parameters, and their associated error limits are approximated. Moreover, the curvature matrix α [see Eq. (3.52b)] is also obtained. Obviously, it is easy to obtain the spectroscopic dielectric function for a parameterized film of substrate by simply using the formula of the dielectric function. However, one also would like to have estimates of the error limits of the dielectric functions. Since the parameters obtained from the fitting procedure are correlated, one must use the error matrix α^{-1} to calculate the error. As an example, if it is assumed that there are three parameters used in the fitting procedure to describe the optical functions of a material, then the squared error for ε (both real and imaginary parts) is given by

$$\sigma_\varepsilon^2 = 4\left(\frac{\partial\varepsilon}{\partial a} \quad \frac{\partial\varepsilon}{\partial b} \quad \frac{\partial\varepsilon}{\partial c}\right)\begin{pmatrix} \alpha_{11}^{-1} & \alpha_{12}^{-1} & \alpha_{13}^{-1} \\ \alpha_{21}^{-1} & \alpha_{22}^{-1} & \alpha_{23}^{-1} \\ \alpha_{31}^{-1} & \alpha_{32}^{-1} & \alpha_{33}^{-1} \end{pmatrix}\begin{pmatrix} \dfrac{\partial\varepsilon}{\partial a} \\ \dfrac{\partial\varepsilon}{\partial b} \\ \dfrac{\partial\varepsilon}{\partial c} \end{pmatrix} \tag{3.53}$$

The elements of the α^{-1} matrix are the elements of the submatrix determined from the whole curvature matrix α by selecting the rows and columns associated with the three fitted parameters a, b, and c. The factor of 4 in Eq. (3.53) is included to determine the 90% confidence limits. If a different number of parameters is used in the model for the dielectric function, then the dimensions of the vectors and the error matrix in Eq. (3.53) will be the number of these parameters.

Although the error limit shown in Eq. (3.53) is rigorously valid only for linear problems with random errors, its calculation for optical function determination from spectroscopic ellipsometry data is an important part in the calculation in that it gives some estimate of the accuracy. Of course, if the value of the χ^2 from the fitting procedure described in Section 3.5 is significantly larger than 1, then any of these results are suspect, and the error limits obtained from Eq. (3.53) are not meaningful. If this procedure

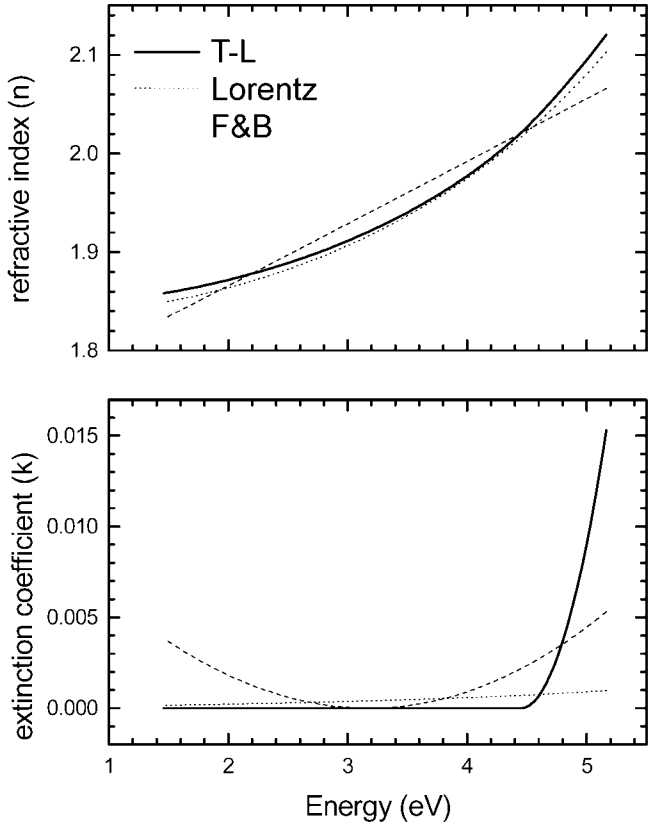


Figure 3.5 The real and imaginary parts of the complex refractive index for SiN obtained from the three parameterizations used to fit the data in Fig. 3.3 and tabulated in Table 3-1.

IS used when χ^2 is much greater than 1, then one should at least multiply Eq. (3.53) by a penalty function (such as the final value of χ^2), which will at least increase the error limits of the estimates of the dielectric functions.

Figure 3.5 shows a plot of the refractive index n and extinction coefficient k obtained for the amorphous SiN film grown on silicon from the spectroscopic ellipsometry data shown in Fig. 3.3. Recall that only the Tauc-Lorentz model fit the experimental data, so the n and k obtained from the other fits are suspect. The error limits in both n and k for the Tauc-Lorentz fit were ~0.002, while the error limits for the other two models were significantly higher if one uses the penalty function described above.

Part of the reason as to why the Lorentz and the Forouhi and Bloomer functions did not work so well can be seen from the k plot shown

in Fig. 3.5. The Tauc-Lorentz function is constrained to have the value of $\varepsilon_2 = 2nk = 0$ for energies below the band gap, while the other two functions are not. For very thick films, even a small value of k can contribute a significant amount of optical absorption for light traversing the film twice. There is also very little difference between the Tauc-Lorentz model and the Lorentz model in the dispersion of n, but there are significant differences between these two and the n obtained from the Forouhi and Bloomer model.

3.6.2 Newton-Raphson Algorithm

The reduction of spectroscopic ellipsometry data to optical functions can be mathematically expressed as finding the roots of the equations:

$$\rho_r(\phi, \lambda_j\, d_1, d_2, \ldots, d_N, n_{aj}, k_{aj}, n_{1j}, k_{1j}, n_{2j}, k_{2j}, \ldots, n_{Nj}, k_{Nj}, n_{sj}, k_{sj}) = 0 \tag{3.54a}$$

$$\rho_i(\phi, \lambda_j\, d_1, d_2, \ldots, d_N, n_{aj}, k_{aj}, n_{1j}, k_{1j}, n_{2j}, k_{2j}, \ldots, n_{Nj}, k_{Nj}, n_{sj}, k_{sj}) = 0 \tag{3.54b}$$

at each wavelength point λ_j where we have assumed that there are N films over a substrate. The measured spectra are given by the real and imaginary parts of ρ [ρ_r, and ρ_i, see Eq. (3.2)] and the angle of incidence is given by ϕ. Each film is characterized by a film thickness d_i and a set of optical functions n_{ij} and k_{ij}, where the i index spans the number of films ($i = 1$ to N) and the j index spans the number of wavelength points in the spectrum. In this problem, two quantities are measured, so up to two parameters can be determined, and all the other parameters in Eqs. (3.54) must be specified.

Since Eqs. (3.54) are inherently nonlinear in all the listed parameters, there are no good, universally convergent algorithms to solve these equations. If one can "guess" at values for the parameters to be solved that are reasonably close to the final solution, then one can use the Newton-Raphson algorithm to converge to the final solution. However, one must always be aware that this convergence is never guaranteed. It is possible that the initial guess is far enough off from the real solution that the algorithm will not converge, or will converge to a false solution. Moreover, it is possible that there is no solution to Eqs.(3.54) in the neighborhood of the initial guess. The reader is referred to the excellent discussion of this in *Numerical Recipes*, Chapter 9.6.

Assume that the problem is the much simpler case of a single film over a substrate, where we wish to determine the optical functions of the

substrate from the data, given that the ambient is air ($n_a = 1$, $k_a = 0$), and the surface overlayer optical functions are known. In this case, Eqs. (3.54) simplify to

$$\rho_r(\phi, \lambda_j\, d_1, n_{1j}, k_{1j}, n_{sj}, k_{sj}) = 0 \tag{3.55a}$$

$$\rho_i(\phi, \lambda_j\, d_1, n_{1j}, k_{1j}, n_{sj}, k_{sj}) = 0 \tag{3.55b}$$

where the only unknowns are n_{sj} and k_{sj}. If these two equations are solved for n_{sj} and k_{sj} at each wavelength point λ_j then we will have determined the optical functions of the substrate. Equations (3.55) can be vectorized

$$\rho(\boldsymbol{n}) = \mathbf{0} \tag{3.56}$$

where $\boldsymbol{n} = (n_s, k_s)$, $\rho = (\rho_r, \rho_i)$, and references to the angle of incidence, wavelength, film thicknesses and refractive indices are assumed to be known and have been suppressed. Define the Jacobian $\mathbf{J}$

$$\mathbf{J} = \begin{pmatrix} \dfrac{\partial \rho_r}{\partial n} & \dfrac{\partial \rho_r}{\partial k} \\ \dfrac{\partial \rho_{ir}}{\partial n} & \dfrac{\partial \rho_{ir}}{\partial k} \end{pmatrix}. \tag{3.57a}$$

The next iteration of the solution vector is then given by

$$\boldsymbol{n}_{new} = \boldsymbol{n}_{old} + \boldsymbol{\delta n}, \tag{3.57b}$$

where

$$\boldsymbol{\delta n} = -\mathbf{J}^{-1}\rho. \tag{3.57c}$$

As with the Levenberg-Marquardt algorithm, the Newton-Raphson algorithm also needs a stopping criterion. If the algorithm does converge and Eq. (3.56) has a solution, then the stopping criterion can be as simple as stopping the iteration when all elements of the $\boldsymbol{\delta n}$ vector are smaller than a chosen error (such as $<10^{-6}$). However, the Newton-Raphson algorithm may *not* converge, depending upon the model, the data, and the starting solution. When this occurs, $\boldsymbol{\delta n}$ may not approach a very small number, indicating algorithm failure. In some other situations, the Newton-Raphson algorithm does converge, but a false solution is obtained because the starting solution was wrong.

3.6.3 Optical Functions of Bulk Isotropic Semiconductors and Insulators

One of the most useful applications of spectroscopic ellipsometry has been the determination of optical functions for a variety of bulk isotropic materials. Naively, one might just interpret the spectroscopic ellipsometry data assuming a simple air/material model. Such an approach has an exact solution, where the dielectric function is given by

$$\varepsilon = \varepsilon_1 + i\varepsilon_2 = \sin^2(\phi)\left\{1 + \left[\frac{1-\rho}{1+\rho}\right]^2 \tan^2(\phi)\right\}, \tag{3.58}$$

where the angle of incidence is given by ϕ and the quantity ρ is defined in Eq. (3.10).

The major problem with this approach is that it ignores the surface layer of the material. All materials have this surface overlayer, which may due to surface roughness, surface oxide, surface reconstruction, etc. Therefore, any realistic model of the sample is more complicated than a simple air/material, and Eq. (3.58) is not valid for any model involving a surface overlayer. However, Eq. 3.58 is quite useful as a limiting case, and often spectroscopic ellipsometry data are transformed from the ρ representation into the pseudo-dielectric function representation using Eq. (3.58). As an example, Figure 3.6 shows the pseudo- refractive index and pseudo-extinction coefficient for silicon covered with 0, 0.8, and 2.0 nm of SiO_2, respectively.

If a very thin, transparent film covers the substrate, then Drude has shown that the measured value of Δ is a function of both the n and k of the substrate, as well as the thickness of the thin film. Drude's expression is given by

$$\Delta = \Delta_o(n_s, k_s) + K\frac{4\pi d}{\lambda}\,\frac{n_f^2 - 1}{n_f^2}. \tag{3.59}$$

The refractive index of the film is given by n_f, and the refractive index and extinction coefficient of the substrate is given by n_s and k_s. The thickness of the film is given by d and the wavelength of light by λ. For a transparent substrate, $\Delta_o = 0°$ or 180°, depending upon whether the angle of incidence is greater than or less than the pseudo-Brewster angle of the substrate. For a substrate with a small value of k, then Δ_o will be x or $180° - x$ where x is a small angle, again depending upon the value of the angle of incidence. Drude also calculated the effect of ψ, but this parameter is much less sensitive to very thin films.

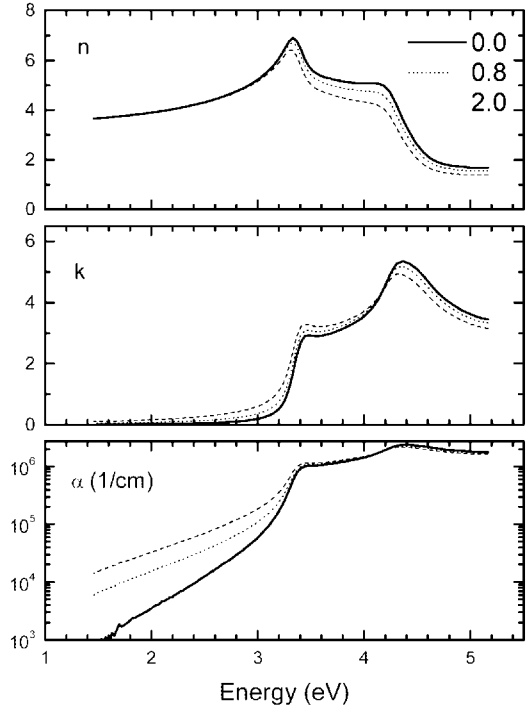


Figure 3.6 Three pseudo-optical functions of silicon for three different values of oxide thickness (in nm). The values of n and k for 0 thickness are the actual optical functions of silicon (from Ref. 20).

Equation (3.59) suggests a strategy for determining the optical properties of semiconductors and insulators from spectroscopic ellipsometry data. The first step is to restrict the analysis range of the spectroscopic ellipsometry data below the band edge of the material, where it is known that $k_s = 0$, and therefore $\Delta_o = 0°$ or 180°. Using some suitable model for the refractive index of the surface overlayer, calculate the thickness of the overlayer from the spectroscopic ellipsometry data. Often, the suitable model is a three-media model consisting of air/overlayer/substrate. The optical properties of the substrate are parameterized using the Sellmeier approximation [Eq. (3.36)] and the optical properties of the overlayer are approximated either as a 50% air/50% substrate Bruggeman effective medium or as a surface oxide.

Once the parameterized model for the surface overlayer has been selected and its thickness determined, then the problem simplifies to solving two equations [Re(ρ) and Im(ρ)] for two unknowns (n_s and k_s) at each wavelength, where the Newton-Raphson algorithm described above can

be used. Moreover, errors in the experimental data ρ, the angle of incidence, and estimated errors in the overlayer thickness and optical functions can and should be used to estimate the propagated error of n_s and k_s.

In using this technique, it is extremely important to measure the imaginary part of ρ very accurately when ρ is near zero, since the second term in Eq. 3.59 can be quite small and must be measured accurately. In addition, sample preparation is also important in that samples with small overlayer thicknesses require less correction and therefore result in smaller propagated errors in n_s and k_s.

Figure 3.6 (for 0 overlayer thickness) shows the values of the real and imaginary parts of the dielectric function of silicon determined from polarization modulation ellipsometry[20] (PME, described in Chapter 6). Since PME experiments can measure N, S, and C, then Δ can be measured accurately at all values, and the correction term given in Eq. (3.59) can be measured very accurately. The model used to determine the optical functions shown in Fig. 3.6 was air/SiO_2/c-Si, where the SiO_2 layer was estimated to be ~7A thick. The error limits are equally important, and are shown in Fig. 3.7, where random errors, as well as systematic errors in the alignment, wavelength, the angle of incidence and surface overlayer thickness were propagated into the final error estimate. Silicon is an indirect band gap material, so the value of k_s is small below the direct band edge, but it is not zero. Therefore, the value of Δ_o is $0° + x$ or $180° - x$ where x is a small value. The value of Δ_0 can be determined from the known values of the absorption coefficient, determined from transmission measurements.

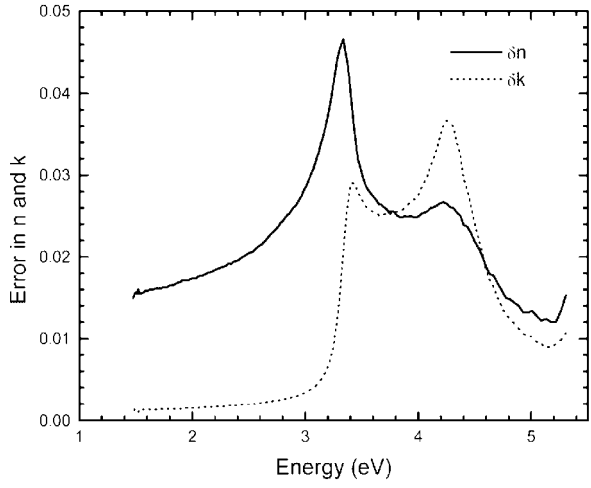


Figure 3.7 The error in the optical functions of silicon shown for the 0.0 nm oxide thickness shown in Fig. 3.6 (from Ref. 20).

Another approach is used by workers using rotating analyzer ellipsometry, where measurements of Δ near 0° or 180° are often not very accurate. The absorption coefficient of Si near 4.25 eV (the E_2 peak) is nearly 2×10^6 /cm, indicating that most of the light is absorbed in the top 5 nm. If one examines the dependence of the peak in the imaginary part of the pseudo-dielectric function given in Eq. (3.58), one finds that the value of $<\varepsilon_2>$ increases as the thickness of the overlayer decreases. Therefore, one can select cleaning and etching procedures that minimize this overlayer simply by maximizing $<\varepsilon_2>$. While this procedure does indicate the best possible surface (that is, the surface that is the cleanest), no metric is given that indicates the thickness of the remaining surface overlayer.

In spite of the importance of accurate values of the optical functions of silicon, there still is considerable controversy concerning the best values. For example, accurate thin film metrology on silicon devices relies on accurate values of the n and k of silicon. However, the above discussion points to some of the difficulties involved in getting extremely accurate values of the optical functions of silicon. First of all, it is not possible to get a perfectly clean silicon surface. Either the surface is oxidized or, if some cleaning procedures are used, the surface is hydrogen terminated. Moreover, surface oxides, surface roughness, and surface reconstructions are dependent upon the face of silicon: the (100) face is different from the (110) face, which is also different from the (111) face. The determination of very accurate values of the optical functions of any material remains an extremely difficult problem.

3.6.4 Optical Functions of Anisotropic Materials

The optical functions of anisotropic materials can also be determined by spectroscopic ellipsometry, but the analysis is considerably more difficult. Generally, each spectroscopic ellipsometry measurement will result in the determination of up to six parameters (the real and imaginary parts of ρ_{pp}, ρ_{sp}, and ρ_{ps}) at each wavelength. Additional data can be obtained by examining different angles of incidence and different orientations of the sample. For uniaxial crystals, there are four independent parameters to determine: n_o, n_e, k_o, and k_e. Therefore, the problem is over-determined, and a procedure such as described in Section 3.5 must be used.

For a simple air/sample model, certain symmetries to the measured data apply.

1) If the sample is uniaxial and the optic axis is in the plane of incidence, then the cross-polarization terms (ρ_{sp}, and ρ_{ps}) will be zero.

2) If the sample is uniaxial and the optic axis is perpendicular to the plane of incidence, then the cross-polarization terms will also be zero.
3) If the sample is uniaxial and the optic axis is in the plane of the sample ($\theta = 90°$), then $\rho_{sp} = \rho_{ps}$ and $\rho_{sp}(\varphi) = -\rho_{sp}(-\varphi)$, $\rho_{ps}(\varphi) = -\rho_{ps}(-\varphi)$, where φ is the azimuthal angle of the optic axis with respect to the plane of incidence.

Although these symmetries hold for the simple air/sample model, they do not necessarily hold if there is a surface overlayer over the sample, which might also be anisotropic.

Spectroscopic ellipsometry studies of anisotropic materials will be discussed in more detail in Chapter 9 by Schubert.

3.6.5 Optical Functions of Thin Films

It is possible to determine the optical functions of thin film materials from spectroscopic ellipsometry measurements, even if it is not possible to find a parametric model that fits the data. The procedure is essentially the same as for bulk materials, once the correct model for the sample is obtained. In this case, the optical functions to be solved are n_f and k_f instead of the substrate optical functions n_s and k_s.

The most difficult problem in determining the optical functions of thin film materials is arriving at a reasonable model for the near-surface region. In some cases, the surface roughness is so thick that it is not possible to extract optical functions for the underlying material, whether it is film or bulk. If the surface roughness is >25 nm thick and the absorption coefficient of a material is $\sim 10^6$ cm^{-1} at some wavelengths, then all of the light is absorbed in the surface roughness region and very little penetrates to the actual material, making the determination of the optical functions of the underlying material nearly impossible.

In some cases, the best solution is to reduce the surface roughness using a polishing technique. This was done with several samples of thin-film silicon grown on oxidized silicon by standard chemical vapor deposition.[21] Depending upon growth conditions, the silicon film can be amorphous, fine-grain polycrystalline, or large-grain polycrystalline. Post-processing of the film can re-crystallize it, creating a nearly single crystal film of silicon.

Performing a spectroscopic ellipsometry measurement on such a sample will result in data such as is shown in Fig. 3.8. At low photon energies (below ~2.6 eV for the data shown in Fig. 3.8), the thin-film silicon is

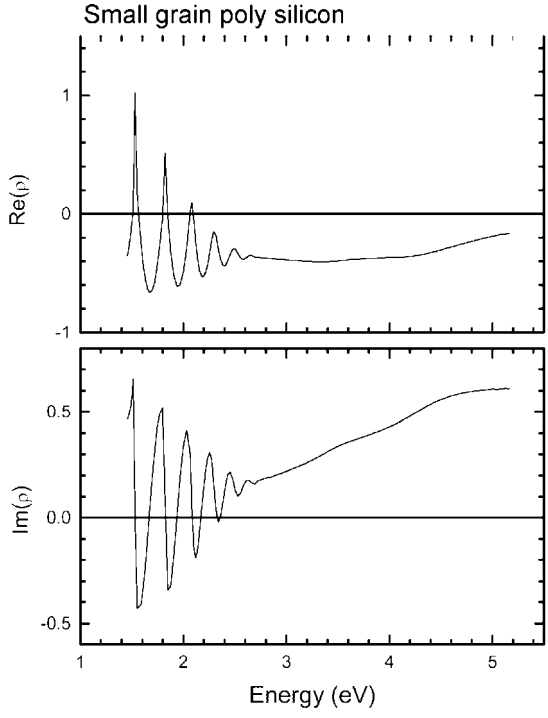


Figure 3.8 The real and imaginary part of complex ρ from a spectroscopic ellipsometry experiment on small-grain polycrystalline silicon grown on oxidized silicon.

semi-transparent, resulting in several oscillations in the data. At higher photon energies, the light is absorbed in the surface roughness and thin-film silicon so no interference fringes are observed.

These observations suggest a strategy to determine the optical functions for thin-film silicon from data such as shown in Fig. 3.8:

1) Restrict the analysis region to the wavelength region of the oscillations and parameterize the optical functions of the thin film. For the data shown in Fig. 3.8, one might use a five-layer model:
   a) Air,
   b) Surface roughness modeled with a 50% air, 50% film Bruggeman effective medium approximation,
   c) Tauc-Lorentz model for the film,
   d) Lorentz model for the amorphous SiO_2,
   e) Crystalline silicon for the substrate.

Fit the data in the restricted wavelength region to the above model, keeping only the values of the thicknesses of the surface roughness, thin-film silicon and the amorphous SiO_2 layer. Also, keep the fitting parameters for the Lorentz model for the amorphous SiO_2.

2) Using the thicknesses and model parameters obtained in step 1), re-analyze the spectroscopic ellipsometry data, solving for n_f and k_f.

The values of n_f and k_f for various forms of thin-film silicon are shown in Fig. 3.9. Note that error limits in n_f and k_f (not shown) can be also be determined in a similar manner as for bulk materials.

This whole procedure works because many of the parameters determined in step 1) are nearly uncorrelated with the other parameters in the fitting procedure. For example, the surface roughness thickness, using the

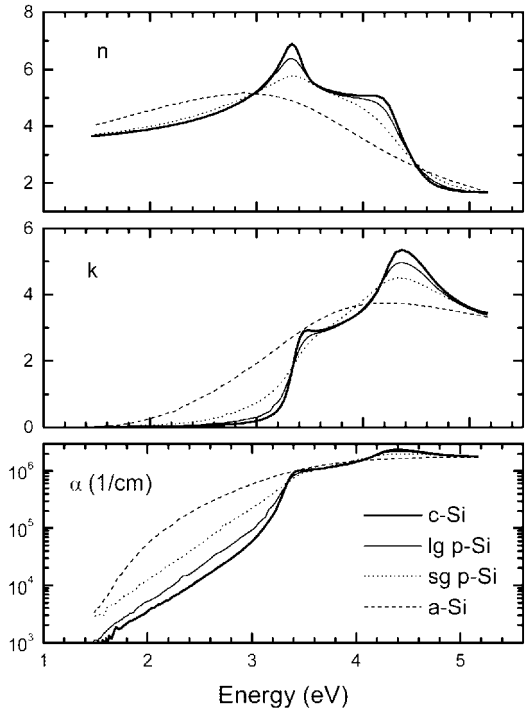


Figure 3.9 The optical functions of several forms of thin film silicon. The values for crystalline silicon are from measurements on bulk samples and are included for reference (from Ref. 21).

50% air, 50% film Bruggeman effective medium approximation, affects the average value of Im(ρ). If there is no surface roughness, then the oscillations in Im(ρ) will be symmetric about 0; any increase in surface roughness will increase the average values Im(ρ). The silicon film and oxide film thicknesses will affect the experimental spectrum in a complicated way, but will be relatively uncorrelated with the other parameters. However, the parameters used to model the optical functions of the thin-film silicon in step 1 can be significantly correlated, but this is not an issue since the values of n_f and k_f are smoothly varying functions of wavelength and the error limits (as determined by the procedure described above in Section 3.6.1 using the error matrix) of n_f and k_f are not large. Note that the values of the optical functions of the SiO_2 layer must only be reasonably valid over the interference region (~1.5 to ~2.8 eV in Fig. 3.8), since no light penetrates to the thin-film silicon/SiO_2 interface at higher energies. Above the interference region, the only parameter that matters is the thickness of the surface roughness.

3.7 Depolarization

Up to this point, it has been assumed that the sample does not depolarize the light beam, and that Eqs. (3.7) can be used to determine the sample Mueller matrix from the sample Jones matrix. For many situations normally encountered in spectroscopic ellipsometry measurements, this is a good assumption. However, there are several situations where the sample does depolarize the light beam. In these cases, there is no rigorous connection between the Jones matrix formalism and the Mueller matrix formalism.

Some classic examples of sample depolarization are:

1) Nonuniform film thickness. If a film on the sample has a nonuniform thickness over the spot size of the ellipsometer, then the total beam will be quasi-depolarized.[22] That is, the light from one part of the incident beam will experience a different film thickness than light from another part of the beam, and will therefore experience a different phase shift upon reflection. The integration of the total light beam performed by the detector will result in a quasi-depolarization.
2) Transparent substrate. If the substrate is transparent, then it is possible for some of the light to be reflected off the back surface of the substrate into the detector.[23] Since the thickness of the average substrate will be

~1000 wavelengths of light, and there will most likely be a very small variation in the thickness of the substrate over the spot size, different positions within the sampling beam will experience different phase shifts. Again, the integration of the total light beam at the detector will result in quasi-depolarization.

3) Rough samples. Very rough surfaces can also depolarize the light beam[24]. This type of depolarization is not noticeable if the roughness is much less than the wavelength of light, where the optical functions can often be approximated using an effective medium approximation of the constituents. If the roughness scale is much larger than the wavelength of light, then a significant fraction of the light will be scattered out of the collection cone of the ellipsometer. However, roughness scales on the order of the wavelength of light become randomly diffractive. This type of roughness may result in some depolarization, and some cross polarization.
4) Monochromator/Spectrograph depolarization. Monochromators and spectrographs are not perfectly monochromatic devices, but rather integrate over a range of wavelengths. Moreover, stray light and 2nd order effects can add light at the detector that is significantly different in wavelength than the test wavelength. Again, light beam integration results in quasi-depolarization.

Several experimental remedies will minimize or nearly eliminate the depolarization effects of 1) through 4) above. If the nonuniformity of the film thickness of a sample is large, then focussing lenses can be used to sample a much smaller area of the sample. Of course, this can lead to errors from the integration of the angles of incidence and from uncorrected strain of the lenses. The effects of transparent substrates can be minimized by roughening the back surface, wedging the substrate so that the back-reflected light is not included in the collection cone of the ellipsometer, or by using index-matching fluid to couple optically the substrate with a thicker piece of similar material. Monochromator/ spectrograph depolarization can be minimized by reducing the slit widths and by minimizing the light intensity from stray light and 2nd order effects using appropriate filters.

Even if the light beam is partially depolarized from the sample, it still may be possible to correct for these effects if the amount of depolarization is measured. If the sample acts as a simple depolarizer, then its Mueller matrix can be given by (see Brouseau)

$$\mathbf{M} = \beta\mathbf{M}_J + (1-\beta)\mathbf{D} = \beta\mathbf{M}_J + (1-\beta)\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.60a}$$

where $\mathbf{M}_J$ is the Mueller-Jones matrix of the non-depolarizing sample and β is the fraction of polarized light in the light beam. If the sample is isotropic, then Eq. (3.60a) becomes

$$\mathbf{M} = \begin{pmatrix} 1 & -\beta N & 0 & 0 \\ -\beta N & \beta & 0 & 0 \\ 0 & 0 & \beta C & \beta S \\ 0 & 0 & -\beta S & \beta C \end{pmatrix}. \tag{3.60b}$$

The measured values of N_m, S_m, and C_m are βN, βS, and βC, respectively, the sum of the squares are

$$(N_m^2 + S_m^2 + C_m^2) = \beta^2. \tag{3.60c}$$

Therefore, the measurement of all three quantities N_m, S_m, and C_m is essential for depolarizing samples, in that it is possible to obtain a measure of the fraction of polarized light reflected from the sample.

Another type of depolarization occurs when the light beam, because of its finite size, samples a range of film thicknesses. In this case, the measured sample Mueller matrix will be

$$\mathbf{M} = \begin{pmatrix} 1 & -N_m & 0 & 0 \\ -N_m & 1 & 0 & 0 \\ 0 & 0 & C_m & S_m \\ 0 & 0 & -S_m & C_m \end{pmatrix}. \tag{3.61}$$

Each of the parameters in Eq. (3.61) is given by an integration over the distribution of film thicknesses over the size of the spot. For example, if the film thickness varies from D_1 to D_2 over the spot size, then the parameter N_m will be given by:

$$N_m = \int_{D_{11}}^{D_2} f(D)N(D)dD, \tag{3.62}$$

where $f(D)$ is the probability distribution function describing the film thickness variation and $N(D) = \cos(2\psi(D))$. Similar expressions can be determined for S_m and C_m. Equation (3.60c) is still valid, and $\beta < 1$.

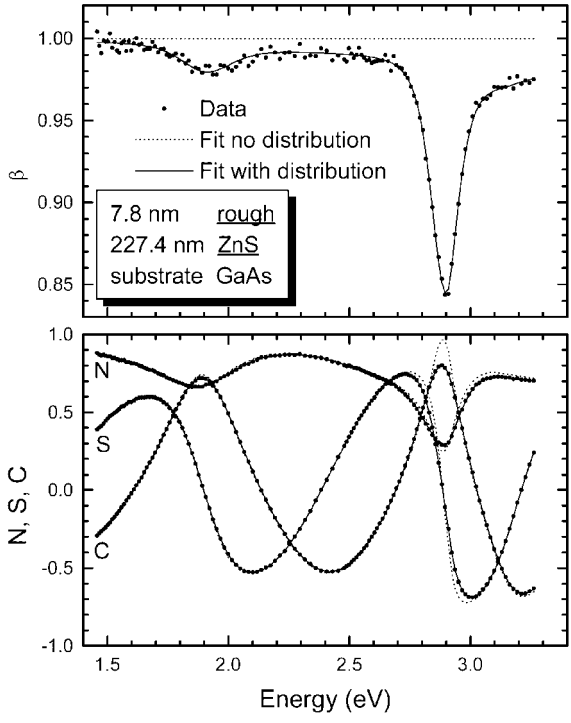


Figure 3.10 The measured ellipsometry parameters *N*, *S*, and *C* for a thin film of ZnS grown on GaAs. Two fits to the data are shown: one with a distribution of ZnS film thicknesses, and the other with a constant thickness of the ZnS film (from Ref. 22).

An example of the effects of this latter type of depolarization[22] on spectroscopic ellipsometry data is shown in Fig. 3.10. A thin film of ZnS was grown on GaAs using pulsed laser ablation. Laser ablation film growth is well known for producing nonuniform film thicknesses, unless several precautions are taken. As can be seen from Fig. 3.10, there is a small dip in β near 1.9 eV, and a much larger dip in β near 2.9 eV. The depolarization is in fact limited to wavelength regions were the measured quantities N_m, S_m, and C_m vary rapidly with wavelength.

The data in Fig. 3.10 was fit to a model that included sample depolarization from non-uniform film thickness. For this fitting procedure, the quantities N_m, S_m, and C_m were fit directly and not the transformed ρ. In order to incorporate depolarization, a distribution of the ZnS film thickness was simulated using a "top hat" probability distribution function. If no depolarization mechanism is included, then the fitted line shown by the dashed line in Fig. 3.10 was obtained, while the inclusion

of depolarization resulted in the solid line in Fig. 3.10. Moreover, the χ^2 was 7.24 for the fit without the film thickness distribution compared to 1.07 for the fit with the film thickness distribution. The distribution of film thicknesses was 227.4 ± 8.9 nm; therefore, a variation of 3.9% in film thickness was responsible for a ~17% depolarization at 2.9 eV! If the film thickness is larger, then even smaller variations in film thickness can result in noticeable decreases in β. In ref. 22, 1.6% variation in film thickness of a 828 nm film resulted in a depolarization as large as 30%!

This type of analysis can also be performed for other types of simple depolarization, such as listed above. However, it is essential to have a measure of the depolarization, which requires that three parameters (such as the N, S, and C parameters) be measured. If only two parameters are measured, small depolarizations such as discussed above will go undetected, although it is possible to detect large depolarizations with conventional rotating analyzer ellipsometers.[25]

Acknowledgements

The author would like to thank Frank Modine, Blaine Johs, and Mathias Schubert for reading and commenting on this chapter. This research was sponsored in part by the Division of Materials Sciences, U. S. Department of Energy under contract No. DE-AC05-00OR22725 with Lockheed Martin Energy Research, Inc.

3.8 Further Reading and References

Optics and Ellipsometry

R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light,* (North-Holland, Amsterdam, 1977).

M. Born and E. Wolf, *Principles of Optics*, *Sixth Edition* Cambridge, University Press, 1980.

P. S. Hauge, "Recent developments in instrumentation in ellipsometry," *Surf. Sci.* **96**, 108–140 (1980).

E. Palik (ed.), *Handbook of optical constants of solids,* Academic Press, Orlando, 1985; *Handbook of optical constants of solids, II* Academic Press, Boston, 1991; *Handbook of optical constants of solids, III* Academic Press, Boston, 1998.

P. Yeh, *Optical Waves in Layered Media*, Wiley, New York, (1988).

D. S. Kliger, J. W. Lewis, and C. E. Randall, *Polarized Light in Optics and Spectroscopy*, (Academic Press, New York) (1990).

H. G. Tompkins, *A User's Guide to Ellipsometry*, Academic Press, New York, 1993.
H. G. Tompkins and W. A. McGahan, *Spectroscopic Ellipsometry and Reflectometry: A User's Guide*, John Wiley & Sons, New York, 1999.
Spectroscopic Ellipsometry, *Thin Solid films,* **233–234** (1993); *Spectroscopic Ellipsometry*, *Thin Solid films,* **313–314** (1998). (Proceedings of the first and second International Conferences on Spectroscopic Ellipsometry)
R. A. Chipman, "Polarimetry" *Handbook of Optics, Vol. II*, (2nd edition, McGraw-Hill, New York, 1995), Chapter 22.
R. M. A. Azzam, "Ellipsometry" *Handbook of Optics, Vol. II*, (2nd edition, McGraw-Hill, New York, 1995), Chapter 27.
U. Rossow and W. Richter, "Spectroscopic ellipsometry," *Optical characterization of epitaxial semiconductor layers,* (Springer-Verlag, Berlin, G. Bauer and W. Richter, eds. 1996), p. 68–128.
C. Brosseau, *Fundamentals of Polarized Light: A Statistical Optics Approach*, Wiley, New York, 1998.

Data Reduction

P. R. Bevington, *Data Reduction and Error Analysis for the Physical Sciences,* McGraw-Hill, New York, 1969.
Y. Bard, *Nonlinear Parameter Estimation,* Academic Press, New York, 1970.
N. R. Draper and H. Smith, *Applied Regression Analysis,* Wiley, New York, 1981.
W. H. Press, B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes, Second Edition,* Cambridge, University Press, 1992.

Numbered References

1. R. Barakat, "Bilinear constraints between elements of the 4 × 4 Mueller-Jones transfer matrix of polarization theory," *Opt. Commun.*, **38** 159 (1981).
2. G. E. Jellison, Jr. "The calculation of thin film parameters from spectroscopic ellipsometry data," *Thin Solid Films*, **290–291**, 40–45 (1996).
3. F. Abelès, *Ann. de Physique,* **5**, 596 (1950).
4. D. W. Berreman, "Optics in stratified and anisotropic media: 4 × 4-matrix formulation," *J. Opt. Soc. Am.*, **72**, 502–510 (1972).
5. P. J. Lin-Chung and S. Teitler, "4 × 4 matrix formalisms for optics in stratified anisotropic media," *J. Opt. Soc. Am. A*, **1**, 703–705 (1984).
6. H. Wöhler, G. Hass, M. Fritsch, and D. A. Mlynski, "Faster 4 × 4 matrix method for uniaxial inhomogeneous media," *J. Opt. Soc. Am., A* **5**, 1554 (1988); H. Wöhler, M. Fritsch, G. Hass, and D. A. Mlynski, "Characteristic matrix method for stratified anisotropic media: optical properties of special configurations," *J. Opt. Soc. Am., A*, **8**, 536–540 (1991).

7. R. Jacobsson, in E. Wolf (ed.) *Progress in Optics*, 5, Wiley, 1965, p. 247.
8. A. R. Forouhi and I. Bloomer, "Optical dispersion relations for amorphous semiconductors and amorphous dielectrics," *Phys. Rev. B*, **34,** 7018–7026 (1986).
9. G. E. Jellison, Jr. and F. A. Modine, "Parameterization of the optical functions of amorphous materials in the interband region," *Appl. Phys. Lett.*, **69**, 371–373 (1996); *ibid* **69**, 2137 (1996).
10. J. Tauc, R. Grigorovici, and A. Vancu, *Phys. Stat. Solidi*, **15**, 627 (1966).
11. S. Adachi, "Optical dispersion relations in amorphous semiconductors," *Phys. Rev. B*, **43**, 12316–12321 (1991).
12. T. Yamaguchi, Y. Kaneko, A. H. Jayalissa, Aoyama, A. vb. Zotov, and V. G. Lifshits, "Empirical dielectric function of amorphous materials for spectroscopic ellipsometry," *J. Appl. Phys.*, **77**, 4673–4676 (1995).
13. D. E. Aspnes, "Modulation spectroscopy/ electric field effects on the dielectric function of semiconductors," *Handbook on Semiconductors, Vol 2*, (North Holland, New York, M. Balkanski, ed. 1980), p. 109–154; P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors,* (Springer-Verlag, Berlin, 1996).
14. F. L. Terry, Jr. "A modified harmonic oscillator approximation scheme for the dielectric constants of $Al_xGa_{1-x}As$," *J. Appl. Phys.*, **70**, 409–417 (1991); P. G. Snyder, J. A. Woollam, S. A. Alterovitz, and B. Johs, *J. Appl. Phys.*, **68**, 5925 (1990).
15. R. J. Elliott, *Phys. Rev.* **108**, 1384 (1957); T. Holden, P. Ram, F. H. Pollak, J. L. Freeouf, B. X. Yang, and M. C. Tamargo, *Phys. Rev. B*, **56**, 4037 (1997); G. E. Jellison, Jr. and L. A. Boatner, "Optical functions of uniaxial ZnO determined by generalized ellipsometry," *Phys. Rev. B*, **58**, 3586–3589 (1998); *ibid*, **65**, 049902(E) (2002).
16. B. Johs, C. M. Herzinger, J. H. Dinan, A. Cornfield, and J. D. Benson, "Development of a parametric optical constant model for $Hg_{1-x}Cd_xTe$ for control of composition by spectroscopic ellipsometry during MBE growth," *Thin Solid Films*, **313–314**, 137–142 (1998).
17. D. E. Aspnes, "Optical properties of thin films," *Thin Solid Films*, **89**, 249–262 (1982).
18. P. H. Roussel, J. Vanhellemont, H. E. Maes, "Numerical aspects of the implementation of effective-medium approximation models in spectroscopic ellipsometry regression software," *Thin Solid Films*, **234**, 423–427 (1993).
19. S. Y. Kim and K. Vedam, "Proper choice of the error function in modeling spectroellipsometric data," *Appl. Opt,* **25**, 2013–2021 (1986).
20. G. E. Jellison, Jr., "Optical functions of silicon determined by two-channel polarization modulation ellipsometry," *Opt. Materials*, **1**, 41–47 (1992).
21. G. E. Jellison, Jr., M. F. Chisholm, and S. M. Gorbatkin, "Optical functions of vapor-deposited thin-film silicon determined by spectroscopic ellipsometry," *Appl. Phys. Lett.*, **62**, 3348–3350 (1993).
22. G. E. Jellison, Jr. and J. W. McCamy, "Sample depolarization effects from thin films of ZnS on GaAs as measured by spectroscopic ellipsometry," *Appl. Phys. Lett.*, **61**, 512–514 (1992).

23. R. Joerger, K. Forcht, A. Gombert, M. Kohl, and W. Graf, "Influence of incoherent superposition of light on ellipsometric coefficients," *Appl. Opt.*, **36**, 319–327 (1997).
24. S. Krishnan, and P. C. Nordine, Mueller-matrix ellipsometry using the division-of-amplitude photopolarimeter: a study of depolarization effects," *Appl. Opt.*, **33**, 4184–4192 (1994).
25. S. Pittal, P. G. Snyder, and N. Ianno, "Ellipsometry study of non-uniform lateral growth of ZnO thin films," *Thin Solid Films*, **233**, 286–288 (1993).

Instrumentation

4 Optical Components and the Simple PCSA (Polarizer, Compensator, Sample, Analyzer) Ellipsometer

Harland G. Tompkins

Chandler, AZ

4.1 General

Ellipsometry was practiced for about 10 years in the 19th century[1] and for the entire 20th century. The field of optics, itself, is much older than this, and many of the components which are currently used have been available for a quite long time. On the other hand, some of the components were developed only recently.

We will deal with methods of obtaining polarized light and shall discuss how the phase difference can be induced to change linearly polarized light into elliptically polarized light. The ellipsometry variables are often wavelength-dependent. Spectroscopic ellipsometry (SE) considers the contribution from many individual wavelengths (or energies) and there are several methods of sorting out the contributions from each wavelength (or energy). The angle-of-incidence is always non-normal, and we shall discuss goniometers briefly.

When polarized light is reflected from a surface at non-normal incidence, a phase shift may be induced, changing the ellipticity of the polarized light. This change is dependent on the material properties of the surface. It is the essence of ellipsometry to analyze the reflected light in order to determine what the surface did to the light beam and to hence draw conclusions about the material.

In order to do this analysis, the polarization state generator often includes the following:

- light source
- polarizer
- phase inducer which may be one of the following
    - none
    - quarter-wave plate
    - compensator
    - phase modulator

Following the reflection, the polarization state detector often includes the following

- phase inducer (see above)
- polarizer (called an "analyzer")
- detector

Since it is essential that the effects of the reflection be analyzed at a single wavelength (or energy) or at many individual wavelengths (or energies), a method of separating white light into its individual components is required. In some cases, this device is incorporated into the light source itself. In other cases, it is incorporated into the detector.

In Fig. 4.1, we show an early optical instrument shown by Drude[2] in his book of 1901. This instrument was not called an "ellipsometer" since this name would not be used until 1945 by Alexandre Rothen,[3] but the instrument has most of the elements of an ellipsometer. The light source is not shown, and the detector was the eyeball of the observer. The polarizer,

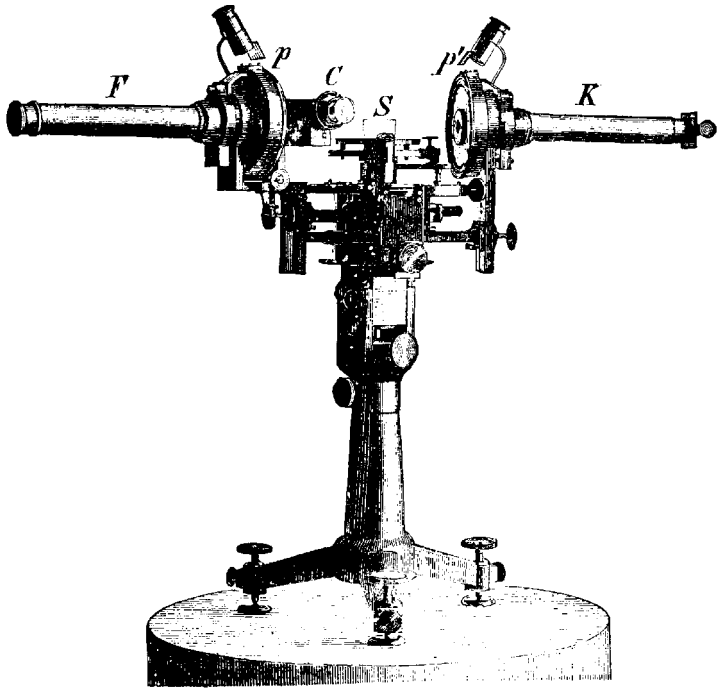


Figure 4.1 Early optical instrument used for studying changes in polarization upon reflection. Telescopes F and K, polarizers p and p′, compensator C, and reflecting surface S. (After Drude[2]).

analyzer, compensator, and reflecting surface can be seen. In this chapter, we shall discuss the devices used in order to obtain polarized light, phase shifters, monochromators, and goniometers.

4.2 The Components

4.2.1 Methods of Obtaining Polarized Light

There are several methods of causing a light beam to be polarized. Perhaps the simplest method is by reflection. This was discovered by a French military officer, E. Malus in about 1808[4] and fully explained by Brewster (circa 1830).[5]

Another method is by dichroism. In one manifestation, organic materials are imbedded into a plastic sheet and then stretched. This is then dyed and the dye molecules selectively attach themselves to the aligned molecules. This causes the absorption of light to be high in one plane and low in another. This is the polarization mechanism of large-area sheet polarizers. In another manifestation, passing light through the crystal tourmaline was observed to give polarized light.[6]

Yet another method is to fabricate a grid of conductive "wires", which are spaced parallel to each other and separated by distances which are comparable to the wavelength of the light of interest. When the electric vector is waving in a direction parallel to the conductive "wires", the electrons can follow the electric field and the electric field is attenuated and/or reflected. When the electric vector is waving in a direction perpendicular to the "wires", the electrons cannot move out of the "wires" and hence the wave continues through the array. This type of polarizer is used frequently in infrared spectrometers.

The most common method of obtaining polarized light in the visible wavelength range is with birefringence, or double refraction. This method was first observed by Bartholinus, in 1670 and further developed by Huygens about 20 years later.[7] To put this in historical perspective, this happened about 50 years after the Pilgrims landed the Mayflower at Plymouth Rock. Galileo had been dead for about 30 years, and Sir Isaac Newton was in the most prolific time of his writing career.

It was observed that when a pencil of light was made to fall onto a calcite crystal, two pencils of light emerged, one slightly offset from the other. It was discovered that each of these light beams had the special characteristics that was later described as being "polarized", with the polarization directions being mutually perpendicular. Sir Isaac Newton described the light beam as having two "sides". Recall that Newton was not a supporter of the undulatory (wave) theory of light, but supported

the corpuscular theory. By analogy of the two "poles" of a magnet, Malus suggested that the two "sides" of the light beam were "polarized" differently.

Over 100 years after Huygens description, Malus was using a doubly refracting crystal when he discovered that reflected light was polarized. He accidentally observed light reflected from the windows of Luxembourg Palace through a calcite crystal and correctly interpreted his observation.

4.2.2 Double Refraction[8]

In an amorphous material or a perfectly symmetric crystal, light propagates at the same speed in all directions and the material is described as being isotropic. For crystals which are not symmetric in all three directions, the possibility exists that the speed of light is different in different directions. For a biaxial crystal, the speed of light (and hence the index of refraction) is different in three mutually perpendicular directions. For a uniaxial crystal, the speed is the same in two of the three directions (i.e., in the plane which contains these two axes), and different in the third direction. We will focus on uniaxial materials.

If unpolarized light is directed through a uniaxial crystal along the direction of the unique axis (called the optic axis), the light emerges unchanged. If the light is directed through the crystal in any other direction, the light is separated into two beams, the ordinary ray and the extraordinary ray, as shown in Figure 4.2. This phenomenon is described as "double refraction" and occurs in many crystals which are not cubic.

Because of the anisotropic nature of the crystal, the index of refraction is different for light traveling along the unique axis compared to light traveling perpendicular to the unique axis. Calcite is a common material which is used in these applications, and the indices of refraction for the two different situations are shown in Figure 4.3.

For calcite, light traveling along the optic axis has an index of refraction of n_o. Light traveling in the plane perpendicular to the optic axis with one specific polarization will have an index of refraction of n_e. Light traveling in any other direction will be separated into two beams. The ordinary ray will have an index of refraction n_o and the extraordinary ray will have an effective index of refraction which is intermediate between n_e and n_o. Accordingly, the index of refraction for the extraordinary ray is a function of angle or direction. The values shown for n_e in Figure 4.3 are then the extreme values. The ordinary ray behaves according to the common version of Snell's law. The extraordinary ray requires a more general version of Snell's law.

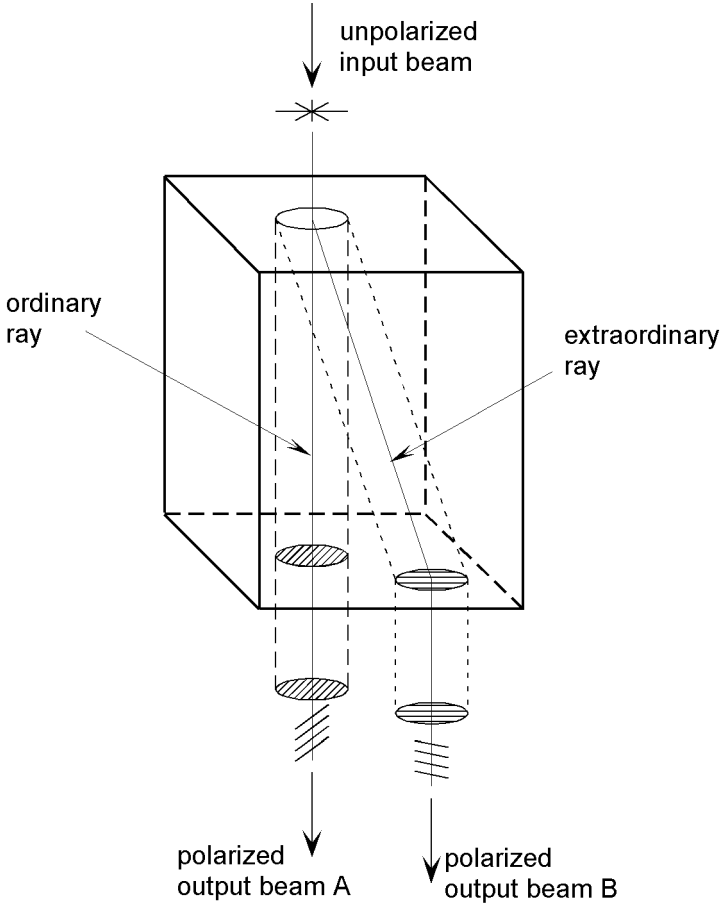


Figure 4.2 Double refraction, illustrated in a generic anisotropic material. Note that the sides of the material do not represent crystal planes.

4.2.3 Calcite Crystals

The primary material used for this purpose is calcite. Historically this was called "Iceland Spar" and was introduced to Europe from Iceland in the 17th century. The large deposits on the east coast of Iceland have long since been exhausted, and most calcite now comes from Mexico, Africa, and Siberia.

In Figure 4.2, there was no reference to crystal planes. In Figure 4.4 we show a representation of the calcite crystal and the optic axis. The crystal is a rhombohedron with faces which are parallelograms with

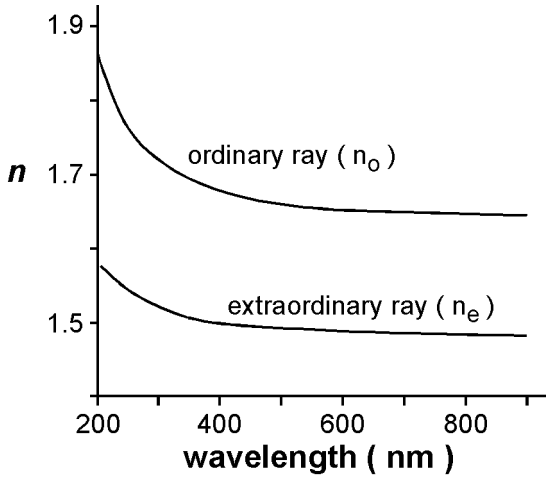


Figure 4.3 The index of refraction *n* for calcite in the direction of the optic axis (shown as the ordinary ray) and in the plane perpendicular to the optic axis (shown as the extraordinary ray).

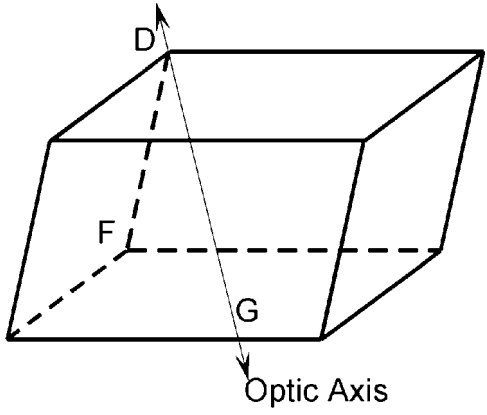


Figure 4.4 Calcite crystal form. The corner labeled D and the opposite corner are the blunt, or obtuse corners. The crystal faces intersect with three angles of ~102° each. At the other six corners the planes intersect at two angles of 78.5° and one angle of ~102°. The optic axis is shown.

angles of 78.5° and 102°. The chemical form of calcite is calcium carbonate ($CaCO_3$).

A line drawn through corner D in such a way that it makes equal angles with the edges intersecting at D defines the principal crystallographic axis, or the "optic axis". We note that the optic axis is a direction

in the crystal rather than a single line. Any line which is parallel to the line bisecting the angles is also an "optic axis". Light which passes through the crystal in the direction of the optic axis is not separated into two beams, but continues as a single beam, undeviated. Unpolarized light which passes through the crystal in the direction perpendicular to the optic axis is separated into two beams. They travel along the same path and hence are undeviated, but they travel at different speeds, however. Light passing through the crystal in any other direction will be separated into two beams which are traveling in different directions (and different speeds), as described above in Figure 4.2.

4.2.4 Polarizers and Analyzers

The terms "polarizer" and "analyzer" have been used in optics for a very long time.[6] The device is often the same kind for both cases, but in this context, the polarizer is the device which sets the state of polarization of a light beam and the analyzer is the device which determines the state of polarization of a beam after it has been altered in some way (e.g., by a reflection).

Although it was understood from the late 1600s that the calcite crystal separated light into two beams which had unique properties, it was not until 1828 that William Nicol[9] devised the "Nicol prism" and was able to remove one of the two beams from the axis of the optical instrument, thus giving a convenient polarizer. It is reported that he himself did not completely understand how it worked.[10]

In making the Nicol prism, the calcite crystal was polished to an appropriate shape. Note that the prism sides did not correspond to the natural crystal planes. The plane which contains the optic axis and the normal to one of the natural crystal planes is called a "principal section" and the light was made to pass through the prism in one of the principal sections.

In Figure 4.5 we show a principal section of a Nicol prism. After polishing to the appropriate shape, the crystal is then cut along the plane which is perpendicular to the principal section and perpendicular to the ends of the crystal (now at 68° from the bottom). The two surfaces are ground and polished optically flat and then the two pieces are cemented together using Canada balsam. This material is a clear transparent substance with an index of refraction which is about midway between that of the *O* and *E* rays.

At the angles shown, the ordinary ray makes a total internal reflection at the calcite/Canada balsam interface. This beam is diverted out of the instrumental axis and is usually absorbed by the surrounding material. On the other hand, the extraordinary ray, with a lower index of

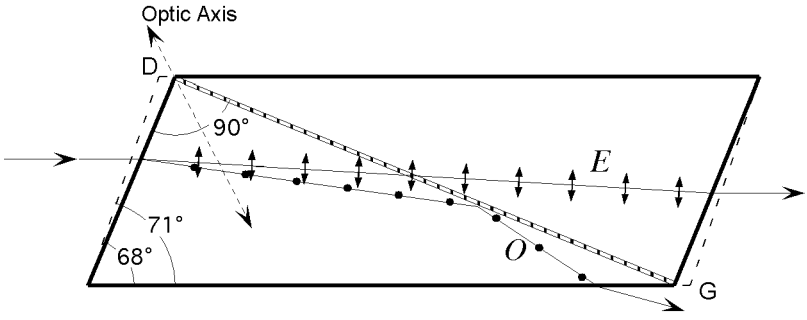


Figure 4.5 Schematic drawing of a Nicol prism, showing the principal section.

refraction than the Canada balsam, does not suffer total internal reflection. It passes through the Canada balsam and into the other calcite part and subsequently out the other end of the Nicol prism. The exiting beam is polarized with the electric vector parallel to the principal section. Note that with a Nicol prism, there is a lateral displacement of the beam. Another disadvantage of the Nicol prism is a rather narrow field angle ($<20°$).

About 50 years after the development of the Nicol prism, another type prism was developed by P. Glan[11] and subsequently improved by S. P. Thompson.[12] One configuration of a Glan-Thompson prism is shown in Figure 4.6. The prism is ground and polished in such a way that the optical axis is in the plane of the entrance face as well as parallel to the diagonal cut. Although this results in much more calcite wastage than does a Nicol prism, there are several advantages. Since the light enters the crystal normal to the surface and to the optic axis, both the ordinary ray and the extraordinary ray continue to move normal to the surface, undeviated. The two rays move with different speeds, however, corresponding to the index of refraction of each respective beam. The angle for the diagonal cut is chosen such that the ordinary ray is totally internally reflected and the extraordinary ray is transmitted, undeviated. For traditional Glan-Thompson prisms, the two halves are cemented together. For some Glan-type prisms (often called Glan-Taylor prisms), the two halves are separated by an air gap rather than being cemented together. Note that the material in the separation between the two parts of the prism does not have to have an index which is intermediate to the ordinary and extraordinary ray. What is required is that the angle is such that one of these two rays makes a total internal reflection and the other does not.

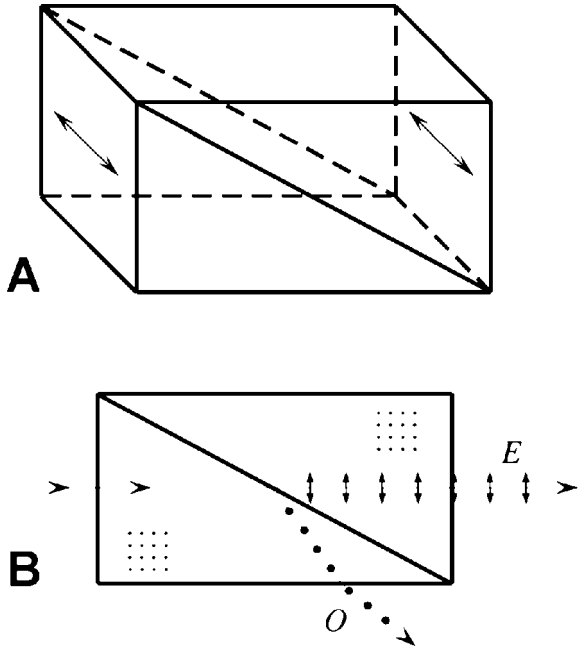


Figure 4.6 A Glan-Thompson prism. Perspective (A) and Cross-sectional (B) drawing. The optic axis is shown by the double headed arrows in (A) and by the matrix of arrow tips in (B).

4.2.5 Wollaston Prisms

In most ellipsometers, we are content to discard half of the light beam and use the light having only one polarization. In some cases, notably the phase modulated ellipsometers, we wish to retain both polarizations, separated into two beams. This is usually included in the polarization state detector part of the ellipsometer. For this we often use a polarizing beam splitting prism, specifically a Wollaston prism, as shown in Figure 4.7

The Wollaston prism consists of two sections which have optic axes as shown. The light ray enters the prism perpendicular to the crystal surface (and perpendicular to the optic axis). Both the ordinary and extraordinary rays continue undeviated, but with different speeds, in much the same way as in the first part of a Glan-type prism. The cleavage angle between the two parts is such that neither the ordinary nor extraordinary ray makes a total internal reflection, but both pass through the interface into the second part of the prism. The optic axis of the second part is

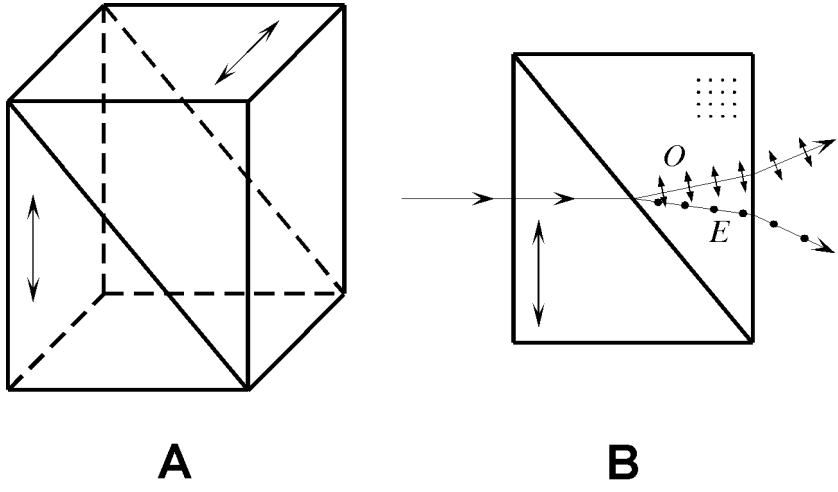


Figure 4.7 The Wollaston prism. (A) shown in perspective. The double arrows show the optical axis of each part. (B) shown from the side. The optical axis of the right part is perpendicular to the page.

perpendicular to the axis of the first part, as shown by the matrix of arrow tips shown in the upper right hand corner (perpendicular to the page). Because the axes of the two parts of the prism are perpendicular, at the interface, the ordinary ray becomes the extraordinary ray and vice versa. Thus the original *O* ray (now the *E* ray) enters a medium of lower refractive index and is refracted away from the normal to the cut. On the other hand, the original *E* ray (now the *O* ray) enters a medium of higher refractive index and is refracted toward the normal to the cut. When constructed properly, the deviation of the two outgoing beams is symmetric about the incoming beam. As they leave the exit face of the prism, the deviations are increased due to Snell's law.

4.2.6 Compensators, Quarter-Wave Plates, and Retarders

4.2.6.1 The Simple Quarter-Wave Plate

Suppose we consider the incoming linearly polarized light beam to consist of two mutually perpendicular components which are in phase. It is sometimes desirable to shift the phase of the beam such that we have elliptically (or circularly) polarized light. We almost always obtain this result

by passing the light through a birefringent crystal (sometimes called a "wave plate") in such a way that the index of refraction for one component is slightly different than the index of refraction for the other component.

Suppose we have a plate where the thickness of the plate is d, and the thickness is such that an integral number of waves fits in the crystal (the wave coming out has the same phase as the wave going in). The wavelength of the light in vacuum is λ and the wavelength in the crystal is λ/n, where n is the appropriate index of refraction. Then the total number of waves in the plate is $\dfrac{d}{\lambda/n}$. If each of the mutually perpendicular components is controlled by a different index (e.g. n_i and n_j), then the phase difference upon emerging would be

$$\delta = 2\pi \frac{d(n_i - n_j)}{\lambda} \tag{4.1}$$

We obtain this separation in several ways, but they all consist of arranging for the two mutually perpendicular components to pass through the wave-plate at different speeds, (i.e., experiencing different values for the index of refraction).

Historically, for single-wavelength ellipsometry (SWE), a quarter-wave plate was used. This optical element would transform linearly polarized light into elliptically or circularly polarized light, depending on the relative orientation of the quarter-wave plate to the plane of polarization of the light. Figure 4.8 shows a simple plate retarder.

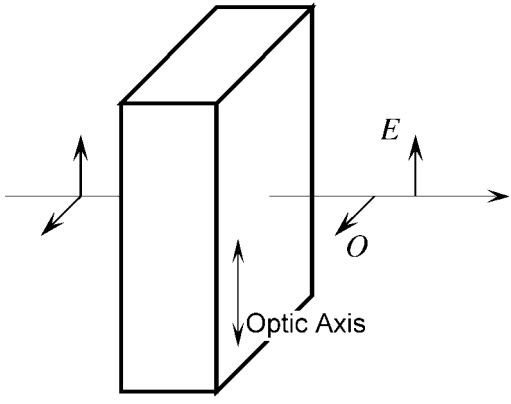


Figure 4.8 Simple retardation plate, made from a negative uniaxial material. The optic axis is shown. The crystal retards the ordinary ray *O* relative to the extraordinary ray *E*.

The crystal is cut such that the optical axis lies in the plane parallel to the face of the retarder plate. The light beam enters normal to the face and optic axis and hence is not deviated. In this example, the light is polarized such that the component parallel to the optic axis has the same amplitude as the component perpendicular to the optic axis. Because the ordinary and extraordinary rays travel at different speeds, the phase fronts become separated. The extraordinary ray travels faster than the ordinary ray (for a negative uniaxial crystal), hence when the light beam emerges from the crystal, the two rays are not in phase. The phase retardation δ is given by

$$\delta = 2\pi N = \pm \frac{2\pi d(n_e - n_o)}{\lambda} \tag{4.2}$$

where d is the thickness of the plate and λ is the wavelength of the light. N is the retardation in fractions of a wavelength. If $N = 1/4$, we have a simple zero-order quarter-wave plate. If $N = 1\ 1/4$, $2\ 1/4$, etc., we have a multi-order quarter-wave plate. If $N = 1/2$, $1\ 1/2$, $2\ 1/2$, etc., we have a half-wave plate.

For wave plates, having multi-orders magnifies the effect of wavelength on the phase retardation, as well as magnifying temperature effects and other effects. Because of this, it is preferable to have a phase retardation of only one quarter-wave. For quartz (a common birefringent material) for red light, $n_e - n_o \approx 0.009$. From the above equation, we calculate that for quartz, a simple zero-order quarter-wave plate would be only 18 micrometers thick. From a structural strength point of view, a device this thin would be difficult to fabricate and to handle. Accordingly, a common practice is to fabricate a "pseudo" zero-order quarter-wave plate which is considerably thicker but can be readily fabricated and handled.

If we attach two thicker crystals together in such a way that the direction of propagation is perpendicular to the fast and slow axes of both (as in the simple quarter-wave plate) but the fast axis of the second crystal is perpendicular to that of the first, a phase shift is induced by the first crystal which is somewhat counteracted by the second crystal. The thicknesses of the crystals can be chosen such that the net effect is to shift the phase by only one quarter-wave.

With a quarter-wave plate, if the two components of the incoming beam are in phase, as suggested in Figure 4.8, the emerging wave is $\pi/2$ or 90° out of phase. If the two components have equal amplitudes, the wave emerging from the wave plate will be circularly polarized. If the amplitudes of the two components are not equal, the resulting wave will be elliptically polarized. Note that if either of the components were zero (i.e., the linearly polarized wave were aligned to either the fast or slow axis) then the emerging wave would remain linearly polarized. With

incident plane-polarized light, one can shift the ellipticity from linear to circularly polarized by simply changing the azimuthal orientation of the wave plate relative to the orientation of the incident polarized light.

We should note that for this type of wave plate, the retardation is a function of not only the thickness, but the wavelength, hence a plate which is a quarter-wave plate for one wavelength will not be a quarter-wave plate for other wavelengths. This implies that this simple retardation plate will work well for SWE, but has some difficulties when used for SE.

4.2.6.2 The Babinet-Soleil Compensator

By varying the thickness of the wave plate, one can select the wavelength for which the retardation is 90°, or conversely, for any given wavelength, one can vary the thickness in such a way that the device is a quarter-wave plate. This can be constructed manually by using wedges of the birefringent material. The Babinet-Soleil compensator is one example of such a device and is shown in Figure 4.9 The wedge is moved with a micrometer screw.

In some of the early manual ellipsometers, used prior to the introduction of laser sources, a variable retarder such as this was used. In principle, one could use this type with spectroscopic ellipsometry, although in practice, it is not commonly used.

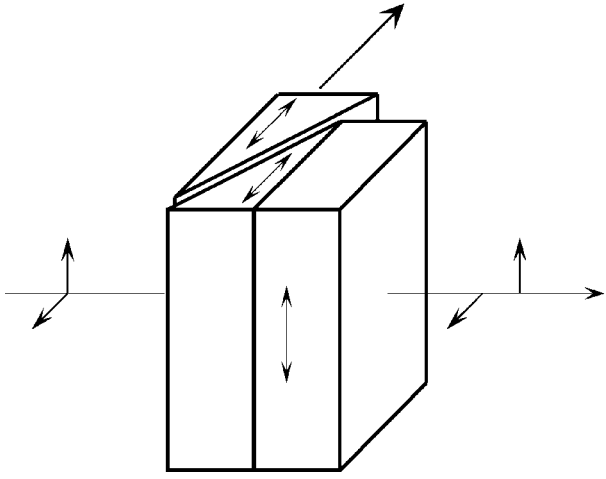


Figure 4.9 The arrangement of a Babinet-Soleil compensator. The wedge is moved with a micrometer.

4.2.6.3 The Multi-Element Retarder

Equation (4.2) gives the phase shift induced by a simple quarter-wave plate. For a given wavelength λ, the thickness d is chosen such that $\delta = \pi/2$. If one uses this "quarter-wave plate" with radiation with a different wavelength, clearly, the phase shift is not a quarter wave. Note from Figure 4.3 that although n_e and n_o vary significantly as a function of wavelength, the difference between them is not a strong function of wavelength. Hence, we would expect that the phase shift for a given retarder would vary roughly as $1/\lambda$. An example of this is shown in Figure 4.10.

It is not essential, for ellipsometry, that the retardance be exactly 90° so long as the calculation software has the correct value for the retardance for each wavelength of interest. On the other hand, no ellipsometric data can be obtained if the retardance is zero or a multiple of 360°, and if the retardance is either 180° or 180° plus a multiple of 360°, the ability to accurately measure the ellipsometric parameter Δ is lost.[13]

From Figure 4.10, we see that if the wavelengths used are simply those in the visible region, this retarder would probably be sufficient. In the UV region, however, the retardance passes through 180°. Accordingly, this simple retarder is not useful when the spectral range extends into the UV region.

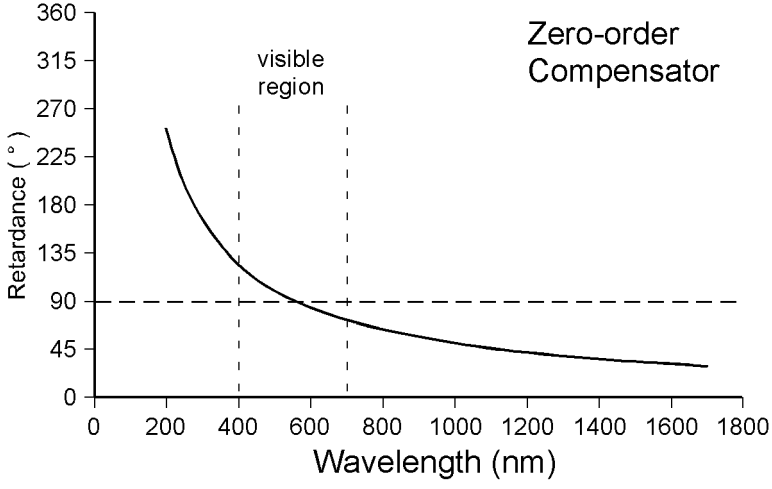


Figure 4.10 The retardance of a simple retarder as a function of wavelength. The thickness was chosen so that the retardance was a quarter-wave in the middle of the visible region.

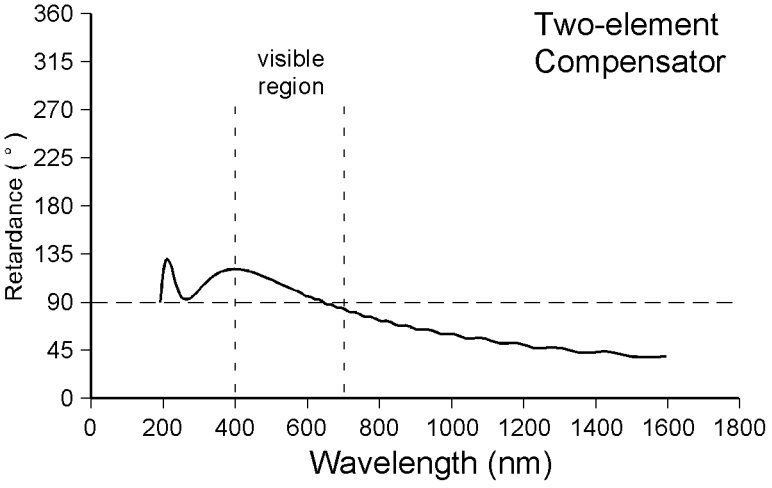


Figure 4.11 Retardance as a function of wavelength for a two-element compensator. Quarter-wave plates for 266 nm and for 633 nm were used. The fast axes were oriented approximately 45° apart. (data taken from Johs, et al.[14]).

A two-element compensator can be constructed using two retarders which are quarter-wave plates for different wavelengths. The fast axes of the two elements are offset from one another.[14] Figure 4.11 shows the retardance curve for one such combination. The retardance of this two-element compensator does not cross 180°, but stays within the 90–135° range in the UV. The trade-off is that calibration is somewhat more complex since the "effective" fast axis varies strongly with wavelength.[14]

4.2.6.4 The Fresnel Rhomb

Using the Fresnel rhomb is a very old method of obtaining circularly polarized light.[15] The basic idea is that when light makes a total internal reflection inside a transparent material, a phase shift occurs, and the amount of phase shift is different for the p-waves and the s-waves. The relative phase shift between the p and s waves is angle dependent. Figure 4.12 shows one configuration.

In this particular example, the parallelepiped (or rhomb) is made of a transparent material which has an index of refraction of 1.51 (e.g., crown glass) in the center of the visible region. The dependence of phase shift is somewhat convoluted,[16] but the net effect is that the dependence of phase shift on wavelength is not nearly as strong as that shown in

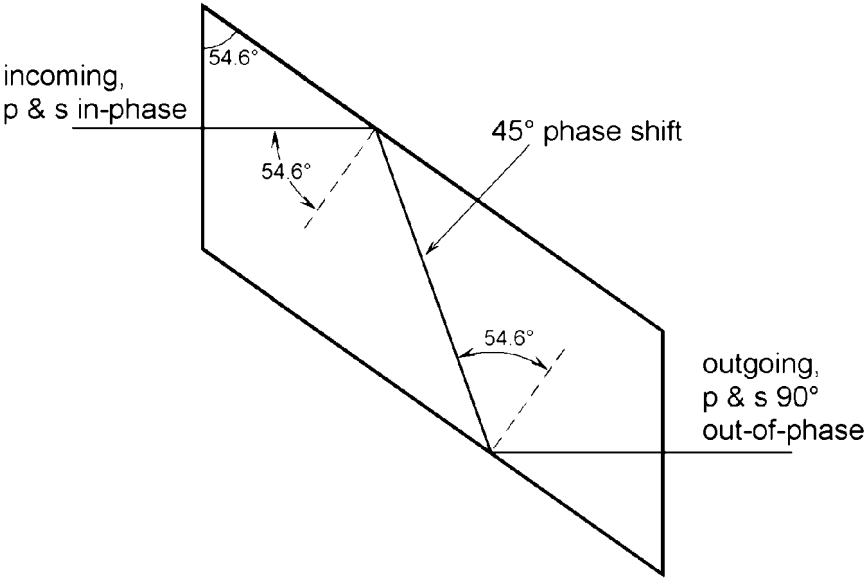


Figure 4.12 The Fresnel rhomb. The angles shown are for a material with index n = 1.51. If the incoming p-waves and s-waves have equal amplitude and are in phase, the light is elliptically polarized ($\lambda/8$ phase shift) after the first reflection and is circularly polarized ($\lambda/4$ phase shift) after the second reflection.

Figure 4.10. For some materials, the phase shift varies less than 2° over the range from 400 nm to 700 nm.[17]

One disadvantage of the Fresnel rhomb shown in Figure 4.12 is that light beam is displaced. If one makes the angle-of-incidence equal to 74.7°, the phase shift is $\lambda/16$ and four reflections are required for a quarter-wave shift. Such an arrangement is shown in Figure 4.13. In this case, the result is no net displacement of the light beam.

4.2.6.5 The Tilting-Plate Compensator

With this type compensator, the retardance is not necessarily 90°. We will find that for rotating analyzer and rotating polarizer ellipsometers, it is often desirable to shift the phase somewhat so that the value of Δ is not particularly close to 180° or 0°. Although it is not necessary to shift the phase 90°, it is necessary to know the value of the shift for subsequent analysis. The Berek compensator is one such tilting-plate compensator,[18] shown in Figure 4.14.

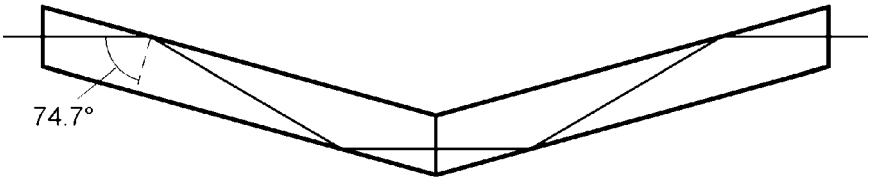


Figure 4.13 A double Fresnel rhomb quarter-wave retarder. The angles shown are for a material with index $n = 1.51$. Each reflection induces a shift of $\lambda/16$ thus giving a phase shift of a quarter-wave after the four reflections. The light exiting the device on the right is on the same axis as the light entering the device on the left.

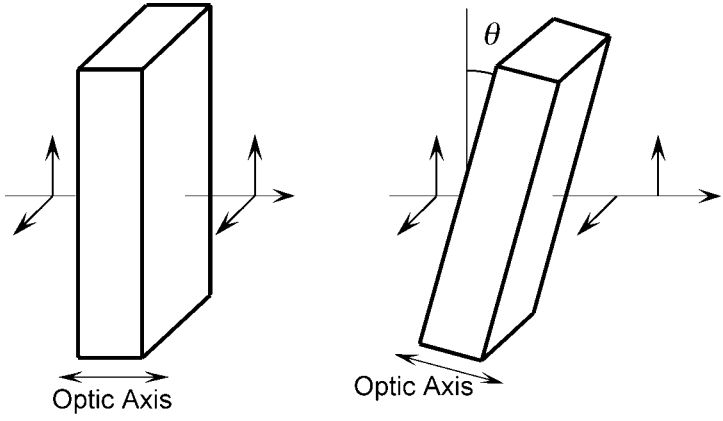


Figure 4.14 The Berek compensator, an example of a tilting-plate compensator.

Whereas in the traditional quarter-wave plate shown in Figure 4.8, the optic axis is parallel to the crystal face, in the case of a Berek compensator, the optic axis is perpendicular to the crystal face. If the crystal is not tilted (as shown in the left side of Figure 4.14), and the light enters at normal incidence, it propagates with a velocity independent of polarization, according the ordinary index, n_o and hence no phase shift occurs. If the crystal is tilted at an angle θ, as suggested in the right side of Figure 4.14, the part of the light beam which is polarized perpendicular to the tilt axis (in this case, in the vertical plane) will become slightly extraordinary, with an effective index of refraction n'_e which is given by

$$\frac{1}{n'_e} = \sqrt{\frac{\cos^2\theta}{n_o^2} + \frac{\sin^2\theta}{n_e^2}}. \tag{4.3}$$

Since the tilt is always a small angle (typically less than 10°), the effective index of refraction n'_e is very close to n_o. The primary effect of the tilt is to distinguish the ordinary ray from the slightly extraordinary ray. The displacement of either ray due to the tilt for a 1 mm thick plate is less than 100 micrometers and the difference in the displacement of the two rays (the separation) is less than 100 Å. However, due to the slight difference in velocity, the two beams come out of the crystal out of phase. The phase difference δ (in radians) is given by

$$\delta = 2\pi(n'_e - n_o)\frac{d}{\lambda} \tag{4.4}$$

where d is the thickness of the plate. For a 2 mm plate of material such as MgF_2, at a wavelength of 500 nm, a tilt of 3° will give a retardance of slightly less than 50°.

It should be noted that if one tilts the crystal the same angular amount about the same axis, but in the opposite direction, the phase shift will be the same (as opposed to being equal to $-\delta$). In order to obtain a phase shift in the opposite direction, one has to tilt the crystal about an axis perpendicular to the one shown in Figure 4.14.

One of the primary advantages of the tilting-plate compensator is that it is very easy to obtain the same effect as having no compensator by simply aligning the beam and the crystal optical axis.

4.2.7 Photoelastic Modulators

In the above described devices, once the device is configured, the phase shift is constant in time. For the photoelastic modulator, the phase shift changes sinusoidally as a function of time. Like the compensators described in the previous section, the phase separation is obtained by having the two perpendicular components travel through a "wave plate" at different speeds. Whereas the different speeds in the previous examples were obtained by using a birefringent crystal at rest, it is obtained in a photoelastic modulator with a time-varying birefringence induced by a time-varying stress in a normally isotropic material. An isotropic material such as fused quartz will become anisotropic when stressed and hence will induce the same kind of birefringence as an anisotropic crystal such as calcite.

The construction of a typical photoelastic modulator[19] is shown in Figure 4.15. The piezoelectric transducer, shown on the left, is a block of crystalline quartz which is cut in a specific orientation (−18°, X cut). A metal electrode is deposited on each of two sides and the transducer is cut in such a way that it resonates when a 50 kHz electric field is applied. The resonance is uniaxial and is directed along the long axis of the crystal. A

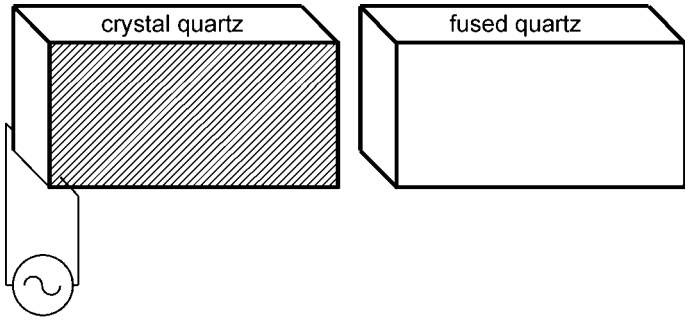


Figure 4.15 The components of a photoelastic modulator.

block of fused quartz is cemented to the end of the transducer. The end dimensions of the two pieces are the same and the length of the fused quartz is such that it too has a 50 kHz fundamental longitudinal resonance. When the two are cemented together, resonance of the transducer causes a periodic strain in the fused quartz.

When the fused quartz is unstrained, it is isotropic. The phase of two perpendicular components of light will remain unchanged. As illustrated in Figure 4.16, when the optical element (the fused quartz) is strained in the longitudinal direction, the index of refraction in that direction changes. The block is not strained in the vertical direction and hence the index of refraction in the vertical direction remains unchanged. When the optical element is in tension (A), the longitudinal component of light travels slightly faster than the vertical component in such a way that the two components are no longer in phase. Similarly, when the optical element is compressed (B), the longitudinal component travels slighter slower than the vertical component.

The light is therefore periodically phase shifted, first in the positive direction and then in the negative direction, at the resonant frequency. The frequency of oscillation is determined by the size and shape of the crystals. The amplitude of the oscillations (and hence the maximum ellipticity) depends on how much voltage is applied at the resonant frequency. The oscillating birefringence effect is at its maximum at the center of the fused quartz element.

4.2.8 Monochromators

For SWE, a monochromatic light source such as a laser is typically used and hence the wavelength is fixed. For SE, we need to obtain the Δ/Ψ information as a function of wavelength (or energy), hence some manner is needed either to measure the light one wavelength at a time

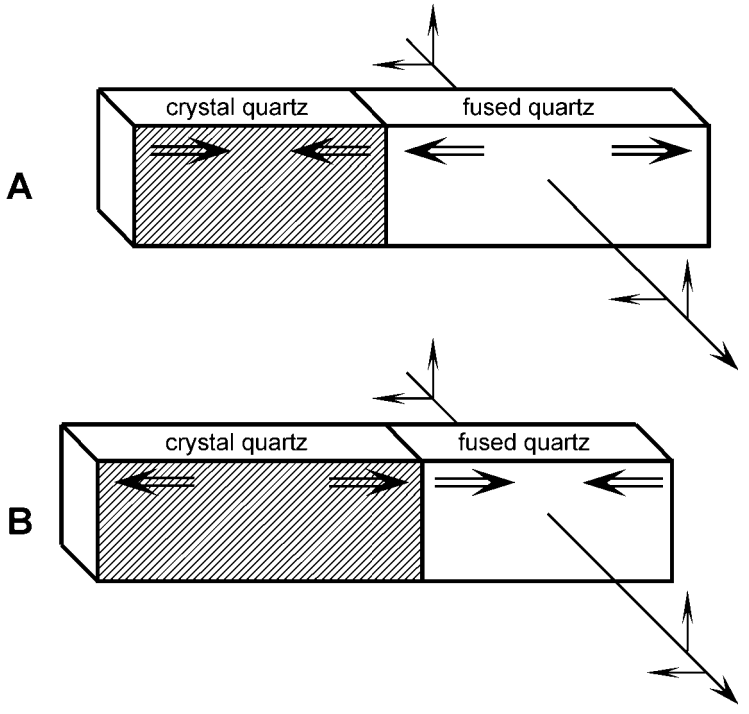


Figure 4.16 As the transducer oscillates, the fused quartz is strained in the longitudinal direction. The induced change in the index of refraction causes the component of light in the longitudinal direction to move either slower (A) or faster (B) than the component in the vertical direction.

or to simultaneously measure the contributions of the individual spectral components.

In some cases, the monochromator is made part of the light source, hence only the light with a single wavelength strikes the sample. In other cases, white light strikes the sample and the light is separated into its spectral components in the detector side of the instrument.

Historically (Newton, circa 1672), light was divided into its spectral components with a prism. Another method of separating the light into its spectral components has been available since roughly the beginning of the 20th century, however, and that is the diffraction grating. H. A. Rowland, in 1882, was able to produce a simple diffraction grating. Lord Rayleigh, in 1888, developed the concept theoretically, suggesting that it should be possible to put a large fraction of the light into a single order (this would later be called a "blazed" grating). By 1910, R. W. Wood had advanced the art of making them.[20]

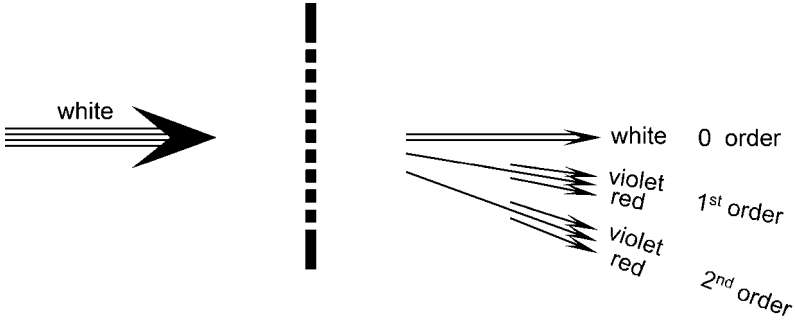


Figure 4.17 Transmission diffraction grating separating white light into its components.

To understand how a diffraction grating separates the light into its spectral components, let us start with a transmission grating, shown in Figure 4.17. A transmission grating consists of a series of parallel slits which are spaced at regular intervals, with the spacing d being comparable to the wavelength of light, λ. Light with a given wavelength will have constructive interference at several selected angles of deviation from the incident direction. The equation for constructive interference for a given wavelength is

$$\sin\theta = m\frac{\lambda}{d} \tag{4.5}$$

where m is an integer with values 0, 1, 2, 3, etc. For m = 0, i.e., the zero order, the light emerges with no deviation. If various wavelengths are present, they all come out together, hence with white incident light, the zero order beam is white. The various other beams are referred to as the 1st order beam, the 2nd order beam, etc., and the light is separated into its spectral components according to the equation above. If the light is not incident in a normal direction, but at an angle ϕ from the normal, then the equation becomes

$$\sin\phi + \sin\theta = m\frac{\lambda}{d} \tag{4.6}$$

One difficulty arises from the multiple order beams. From the equations above, we see that light for the first-order beam with wavelength λ will occur at the same angle as light for the second-order beam for wavelength of $\lambda/2$. Light at 400 nm (violet) from the first order will coincide with light at 800 nm (near IR) from the second order. Filters are sometimes used to separate these two components.

The effect shown in Figure 4.17 is that of a transmission grating. Diffraction can also occur with reflection in much the same manner. If the surface of the grating shown in this figure were made of a reflective material, one would also observe various order diffraction from reflection since the part of the grating between the slits has the same spacing and periodicity as the slits themselves. Reflection gratings are somewhat easier to fabricate from a structural strength point-of-view.

It is the task of any dispersion device to exclude all wavelengths except the desired one. For a grating, one must also exclude all except one order. Consequently, it is often desirable to enhance one order beam compared to the other order beams (for example, we might want to make the 1st order beam very intense compared to the 2nd, 3rd, or greater orders.) This can be done by tailoring the shape of the reflecting surface. Figure 4.18 shows a cross section of a reflection grating with controlled groove form. This type grating concentrates most of the intensity into one order (usually the 1st or 2nd) on one side of the incident beam. With proper choice of shape, gratings can be produced which will show a *blaze* of color at any desired angle, hence they are described as "blazed gratings".

For ellipsometry, the grating can be used with slits so that only the desired wavelength passes through to the detector. This is often done with a double-monochromator with filters, choppers, etc. They also can be used with a series of detectors (i.e., a charge coupled detector array) so that a set of chosen wavelengths can be read simultaneously.

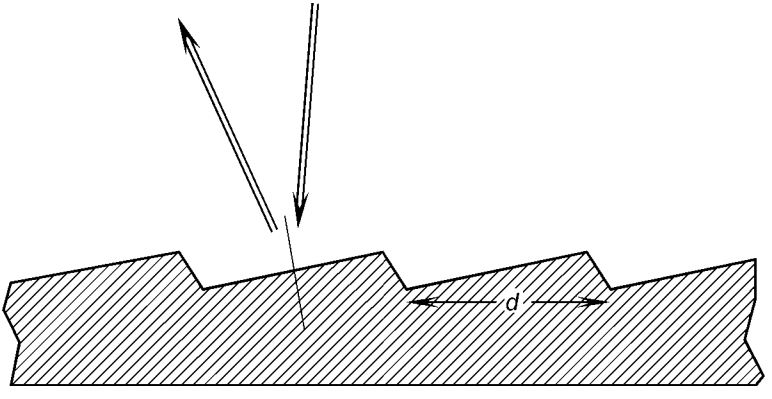


Figure 4.18 Cross section of a blaze reflection grating.

4.2.9 Goniometers

The measurement of angles is one of the key issues of ellipsometry. This involves the angle-of-incidence as well as the azimuthal angle position of the polarizers, analyzers, and retardation plates. Measurement of angle is a very old technology. Historically, the measurement of angle was used in surveying, navigation, and astronomy. The 360-degree division of a circle originated in the astronomy of the Babylonians. The further subdivision into minutes and seconds based on a sexagesimal system (e.g., 15° 7′ 30″) also dates to that time.[21] The decimal notation (e.g., 15.125°) could not have been used prior to the early 1600's when John Napier developed the concept of the decimal notation as part of his development of logarithms.[22]

It is interesting to note that there exists another division of the circle, sometimes used by mariners, which divides the compass into 32 points. This results from bisecting the four cardinal directions, then repeating the bisecting process two additional times. In international waters, when one vessel approaches another more than "two points abaft the beam", it is considered a "passing" situation rather than a "meeting" situation. The 360-degree system has persisted for all other angular measurements, however, probably due to the similarity to the number of days in the year.

It was the work of the Hellenistic astronomer Ptolemy (circa 100 − 170 AD) that established the methods of angular measurement (for astronomy) for many centuries to come.[23] There is very little written historically about the manufacture of the instruments for angular measurement prior to the late 1700s. The art of graduating the circle was carried out in secrecy, with each artisan having a method of his own until in 1767, the British Board of Longitude persuaded John Bird to set down his methods in print for the benefit of the nation.

Many of the early instruments were segments of a circle and the possibility of systematic error was great. The graduation of an entire circle must close on itself, and hence the possibility of systematic error was reduced. Kepler notes that Ptolemy could measure only to an accuracy of about one-sixth of a degree (or to 10 minutes of arc).[24] If one considers the entire range to be 180°, this represents an uncertainty of about one in 1000, which doesn't seem too bad for instruments used almost 2000 years ago.

Tycho Brahe had established an observatory on the island of Hven in the Danish sound in 1575. His innovations rendered a single minute of arc a sensible quantity. The best astronomical quadrants were made to have tolerances down to about 2 seconds of arc by 1775, and by 1820, the circles were measuring down to fractions of a single second.[23]

Technological improvements included the vernier scale, developed by Pierre Vernier[26] in 1631, and the micrometer screw and worm gear,

developed in late 1700s.[27] Thus, by the time of Drude, in the late 1800s, the technological abilities for angular measurements exceeded the needs for the ellipsometric technique. It might be noted in passing that an inexpensive hand-held global positional system receiver, along with the necessary artificial satellite system, gives a longitudinal angular position on the earth's surface to within few meters, which represents an angular measurement uncertainty of about one part in ten million.

The remaining improvements in the technologies needed for ellipsometry then became speed and automation. The development of the stepper motor in the late 20th century allows current instruments to position themselves quickly and reproducibly with angular accuracy in excess of 0.01°. The limiting step in most ellipsometry is not angular measurement, but component non-ideality and readout non-ideality. The other primary limitation is how well the sample conforms to the model of a uniform material with plane parallel interfaces.

4.3 Ellipsometer Component Configurations

4.3.1 Early Null Ellipsometer Configurations

For the first three quarters of the 20th century, ellipsometry was primarily done with a single wavelength, and with a nulling technique.[28] In this configuration, the primary measurements which were made were the azimuthal angles of the polarizer, analyzer, quarter-wave plate, and the angle-of-incidence. In this configuration, quantitative consideration of the intensity of the light was minimal; the only concern was to determine when the null (the intensity minimum, hopefully zero) had been reached.

The rudimentary concept is simple. Polarized light is incident on a sample. The reflection induces an ellipticity. The quarter-wave plate can be rotated until the ellipticity is canceled. The final polarizer determines when this is the case, and it also gives the value of the azimuthal angle of the linear polarized light after emerging from the quarter-wave plate. The azimuthal angles of the polarizer, quarter-wave plate, and analyzer are then used to calculate what the sample did to the probing light beam. Unfortunately, the equations are somewhat cumbersome for this configuration.

In actual practice, a slightly different configuration is used, as shown in Figure 4.19. This simplified the calculation significantly. Rather than have the polarizer fixed at a chosen angle, the quarter-wave plate Q was placed before the reflection and it was held fixed at 45°. The polarizer P and analyzer A were then rotated interactively until null was reached.

With the development of microprocessors, this same configuration was used, with the microprocessor driving the optical instruments and deciding when null had been reached. The microprocessor could also then do some regression analysis to determine sample parameters such as thickness and index for simple sample geometries.

4.3.2 Photometric Ellipsometer Configurations

Because the nulling process was somewhat slow, a different configuration was developed.[28] This is shown in Figure 4.20.

In this case, the polarizer is fixed at an azimuth of 45°. The quarter-wave plate (Q) is held at a fixed angle such that when it is in place, the emerging light is circularly polarized. The quarter-wave plate can be either in the beam, or removed from the beam. After the reflection, the light is, in general, elliptically polarized. The analyzer rotates continuously, hence this type ellipsometer is called the rotating analyzer ellipsometer (RAE).

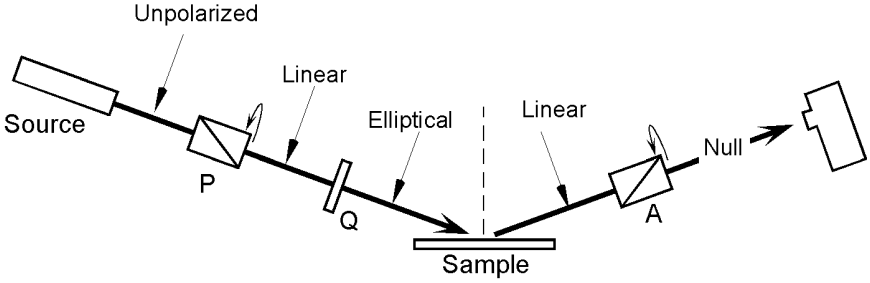


Figure 4.19 The configuration of a nulling single-wavelength ellipsometer.

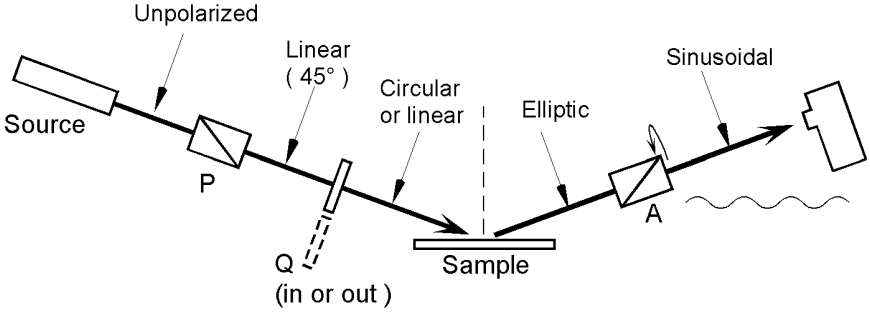


Figure 4.20 The configuration of a photometric single-wavelength ellipsometer.

RAEs are discussed in detail in a subsequent chapter. We simply introduce them at this point. The intensity of the light emerging from the analyzer depends on how the polarizer is oriented with respect to the ellipticity of the light beam. Continuously rotating the analyzer will result in a sinusoidally varying intensity which can be Fourier analyzed to determine the ellipsometric parameters Δ and Ψ. Note that in this case, the magnitude of the light beam is significant, thus this is referred to as a "photometric" instrument. It has been shown by Aspnes[29] that, with this method, the uncertainty in the value of Δ is the greatest when the emerging light is linearly polarized (the measured value of Δ is near zero or 180°). Inserting or removing the quarter-wave plate will shift the value of Δ by 90°, thus assuring that for one of the two situations (in or out) a reliable value of Δ can be obtained.

4.3.3 Spectroscopic Ellipsometers

For spectroscopic ellipsometry, a white light source along with a monochromator of some sort is used to obtain ellipsometric measurements at many different wavelengths. In some cases, the monochromator is positioned prior to the polarizer and in some cases, after the analyzer. For SWE, the quarter-wave plate is usually a simple crystal which retards the beam 90° for that particular wavelength. Clearly, this configuration will not work for SE, which uses many wavelengths.

In the configuration shown above for SWE, half of the measurements are made with the quarter-wave plate in the beam and half are made with it out of the beam. Psi is determined without the quarter-wave plate. The quarter-wave plate is required for single-wavelength ellipsometry for those instances where Δ is near zero of 180°[29]. For many spectroscopic ellipsometers, no compensator (or quarter-wave plate) is used. We simply toss out the data for those wavelengths where Δ is near zero or 180° and do our regression analysis with the remainder of the data. Two such configurations are shown in Figure 4.21.

For any given wavelength, the white light source is not nearly as intense as a laser source. Some manufacturers use chopping in the white light part of the beam to reject room light. In this case, the intensity of the light beam after the analyzer will have two sinusoidal waves superimposed, one from the rotating polarizer or analyzer and the other from the chopper. Fiber optics allows us to position the monochromator at a convenient location, rather than mounting it on the goniometer arm. The fiber maintains intensity information, but does not maintain phase information.

For these configurations, the position of the fixed element, (the polarizer in Fig. 4.21A and the analyzer in Fig. 4.21B) is not held fixed at 45°,

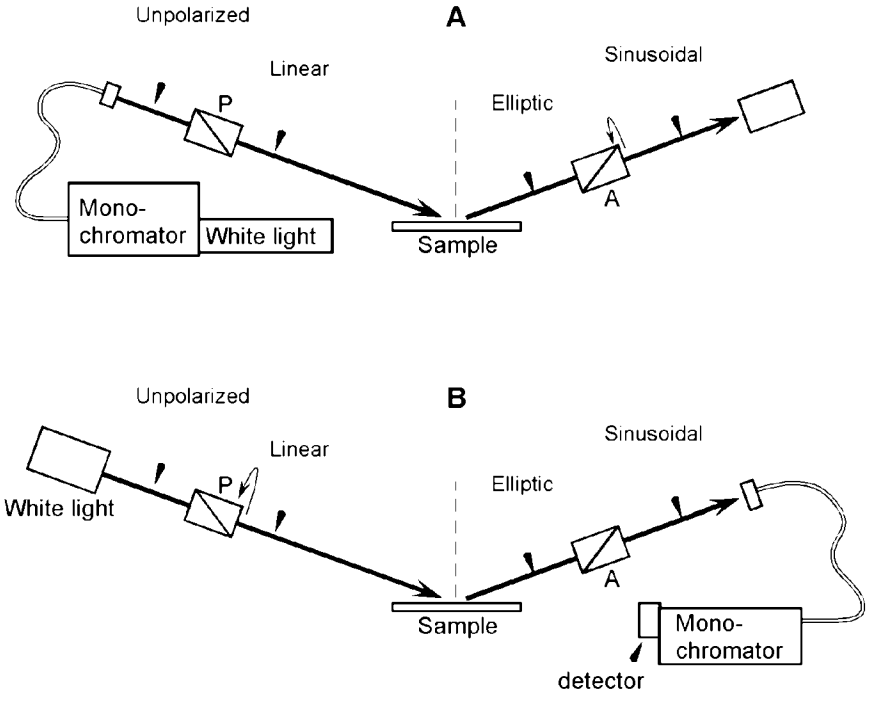


Figure 4.21 (A) Configuration of a rotating analyzer spectroscopic ellipsometer, and (B) configuration of a rotating polarizer spectroscopic ellipsometer. In both cases, the light entering the detector has a sinusoidal component due to the rotation of the polarizer or analyzer.

as was the case for SWE. The uncertainty in the data is reduced significantly if the azimuthal position of the fixed element has a value near that of Ψ. Accordingly, based on the previous measurement, the fixed element is positioned at a value near Ψ prior to making the measurement. This is referred to as "polarizer (or analyzer) tracking."

There are some applications where most of the Δ values will be near zero or 180°. Examples are the analysis of a dielectric substrate, or a semiconductor substrate in the transparent region of the spectrum. A classical example is the analysis of a polymer film on a microscope slide.[30] In this case, we cannot simply throw out the data where Δ is near zero or 180° since this represents almost all of the data. In this case, some sort of compensator is required to move the measured values of Δ to a more favorable part of the Δ/Ψ domain. As shown in Figure 4.22, the compensator is positioned in the incident beam just after the polarizer, in a manner similar to the SWE instrument shown

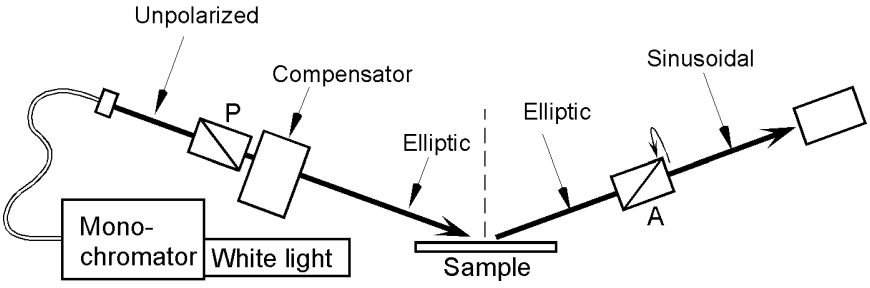


Figure 4.22 Configuration of a rotating analyzer spectroscopic ellipsometer which employs a compensator.

previously. This is not a simple crystal, however, since many different wavelengths are involved. Also, it is not particularly important that the phase difference induced be 90°. It is simply necessary for the compensator to move the measured Δ value away from zero or 180° sufficiently far to reduce the uncertainty. It is also necessary that the controlling computer and software know how much phase shift was induced. When the value of Δ is in a favorable region, it is important not to shift it to a less favorable region. Accordingly, the retarder must be under computer control and be able to induce the proper phase shift to improve the quality of the data.

Various different kinds of compensators are used for this function in SE. These include the tilting-compensator, a compound quarter-wave plate, or a Fresnel rhomb.

4.3.4 Other Configurations

There are a variety of other configurations which can be used in spectroscopy ellipsometry. These include the rotating compensator ellipsometer, multi-wavelength ellipsometer, and the phase-modulated ellipsometer. The components of a rotating compensator ellipsometer are configured analogously to the rotating analyzer except that the rotating element is the compensator. The equations for analyzing the light intensity are also different. The multi-wavelength ellipsometer uses an array of detectors and captures the data for each wavelength simultaneously. This type of ellipsometer is described in detail in a subsequent chapter. The phase-modulated ellipsometer is also described in a chapter later on in the book.

4.4 References

1. P. Drude, *Annalen der Physik und Chemie*, **36**, 532 (1889); P. Drude, *Annalen der Physik und Chemie*, **36**, 865 (1889).
2. *Theory of Optics*, P. Drude, Longmans, Green & Co. New York (1901), p 258.
3. A. Rothen, *Rev. Sci. Instruments*, **16**, 26 (1945).
4. E. Malus, *Memoires de Physique et de chemie de la societe d'Arcueil*, Vol. 2, pp 143, etc., 254, etc., Paris (1809). This is reproduced in *The Principles of Physical Optics*, E. Mach, Dover Publications, New York (1926).
5. *Treatise on Optics*, D. Brewster, Longman, Rees, Orme, Brown, and Green, London (1831).
6. *Polarisation of Light*, W. Spottiswoode, Macmillan & Co., London (1874), p 2.
7. *The Principles of Physical Optics*, E. Mach, Dover Publications, New York (1926), p 186.
8. Double refraction, calcite crystals, and calcite polarizers are described in a variety of sources. These include *Fundamentals of Optics*, 3rd Edition, F. A. Jenkins and H. E. White, McGraw-Hill, New York (1957), pg 496; *Handbook of Optics* Volume II, 2nd Edition, Edited by Michael Bass, McGraw-Hill, New York (1995), Chapter 3. An additional source is *Optics Guide 5* published by Melles Griot, 1990, Chapter 14. Historical references include *Treatise on Optics*, D. Brewster, Longman, Rees, Orme, Brown, and Green, London (1831); *Polarisation of Light*, W. Spottiswoode, Macmillan & Co., London (1874); *Theory of Optics*, P. Drude, Longmans, Green & Co. New York (1901).
9. W. Nicol, *Edinb. New Phil. J.*, **6**, 83 (1828) as quoted in A. Johannsen, *Manual of Petrographic Methods*, 2nd ed. Hafner, New York, 1968, p. 158 (originally published in 1918).
10. *Fundamentals of Optics*, 3rd edition F. A. Jenkins and H. E. White, McGraw-Hill, New York (1957) pg 500 footnote.
11. P. Glan, *Carl's Report*, **16**, 570 (1880).
12. S. P. Thompson, *Phil Mag.*, ser. 5, **12**, 349 (1881); S. P. Thompson, *Proc. Opt. Conv.* (1905) p 216.
13. J. Opsal, J. Fanton, J. Chen, J. Leng, L. Wei, C. Uhrich, M. Senko, C. Zaiser, D. E. Aspnes, *Thin Solid Films*, **313–314**, 58 (1998).
14. B. Johs, J. Hale, N. J. Ianno, C. M. Herzinger, T. Tiwald, and J. A. Woollam, *SPIE Proceedings*, Vol 4449 (2001).
15. The Fresnel rhomb was first fabricated by Fresnel and is described in *Treatise on Optics*, D. Brewster, Longman, Rees, Orme, Brown, and Green, London (1831). Modern treatments are given in most optics textbooks or handbooks. These include *Fundamentals of Optics*, 3rd Edition, F. A. Jenkins and H. E. White, McGraw-Hill, New York (1957), p 517 and *Handbook of Optics*, Volume II, 2nd Edition, Edited by Michael Bass, McGraw-Hill, New York (1995), p 3.54.
16. The equations are given in *Fundamentals of Optics*, 3rd Edition, F. A. Jenkins and H. E. White, McGraw-Hill, New York (1957), p 517.

17. Product literature from Halbo Optics, 83 Haltwhistle Road, South Woodham Ferrers, Chelmsford CM3 5ZA, UK.
18. Most of the information on tilting plate compensators was taken from *Handbook of Optics*, 1st Edition, edited by W. G. Driscoll and W. Vaughan, McGraw-Hill, New York (1978), p 10–140, from U. S. Patent # 5,757,494, by S. E. Green, G. M. Herzinger, B. D. Johs, and J. A. Woollam, and from product information from New Focus, Inc. 2630 Walsh Ave., Santa Clara, CA 95051.
19. Most of the information on phase modulators was taken from S. N. Jasperson and S. E. Schnatterly, *Rev. Sci. Inst.*, **40**, 761 (1969), from G. E. Jellison, Jr. and F. A. Modine, *SPIE* **Vol 1166** "Polarization Considerations for Optical Systems II" (1989) pg 231, and from product literature from Hinds Instruments, Inc, Hillsboro OR 97124.
20. This historical information was taken from *Concepts of Classical Optics*, J. Strong, W. H. Freeman & Co., San Francisco (1958) pg 218.
21. *A History of Mathematical Notations*, Florian Cajori, Dover Publications, New York, reproduction of book published in 1929. paragraph *511*.
22. *Napier Tercentenary Memorial Volume*, edited by C. G. Knott, Longmans, Green and Company, London (1915) p 77.
23. *The Divided Circle*, J. A. Bennett, Phaidon-Christie's, Oxford (1987), p 8.
24. As described in *Dividing the Circle*, A. Chapman, John Wiley & Sons, New York (1995), p 20.
25. Ibid, p 146 and p 150.
26. *La Construction, l'Usage et les Propriétéz de Quadrant Nouveau de Mathématique*, P. Vernier, Brussels (1631).
27. *Dividing the Circle*, A. Chapman, John Wiley & Sons, New York (1995), p 20.
28. Configurations for single-wavelength ellipsometers are described in depth in *A User's Guide to Ellipsometry*, H. G. Tompkins, Academic Press, New York (1993), Chapter 2.
29. D. E. Aspnes, *Journal of the Optical Society of America*, **64**, 639 (1974).
30. "Optimizing the Ellipsometric Analysis of a Transparent Layer on Glass". H. G. Tompkins, S. Smith, and D. Convey, *Surface and Interface Analysis*, **29**, 845 (2000).

5 Rotating Polarizer and Analyzer Ellipsometry

Robert W. Collins

Department of Physics and Astronomy
The University of Toledo

Ilsin An and Chi Chen

Department of Physics and Materials Research Institute
The Pennsylvania State University

5.1 Introduction

The complete reflection ellipsometry measurement entails five basic steps: (1) generation of an incident light beam in a polarization state known accurately in advance through instrument calibration, (2) specular reflection of the beam from a sample surface leading to an emergent beam in a new polarization state, (3) accurate analysis of the new polarization state again relying on instrument calibration information, (4) calculation of beam/sample interaction parameters such as the ellipsometric angles (ψ, Δ) from the information on the incident and reflected beam polarization states, and (5) computation of optical and structural information on the sample such as complex dielectric functions and/or thicknesses from the interaction parameters.[1,2] Steps (1) through (4) depend sensitively on the details of the instrument design, whereas step (5) depends only on the mode of data collection and the parameter representation chosen in the experiment.

In steps (1) and (3), the four components of the Stokes vector describe the most general polarization states for the incident and reflected beams.[1] However, if the incident and reflected beams are completely polarized (so that no depolarization by the sample occurs), and furthermore if one is not interested in the irradiance in the beams (which determine the reflectance of the sample), then two quantities are sufficient to describe each of these states. The two quantities are the azimuth angle Q ($-90° < Q \leq 90°$) of the semimajor axis of the polarization ellipse and the ellipticity angle of the ellipse $\chi = \tan^{-1} e$ ($-45° \leq \chi \leq 45°$; $-1 \leq e \leq 1$), where e is the ratio of the semiminor axis to the semimajor axis of the ellipse. The angle Q is measured from the plane of incidence in a counterclockwise-positive sense looking

opposite to the beam direction. Thus, if the incident beam polarization state is known, then the determination of (Q, χ) for the reflected beam provides two beam/sample interaction parameters, typically the ellipsometric angles (ψ, Δ). In fact, (ψ, Δ) are defined by $\tan\psi \exp(i\Delta) = r_p/r_s$, where r_p and r_s are the complex amplitude reflection coefficients (the reflected to incident electric field ratios) for the p and s linear polarization directions (parallel and perpendicular to the plane of incidence, respectively).

The key step in the development of ellipsometry as an important optical tool for the characterization of specularly reflecting surfaces and thin films was the automation of measurement methodologies that had previously required numerous manual steps and several minutes in order to obtain a single (ψ, Δ) pair. Progress in the automation of ellipsometric instrumentation began in the early 1960s and continued through the 1970s, with powerful automatic instruments and software becoming available commercially soon thereafter. The development of ellipsometric instrumentation over this period can be appreciated by consulting the four conference proceedings on this subject.[3–6] The introduction of inexpensive laboratory computers, assuming the tedious tasks of data collection and analysis, stimulated the rapid development of ellipsometry over this period. This progress was significant in two respects. First, computer control of ellipsometers allowed extension of ellipsometry into the spectroscopic domain using a broad-band source and a scanning monochromator.[7] As a result, one could measure optical functions of materials continuously with high spectral resolution from the near-infrared to the ultraviolet. Without automation, manual instruments could be used to collect only a relatively small number of (ψ, Δ) pairs versus wavelength (not the >100 pairs possible with automatic instruments). Second, computer control also allowed extension of ellipsometry into the time domain using high speed data acquisition for in situ studies of thin film growth kinetics and surface modification.[8,9] Without automation in this case, manual instruments could be used to probe only static or very slowly changing surfaces (>10 min time variations).

Figure 5.1 shows a generic ellipsometer design that includes spectroscopic measurement capabilities.[1,7] Ellipsometers can be classified as either compensating (also called null) or photometric. In compensating ellipsometry the generic design of Fig. 5.1 is applicable, and the ellipsometric angles (ψ, Δ) are determined directly from the adjustable angular settings of the optical components required to achieve a null signal at the detector. Such settings include the azimuth angles of the transmission axes of the polarizer P' and analyzer A', and that of the fast axis of the compensator C', along with the retardance of the compensator δ_C. (The azimuth angles are measured as shown in Fig. 5.1 in a counterclockwise positive sense looking opposite to the beam direction, in the same way as the tilt angle Q of the polarization ellipse.) In photometric ellipsometry,

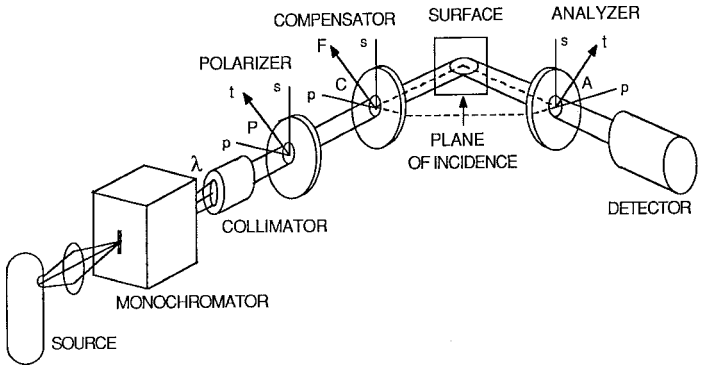


Figure 5.1 A generic ellipsometer design that includes a spectroscopic measurement capability.

(ψ, Δ) are determined from measurements of the time-dependence of the irradiance at the detector also using the known values of the optical element settings or modulation phases. Compensating and photometric ellipsometers can be designed to be either manual or automatic, and either single wavelength or spectroscopic. In Section 5.2 where different instruments designs will be compared in greater detail, the exclusive focus will be on automatic instruments with spectroscopic measurement capabilities.

The two most common ellipsometers as judged by commercial availability are based on rotating polarizer/analyzer and phase modulation principles. The sequences of these two ellipsometers are very similar. As in the generic configuration of Fig. 5.1, the rotating polarizer or analyzer ellipsometer includes a light source, collimating optics, a polarizer, and an optional compensator all on the polarization generation arm of the instrument, and a single polarizer, called an analyzer, optional focusing optics, and irradiance detection system all on the polarization detection arm of the instrument.[10–12] For the rotating polarizer configuration, the polarization state of light on the polarization generation arm of the instrument is modulated by rotating the polarizer continuously during measurement at a constant angular frequency ω, typically ranging between 10 and 100 Hz. The compensator angle and retardance setting (if available) as well as the analyzer angle are fixed during measurement. For the rotating analyzer configuration, the analyzer on the polarization detection arm of the instrument is rotated continuously, while the polarizer angle and compensator settings (if available) are fixed. Finally, for the phase modulation ellipsometer, the compensator in the generic design of Fig. 5.1, optional for the rotating polarizer and analyzer instruments, is replaced by the key element, a piezo-birefringent phase modulator.[13,14] Usually operating at a frequency of 50 kHz, this element modulates the incident beam polarization state while the polarizer and analyzer angles are fixed during the measurement.

In this article, automatic rotating polarizer and analyzer ellipsometers will be discussed in detail. As noted above, the discussion will focus primarily on automatic instruments with spectroscopic capability, based on the recognition that wavelength is the single most powerful variable parameter to use for multiple determinations of (ψ, Δ) under different conditions.[7]

In Section 5.2 of this chapter, a broad overview of the different ellipsometers is provided including a general comparison of the design features and their relative advantages and disadvantages for various applications.

Section 5.3 will address the practical details of rotating polarizer and analyzer instrument designs, including the issues that impact the selection of either the rotating polarizer or rotating analyzer instruments. This section will also address the selection of optical components, the spectral range accessible to the different component selections, the alignment of optical elements and the overall instrument, and, finally, the design of the instrument control and data acquisition electronics and related instrumentation.

In Section 5.4, the mathematical descriptions of the polarization generation and detection arms will be provided for both the rotating polarizer and analyzer instruments using Stokes vectors and Mueller matrices. In addition, the incorporation of a compensator into both instrument designs will be treated in a general way through the adoption of an arbitrary Mueller matrix in the mathematical description. In this section, the theoretical treatment will be applied to establish data reduction procedures for instruments with and without compensators. Modifications of the data reduction equations are described that include the effects of errors without approximation. The appropriate parameters that characterize the errors are determined from measurements of standard samples as described in this section or from calibration procedures as described in Section 5.6.

In Section 5.5, the factors that control the precision of the rotating polarizer and analyzer ellipsometers are described. In particular, optimum approaches for operating the instrument so as to achieve the highest possible precision and accuracy are discussed.

Calibration of the rotating polarizer and analyzer ellipsometers is discussed in Section 5.6, and the expansion of such procedures to include the compensator is treated in this section as well. An important capability of the calibration procedures involves the determination of a number of correction parameters, in addition to the polarizer, compensator, and analyzer scale reading corrections. These correction parameters characterize the effects of optical activity in the optical elements, apparent dichroic effects in the compensator, source polarization, polarization sensitivity in the detection system, and errors in the electronic system.

A summary of this chapter and anticipated future directions in rotating-element ellipsometry will be provided in Section 5.7. The status of various rotating polarizer and analyzer instruments for use in specialized applications will be described briefly along with their current limitations. These applications include variable-angle spectroscopic ellipsometry, infrared and ultraviolet spectroscopic ellipsometry, different forms of imaging ellipsometry, and real time spectroscopic ellipsometry. The emphasis here is on the design commonalities of these instruments and how their limitations can be overcome in future instrumentation research.

5.2 Comparison of Ellipsometers

Two forms of automatic or self-compensating ellipsometers have been developed. The first is a straightforward extension of the classical null ellipsometer that involves attaching stepping motors to the adjustable elements.[15] This instrument has the same form as the generic ellipsometer of Fig. 5.1 and includes a collimated quasi-monochromatic beam originating from a broad-band source and monochromator, a stepping motor controlled polarizer and a fixed compensator (not optional in this case), the reflecting sample, a stepping motor controlled analyzer, and a photomultiplier tube (PMT) detection system. The compensator is usually set at or near the quarter wave point $\delta_C = 90°$ with $C' = \pm 45°$. Measurements are made with this system by adjusting the angle of the polarizer under computer control such that, upon reflection, the sample exactly reverses the ellipticity angle imparted to the beam by the polarizer/compensator combination. Then the reflected light is linearly polarized and can be fully extinguished (nulled) by adjusting the angle of the analyzer also under computer control. The resulting angles of the polarizer and analyzer at the null provide the ellipsometric angles (ψ, Δ). Although the manual process of nulling requires several minutes of tedious operation, the automatic stepping motor system can provide (ψ, Δ) in 1 s or less.

In this ellipsometer, it is also possible to place the compensator after the sample, thus using the compensator to exactly reverse the ellipticity angle imparted by the sample. If one has a choice, however, it is generally best to place the compensator before the sample. When positioned after the sample, the compensator may exhibit slightly different polarization modification characteristics (ψ_C, δ_C) due to small differences in beam path generated in turn by small differences in sample alignment. When positioned before the sample, these polarization modification characteristics are immune to sample alignment variations, and will only vary due to source alignment changes, which occur only when lamps are changed. Here, the compensator transmission characteristics (ψ_C, δ_C) are defined

analogously to the sample reflection characteristics by $\tan\psi_C \exp(i\delta_C) = t_F/t_S$. Here t_F and t_S are the complex amplitude transmission coefficients (the transmitted to incident electric field ratios) for the F and S linear polarization directions (parallel to the fast and slow axes of the compensator, respectively). In the following discussions of different ellipsometer designs, it is assumed that the compensator (if incorporated) is placed *before the sample*.

A schematic of the second version of the self-compensating ellipsometer and its associated electronics are shown together in Fig. 5.2. In this instrument, the null is established by changing the tilt angle Q of the polarization state at two stages in the beam path (where the polarization state is linear at the desired null settings): (i) between the initial fixed polarizer and compensator and (ii) between the sample and the fixed analyzer. This change in Q is performed for the instrument of Fig. 5.2 by two Faraday cells mounted after the polarizer and before the analyzer,

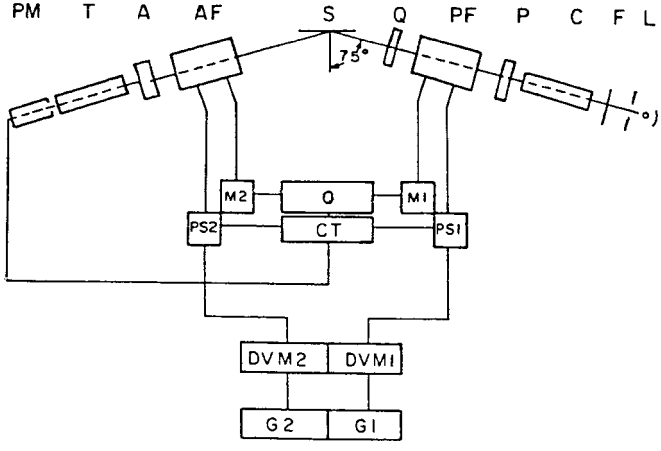


Figure 5.2a Block diagram of the optical components of a self-compensating ellipsometer based on Faraday cell polarization compensation and modulation. This instrument is designed for high speed operation at a selectable wavelength. (Key: A – analyzer; AF – analyzer Faraday cell; C – collimator; CT – controller; DVM1 – digital voltmeter indicating polarizer rotation; DVM2 – digital voltmeter indicating analyzer rotation; F – interference filter for wavelength selection; G1 – galvanometer - oscillograph for polarizer rotation; G2 – galvanometer - oscillograph for analyzer rotation; L – arc lamp; M1 – polarizer modulation power supply; M2 – analyzer modulation power supply; O – oscillator; P – polarizer; PF – polarizer Faraday cell; PM – photomultiplier tube; PS1 – polarizer compensation power supply; PS2 – analyzer compensation power supply; Q – quarterwave plate; S – sample; T – telescope). [After Ref. 16.]

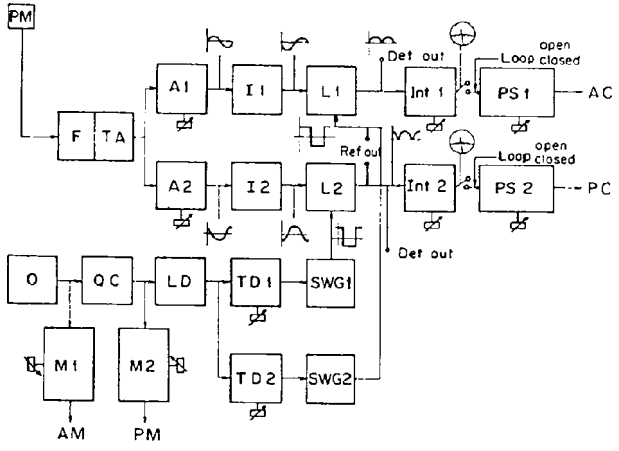


Figure 5.2b Block diagram of the electronic system of a self-compensating ellipsometer based on Faraday cell polarization compensation and modulation. This instrument is designed for high speed operation at a selectable wavelength. (Key: A1, A2 – amplifiers; AC – analyzer Faraday solenoid for compensation; AM – analyzer Faraday solenoid for modulation; F – follower; Int 1, Int 2 – integrators; I1, I2 – inverting operational amplifiers; L1, L2 – phase sensitive detectors; LD – level detector; M1, M2 – modulation amplifiers; O – oscillator; PC – polarizer Faraday solenoid for compensation; PM – polarizer Faraday solenoid for modulation; PS 1, PS 2 – compensation power supplies; QC – quadrature circuit; SWG1, SWG2 – square wave generators; TA – tuned amplifier; TD1, TD2 – multivibrators. [After Ref. 16.]

although Pockels cells have also been used for this purpose. In a Faraday cell, a magnetic field H is applied to a transparent core material parallel to the direction of propagation of the light beam through the core. Upon transmission through the core, the electric field vector associated with a linearly polarized beam is rotated by an angle ΔQ that is proportional to H and to the length of the core. In the ellipsometer of Fig. 5.2, compensation and modulation currents are supplied to separate coils of the Faraday cells. The compensation currents rotate the linear polarization state, whereas the modulation currents generate an error signal at the detector that drives a control loop and directs the system to the desired null. For this instrument a single measurement of (ψ, Δ) can be obtained in 1 ms. In fact, the instrument is so fast that it has been applied for real time spectroscopic ellipsometry incorporating a rotating interference filter for high-speed spectral scanning.[17]

The phase modulation ellipsometer has the same series of elements as the stepping motor-controlled self-compensating ellipsometer with the

exception that the fixed compensator in the latter is replaced by a birefringent phase modulator such as a piezobirefringent plate or a Pockels cell. For these elements, the retardance varies sinusoidally as a function of time rather than being fixed as for the compensator of the self-compensating ellipsometers. In the phase modulation ellipsometer, both the polarizer and analyzer angles are fixed during the measurement. Two different electronic system designs are possible for this ellipsometer. The output signal of the detector, usually a PMT, can be analyzed either by phase sensitive detection[13,14] or by high speed digitization followed by discrete Fourier analysis.[18] Figure 5.3 shows an example of an ellipsometer and associated electronics based on the latter design. In either design, the dc level as well as the ω and 2ω harmonic constituents of the waveform provide the ellipsometric angles (ψ, Δ). For the most common phase modulation instrument based on a silica piezobirefringent plate, the modulation frequency $\omega = 50$ kHz provides a minimum measurement time of 20 μs when a single cycle is Fourier analyzed. Such high speed data collection even allows one to perform real time spectroscopic ellipsometry in a serial fashion by using a photodiode array detection system and high-speed serial pixel scanning for digitization.[19]

For the rotating polarizer and analyzer ellipsometers, the series of optical elements is also the same as that of the stepping-motor controlled self-compensating ellipsometer; however, for the former two instruments the compensator is optional.[10–12] In fact, the compensator is removed for the simplest, nearly achromatic version of the instrument. In the original rotating polarizer or analyzer ellipsometer designs, the polarizer or analyzer is mounted within the hollow shaft of a dc motor and rotated continuously at a stabilized frequency value within the range of 10 to 100 Hz.[20] Fourier analysis provides the dc and 2ω Fourier coefficients of the waveform which can provide (ψ, Δ) with a minimum measurement time ranging from 5 ms (at 100 Hz) to 50 ms (at 10 Hz). Figure 5.4 shows a high precision rotating analyzer instrument and its associated electronic circuitry. Although slower than the highest speed self-compensating and phase modulation ellipsometers, the rotating element ellipsometers are well suited for use with integrating photodiode array detectors whereby several spectra consisting of waveform integrals can be collected per optical cycle (π/ω) of the rotating element.[21] From these several detector scans, spectra in (ψ, Δ) can be deduced.

With this brief instrumentation background as illustrated in Figs. 5.2 through 5.4, the advantages and disadvantages of the rotating polarizer and analyzer instruments relative to the most widely used alternatives will be discussed. Because the ellipsometers most successful in advanced research are based on self-compensating, phase modulation, or rotating element principles, this is often the most important choice that the prospective ellipsometry user faces in building or purchasing an instrument.

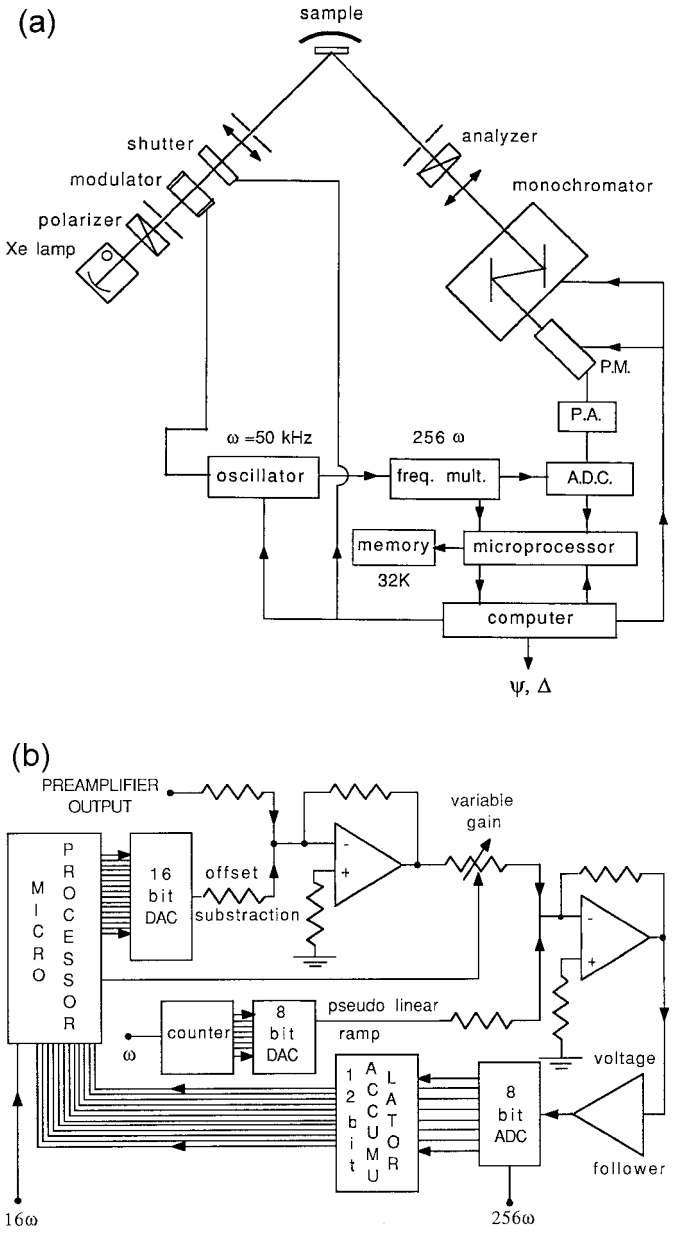


Figure 5.3 Schematic diagrams of (a) the optical configuration and (b) the signal processing system for a polarization modulation spectroscopic ellipsometer designed for real time studies of thin film preparation at a selectable photon energy. [After Ref. 18.]

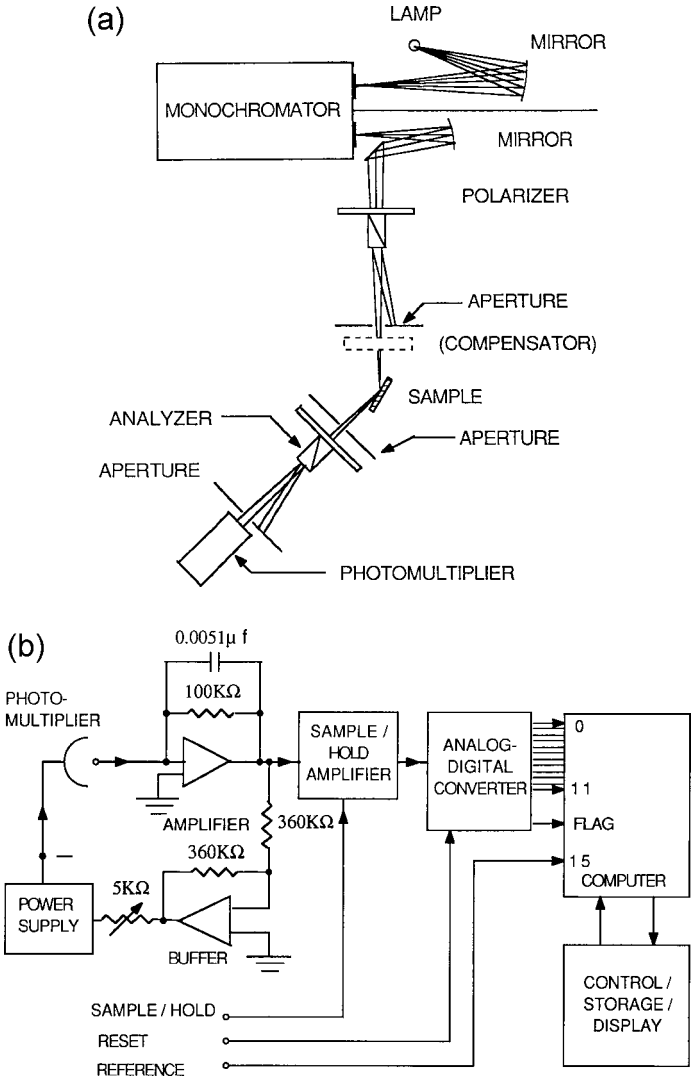


Figure 5.4 Schematic diagrams of (a) optical elements and (b) signal processing system for a rotating analyzer spectroscopic ellipsometer designed for high precision measurements of the optical properties of solids. [After Ref. 11.]

The precision of all three automatic systems has been evaluated in detail, considering the most favorable sample (ψ, Δ) values for each.[22] Based on this evaluation, however, it is clear that no one system has a major advantage over the others. Typical precision values are in the range $2\delta\psi \approx \delta\Delta \approx$ (0.001° to 0.005°. For the self-compensating ellipsometer,

however, a higher beam irradiance is needed to reach the shot-noise limit; at low irradiances the precision is reduced due to the detector dark current and its associated noise. This problem may have a secondary effect in reducing the spectral range accessible to the instrument. One of the primary issues in assessing the advantages and disadvantages of the instruments involves the range of so-called "favorable" (ψ, Δ) values. In this respect, the self-compensating instrument has a clear advantage in that it measures (ψ, Δ) directly through the polarizer and analyzer angular values at null. Thus, this instrument achieves its maximum precision irrespective of the values of (ψ, Δ) for the sample. As a result, it is an effective instrument for the measurement of transparent dielectrics, semiconductors above and below the fundamental gap, and metals. In contrast, for the simplest versions of the other two automatic instruments, the maximum precision is achieved over only specific ranges of (ψ, Δ). As a result, instrumentation complexities have been introduced in these cases to eliminate the ranges of low precision.

For the phase modulation ellipsometer, two common instrument configurations are employed. These are identified as configurations II and III, following the original notation of Bermudez and Ritz.[14] In configuration II, the polarizer, modulator, and analyzer angles are set at the possibilities (P', M', A') = (±45°, 0°, ±45°) or (±45°, 90°, ±45°). In this configuration, $S \equiv \sin 2\psi \sin\Delta$ and $C \equiv \sin 2\psi \cos\Delta$ are measured and so the instrument exhibits poor precision in ψ when $(d/d\psi)(\sin 2\psi)$ vanishes, or when $\psi = 45°$. Thus, this configuration has poor precision in ψ for highly metallic surfaces in which case (ψ, Δ) ~ (45°, 90°). In configuration III, (P', M', A') = (0°, ±45°, ±45°) or (90°, ±45°, ±45°), and S and $N \equiv \cos 2\psi$ are measured. In this case, the system exhibits poor precision in Δ when $(d/d\Delta)(\sin\Delta)$ vanishes or when $\Delta = \pm 90°$, again a problem for highly metallic surfaces. Such problems are eliminated for measurements in both configurations II and III in which case the poor sensitivity ranges can be avoided completely. However, rather than rotating the polarizer and modulator from one set of values to another in order to access the two configurations and (N, C, S), Jellison has adopted a beam splitting analyzer consisting of a Wollaston prism.[23] By measuring the irradiance in the orthogonal polarization states at two PMT detectors, N, S, and C are accessible in one optical configuration, an important enhancement to the phase modulation instrument.

For the rotating polarizer and rotating analyzer ellipsometers, incorporating compensators in each case with $C' = 0$, the measured quantities are $\tan\psi$ and $\cos(\Delta + \delta_C)$. In these cases, the precision in measuring Δ is poor when $\Delta = -\delta_C$ or $\Delta = \pm 180° - \delta_C$ [Ref. 24]. In contrast, in the absence of a compensator ($\delta_C = 0°$), the precision of the rotating polarizer and analyzer instruments is poor when $\Delta = 0°$ or $\Delta = \pm 180°$. Thus, weakly absorbing dielectric materials cannot be measured with precision

for these instruments in the absence of a compensator. When the compensator is included and is set for quarter wave operation, the precision becomes poor when $\Delta = \pm 90°$, preventing precise measurement of metallic materials. For the rotating polarizer and analyzer instruments, performance is significantly enhanced with an automatically adjustable compensator. In this enhancement, an initial measurement of (ψ, Δ) yields lower precision results which are then improved by a second measurement with δ_C set at $\delta_C = (-\Delta \pm 90°)$ or $\delta_C = (-\Delta + 270°)$.[25]

Based on the previous discussions, enhancements can be made to the phase modulated ellipsometer and the rotating polarizer/analyzer ellipsometers to ensure high precision operation irrespective of (ψ, Δ), as is possible with the self-compensating instrument. It should also be noted that in ranges where the precision is poor, accuracy will be poor as well since experimental errors will increase in these ranges. Thus the enhancements just described improve overall performance of the instruments, not simply the precision. Next, the issues of accuracy, speed for real time applications, and spectral range for specroscopic measurements will be discussed for the three instruments.

The highest accuracy ellipsometric measurements, typically within the range 0.01° to 0.05° in (ψ, Δ), are achieved through four-zone averaging with a compensating or null ellipsometer, since the averaging is known to eliminate many possible instrument errors.[1] With the trend toward automation and higher speed necessitated by the desire to perform real time and spectroscopic ellipsometry, the four-zone average is rarely used in self-compensating ellipsometry.[17] However, the self-compensating ellipsometer does maintain several advantages over its photometric alternatives. First, as noted above, high accuracy can be maintained over all regions of (ψ, Δ), in contrast to the unenhanced designs used for the phase modulation and rotating polarizer/analyzer instruments. Second, because self-compensating instruments seek a null at the detector, they are not adversely influenced by detection system nonlinearities as are the photometric instruments.

The primary optical element errors in all instrument designs arise due to the compensator[26] or phase modulator.[27] Introducing such an element is not unlike incorporating a second unknown sample with wavelength dependent properties and possible non-idealities into the instrument (in addition to the sample to be measured at the central ellipsometer axis). Documented non-idealities for compensators include multiple reflections between the faces of the device, beam displacements and deviations, plate misalignment particularly for biplate designs, dichroic characteristics, slight ellipticity of the normal modes due to optical activity for quartz devices, and the temperature dependence of the retardance. Non-idealities for vitreous silica phase modulators include multiple reflections, higher

harmonic generation, static and non-uniform birefringence, nonuniformity or instability of the principal axes of birefringence, and temperature dependent effects. Although expert developers of ellipsometry instrumentation have established methods of minimizing or correcting these problems, it is generally recognized that they collectively represent the primary source of inaccuracy in these ellipsometers at the level of 0.05° to 0.5°.

In such cases where accuracy is limited by the compensator, the rotating polarizer and analyzer ellipsometers without such elements can provide the highest accuracy design, as long as Δ for the sample lies within the range given by $20° < |\Delta| < 160°$. To reach such high levels of accuracy for these instruments, provisions for eliminating the remaining dominant errors must be adopted. These include detection system nonlinearity, residual source polarization for the rotating polarizer design, and detector polarization sensitivity for the rotating analyzer design. These issues will be discussed further in Sections 5.3, 5.4, and 5.6. It is no accident that the first continuously spectroscopic dielectric function data for absorbing materials, in particular for semiconductors above the lowest direct gap, were obtained with a rotating analyzer instrument without a compensator.[28] Although such dielectric function data are continually being improved over the years, the early results for many materials have remained a standard to which recent data sets are being compared.[29,30]

Besides the issues of precision and accuracy, those of time resolution for real time measurements and spectral range for spectroscopic measurements are important in comparing ellipsometer designs. These issues will be discussed in the next several paragraphs.

The phase modulation ellipsometer is the clear choice in real time, single wavelength applications in which high speed is the primary requirement. An ellipsometer using a Pockels cell modulator has provided a measurement time for successive (ψ, Δ) values of 4 μs,[31] whereas the conventional piezobirefringent modulator provides a minimum measurement time of 20 μs.[18] In contrast, the self-compensating and rotating polarizer/analyzer instruments are slower (1-50 ms per measurement). Applications for high-speed data acquisition in phase modulation ellipsometry include electrochemical reactions at surfaces, laser annealing and ion bombardment modification of surfaces, and the deposition and etching of thin films at high rates.[32] For such instruments, however, one must consider the application and in particular the required sensitivity in surface processes since the precision of measurement decreases significantly for measurement times less than 1 ms.[33] In this range, the submonolayer precision characteristic of ellipsometry achieved when $2\delta\psi \approx \delta\Delta < 0.1°$ in fact may be lost.

The moderately high speed of the self-compensating ellipsometer (1 ms per measurement) has been employed extensively in electrochemical studies for real time single wavelength and serially scanning spectroscopic ellipsometry.[35,36] Because such measurements are susceptible to random and systematic errors due to ambient light that can mask the null position, self-compensating ellipsometers have serious disadvantages for real time analysis of thin films in plasma deposition processes or in high temperature processes involving thermally emitting samples.

Finally, it should be pointed out that the highest data acquisition speed for full spectroscopic ellipsometry, ~15 ms per (ψ, Δ) spectra, has been attained with an instrument exhibiting the lowest speed for real time single wavelength measurements, namely the rotating polarizer ellipsometer.[34] For this instrument, the readout cycle of a photodiode array detector can be synchronized to the polarizer rotation cycle so that waveform spectra can be collected in parallel from a minimum of three readouts, each requiring 5 ms. Thus, the rotating polarizer/analyzer instruments provide the greatest promise for high-speed measurements where full spectroscopy is desired. In fact, in a 5 ms photodiode readout time, up to 1024 wavelengths can be collected using a commercial photodiode array system. Pixel grouping in the software can be used to tailor the spectrum to the desired wavelength or photon energy resolution.

In assessing the advantages and disadvantages of the three instrument designs for spectroscopic applications in general, one must consider the limitations on the spectral range due to the optical elements used in these designs. Here, it is assumed that a broad spectral range from the near-infrared to the ultraviolet is desired in order to detect interband transitions for metals and semiconductors with a wide range of band gaps. Because polarizers and compensators can be fabricated from MgF_2, an optical material with a very wide range of transmission from ~100 to 1000 nm and beyond, stepping motor controlled self-compensating ellipsometers and rotating polarizer/analyzer ellipsometers in principle can cover this full range, providing suitable sources (e.g., Xe arc, D_2 discharge, and W filament lamps) and detectors (e.g., PMTs and Si and Ge photodiodes) are incorporated. The problem with the self-compensating ellipsometer is that noted earlier, namely, its lower precision and accuracy under the low light levels typically encountered in the extremes of the spectra at short and long wavelengths.

In designing an instrument with a wide spectral range, compensator selection can be a limiting factor, as the retardance may cross the half wave points at the upper and lower reaches of the spectrum.[37] Thus, one may seek to adjust the compensator retardance as one scans through the spectrum. The traditional compensator for this purpose has been of the

Babinet-Soleil design constructed from quartz, a material that restricts the short wavelength limit of the spectral range compared to MgF_2 (and also exhibits troublesome optical activity). In a recent advance, however, the need to adjust the compensator to remain near $\delta_C = 90°$ (quarter wave) has been eliminated in the rotating analyzer configuration by the use of a vitreous silica, three-reflection achromatic compensator, while also restricting the short wavelength limit.[38] More recently still, auto-compensators of the tilting-plate or Berek design that allow the use of MgF_2, and hence the widest spectral range, have been incorporated in rotating polarizer/analyzer instruments, not only to maintain operation near quarter wave points, but also to maintain high precision and accuracy independent of the sample Δ value as described above.[39]

Because of the use of the vitreous silica piezobirefringent plate, the most popular phase modulation ellipsometer is not as suitable for extension of the instrument well into the ultraviolet as are the rotating polarizer and analyzer ellipsometers. In addition, the self-compensating ellipsometer based on Faraday rotation should be excluded from consideration, if the goal is to reach short wavelengths. The high Verdet coefficient glasses required for the Faraday cell cores (e.g., SF6 glass) limit the lowest accessible wavelength to 350 nm.[17] In summary, if the widest possible spectral range is desired, especially deep into the ultraviolet, the rotating polarizer/analyzer ellipsometers are the best choice since they can be designed solely with MgF_2 optical elements and can be operated at low light levels.[40]

5.3 Instrumentation Issues

5.3.1 Optical Configuration

There are at least four possible variations of the optical configuration for the rotating element ellipsometers. First, either the polarizer or analyzer can rotate, and in principle the compensator can be placed either before or after the sample. The placement of the compensator just before the sample has been justified in detail earlier. The choice between the rotating polarizer and rotating analyzer ellipsometer configurations, however, can be a difficult one, as it depends sensitively on the desired application. Here selection criteria are discussed, as well as the basic optical design of each of the separate instruments. As will be presented in Sections 5.4.2 and 5.4.3, residual polarization of the source in the rotating polarizer ellipsometer[41] and polarization sensitivity of the detection system in the rotating analyzer ellipsometer[42] are the predominant errors that must be minimized for high accuracy operation.

The first rotating element ellipsometer designed for high precision and accuracy, as well as for a wide spectral range, was based on the rotating analyzer design as shown in Fig. 5.4.[11] Initially, instruments such as this one will be considered that do not employ fiber optics in coupling light from the source to the polarizer and from the analyzer to the detector. For such instruments, errors due to weak source polarization in front of a rotating polarizer tend to be greater and more difficult to minimize than the corresponding errors due to polarization sensitivity of the detection system behind a rotating analyzer. Errors due to polarization sensitivity of the detection system in the latter configuration are minimized through wavelength selection using a monochromator before the fixed polarizer as in Fig. 5.4 and using a single PMT as a detector behind the rotating analyzer. This avoids the severe polarization sensitivity of the detection system that would result with a spectrometer after the rotating analyzer. In addition, for the detection system of a rotating analyzer ellipsometer, side-viewing PMTs are to be avoided due to the dependence of the transmittance and reflectance characteristics of the PMT envelope on the linear polarization state when the beam impinges non-normally at their interfaces. In contrast, for a front-viewing PMT the full light beam crosses the envelope at normal incidence and the only polarization dependence then arises from anisotropic optical effects in the detection system.

For ex situ measurements on an optics bench without fiber optics, the rotating analyzer ellipsometer with a single PMT behind the rotating analyzer leads to the highest accuracy.[11] In real time measurements of thin films in light emitting environments, this design is susceptible to serious errors due to collection of stray light. For stray light sources that are unchanging with time, corrections are straightforward. When the stray light levels vary rapidly, however, correction is generally not feasible. In this case, it is important to select the wavelength on the detection side of the instrument with a spectrometer after the analyzer. This instrument redesign along with high f/# optics to couple the collimated beam to the input slit of the spectrometer significantly reduces stray light. It is then necessary to fix the analyzer and revert to the rotating polarizer configuration.[12] In the rotating polarizer configuration, the lamp is aligned in its initial setup to minimize source polarization. If data collection is designed to provide the amplitudes of the 4ω Fourier coefficients in real time during lamp alignment, then the minimization of these coefficients can be used as a criterion to minimize source polarization. The difficulty of the source polarization problem is that the required corrections may gradually change as the lamp degrades and abruptly change when the lamp is replaced. In contrast, detector polarization sensitivity remains essentially unchanged with long-term use of the instrument. Additional details of the correction procedures are provided in Sections 5.4 and 5.6. Finally,

reversion to the rotating polarizer configuration imposes more stringent requirements on the maximum tolerable periodic beam deviations as the polarizer rotates. This difficulty can be suppressed with judicious use of apertures along the beam path.

In general, fiber optics systems on the source and detection sides are not desirable if the widest possible spectral range is sought. Such systems lead to enhanced reflection losses in the coupling optics and absorption losses in the fiber itself, particularly in the ultraviolet range. However, relatively long fibers can act as polarization scramblers in which case one can add the spectrograph at the end of the output fiber on the detection side of the rotating analyzer ellipsometer. Other pseudo-depolarizing elements such as opal glass plates or quartz wedges can also be used after the rotating analyzer to perform the same functions again, with possible reductions in spectral range. In this way, one can avoid having to change the ellipsometer configuration to the rotating polarizer system when stray light rejection is sought for in situ and real time measurements.

5.3.2 Optical Components and Spectral Range

The selection of optical components depends sensitively on the desired application just as does the selection of the instrument configuration described in Section 5.3.1. Here, components will be considered that provide a wide spectral range spanning the near-infrared to ultraviolet. Ellipsometers that operate in the infrared and vacuum ultraviolet are discussed in specialized chapters.

Three sources are possible for rotating element spectroscopic ellipsometry from the near-infrared to ultraviolet: (i) tungsten-halogen lamps that span the near-infrared to visible range (400 to $>$1000 nm); (ii) Xe lamps that span the near-infrared to near-ultraviolet range (200 to 800 nm); and D_2 lamps that span from the violet deeper into the ultraviolet ($<$150 to 400 nm). In general these lamps employ vitreous silica envelopes, however, in order to extend below ~175 nm, a MgF_2 window on the D_2 lamp is required. Because of its wide range, the most common single source is the Xe lamp. The most common Xe lamp used for ellipsometry operates at 75 W with a short arc. Higher power lamps are operated with larger electrode gaps, in which case the arc is susceptible to wandering. This leads to undesirable non-random noise fluctuations. Furthermore, although the total irradiance output of these lamps increases with increasing power rating, the brightness of the arc does not increase proportionally due to the larger emission volume. Even with the 75 W short arc lamp, recent studies have suggested that the lamp axis should be oriented orthogonal to the long direction of the monochromator or spectrometer

slits in order to minimize noise due to arc instability.[43] In addition to single sources, a dual combination of a Xe lamp placed behind a D_2 lamp provides a very wide spectral range and has been used for real time spectroscopic ellipsometry in the rotating polarizer configuration, spanning the range from 190 to 825 nm in a single parallel measurement.[40] By placing an adjustable aperture between the two lamps, the spectral output of the system can be flattened, which is a desirable source attribute for real time spectroscopic measurements.

Mirror optics such as the off-axis paraboloid are ideal for focusing the light from the lamp directly onto the fiber optic or monochromator input (for the rotating analyzer configuration) or onto the pinhole of a collimator (for the rotating polarizer configuration). This prevents chromatic aberrations that can lead to variations in alignment versus wavelength across the ellipsometric spectra. Alternatively, achromatic lenses are possible if a co-linear optical bench is desired. In some cases, one can make effective use of chromatic aberrations in focusing lenses to flatten the spectral output of the source. For example, if the source is too intense in the red relative to the blue, one can set the focus to maximize the blue throughput. In establishing the beam path through the entire instrument, the light from the monochromator or directly from the source can be focused on the sample or the detector, or it can be collimated. If a focusing approach is adopted, one must ensure that the beam does not exceed the specified entrance cones of the optical elements. In addition, the incident angle dependence of the ellipsometric spectra must be sufficiently weak to justify such an approach. If chromatic aberrations are present in the beam, it is helpful to use apertures to define the beam at various stages [usually at the polarizer, compensator (if used), analyzer, and detector]. Apertures also assist in alignment as described in Section 5.3.3.

In selecting a monochromator or spectrometer for rotating element ellipsometry, one generally relies on commercial instrumentation with the key characteristics being spectral range, spectral resolution, and stray light. Perhaps the most useful instruments for spectroscopic ellipsometry employ a prism-grating combination, an approach that is no longer widely available. Now most commercially available monochromators and spectrometers employ single or multiple gratings. In establishing the spectral range of the grating instruments, the selection criterion is the grating blaze. If one is using the instrument with a Xe source to span the range from 200 to 800 nm, a 250 or 300 nm blaze should be selected.[40] Otherwise the combined sharp fall-off of the Xe source output with that of the grating efficiency at shorter wavelength will prevent one from extending the spectral range below 250 nm in the ultraviolet. Such a sharp fall-off in monochromator output in rotating analyzer systems can lead to alignment errors as the image of the source spot moves from the center of the exit slit to the

edge with decreasing wavelength. This is one advantage of placing a spectrometer after the analyzer in rotating element ellipsometry.

An additional consideration in reaching the lower wavelength limit of 200 nm is stray light in the monochromator/spectrometer. In general, it is desirable to have two or more grating stages in order to eliminate stray light. A single grating is only successful over a wide range (200–800 nm) if dual sources are used and the spectral output is flattened with an aperture between the sources. The reason is that the stray light output is approximately proportional to total flux entering the input slits, and when the spectral distribution is flat, the ratio of the true light to stray light at a given wavelength is maximized.[44] In addition, with a flat spectral distribution, the image shifts that affect alignment as noted above are eliminated. For instruments with more than one grating stage, the achievable spectral resolution is normally sufficient; however, for single grating instruments care must be taken to ensure that the width of the sharpest optical structures is not limited by the monochromator. A discussion of the optimum slit width as established through a Fourier analysis of semiconductor critical points can be found in the literature.[45]

When using grating monochromators, higher order grating reflections can be troublesome and if not removed, beam components of wavelength λ/n (n: order integer) can follow the same beam path as the desired wavelength λ and lead to significant errors in the measured spectra. In wavelength-scanning spectroscopic ellipsometry, a filter wheel with one or more order sorting filters is needed in front of the monochromator to eliminate this problem. It is a challenge to find filters that eliminate higher orders while retaining a relatively flat spectral output versus wavelength. For real time spectroscopic ellipsometry the filters must be added within the spectrograph.[40] Incorporating a prism stage in the monochromator or spectrograph eliminates the need for order sorting, but such an instrument design drastically limits the possible commercially available sources.

Most spectroscopic ellipsometers use one of three different polarizer designs, the Glan-Taylor made from calcite with an air gap between the two elements, the Rochon made from quartz, and the Rochon made from MgF_2. A detailed description of the operation of the Glan-type polarizers is given elsewhere in this Handbook.[39] In these polarizers, the optic axes of the two polarizer elements are parallel; in addition, they are parallel to the entrance and exit surfaces of the two elements. The linear polarization state perpendicular to the optical axis (the ordinary ray) is eliminated by total internal reflection at the initial calcite/air interface due to the higher index of refraction of calcite for this polarization state. This unwanted beam can be either directed out of the polarizer through a hole in the side of its mount, or absorbed by an index matching black coating at the side of the polarizer. In contrast, in a Rochon polarizer, the optic axis of the

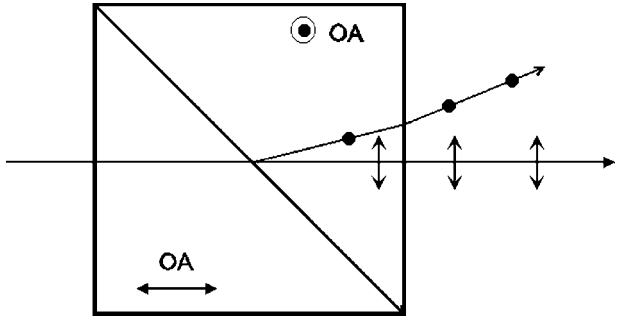


Figure 5.5 A schematic diagram of a Rochon polarizing prism. Such elements can be fabricated from quartz or MgF_2 and have been used extensively in spectroscopic ellipsometers to achieve broad spectral ranges.

first polarizer element is parallel to the beam propagation direction, whereas the optic axis of the second element is perpendicular to the optic axis of the first, parallel to the interface between the two, and also parallel to the entrance and exit surfaces (see Fig. 5.5).[37] At the interface to the two elements of a Rochon polarizer, the extraordinary wave, with polarization parallel to the optic axis of the second element, is deviated at the interface, whereas the ordinary wave is not deviated and constitutes the polarized output beam. An aperture after the polarizer is used to eliminate the unwanted beam.

If spectral range is not a consideration, then the Glan-Taylor polarizers provide the best overall performance due to the relatively wide acceptance angle and the large exit beam angular separation. Thus, the extinction ratio of these polarizers are the best $\sim 10^{-5} - 10^{-6}$. However, if one requires spectroscopic data below about 250 nm, then quartz or MgF_2 Rochons are required. In fact MgF_2 is the best for reaching below 200 nm. The most serious problem for the quartz Rochon polarizers arises from the optical activity of quartz.[46] The simplest manifestation of this effect is the rotation of the incident linear polarization state for propagation along the optic axis, e.g., for propagation within the first element of the Rochon polarizer. Because of this effect, the orientation of the polarizer must be reversed when it is used as an analyzer. To understand this, it should be noted that the first element of the polarizer acts as a partial depolarizer of the incident beam and the second element of the analyzer acts as a partial depolarizer of the exiting beam. The depolarization results from the different distances that different beam components travel within the first polarizer and second analyzer elements. Partial depolarization also occurs versus wavelength due to the optical rotatory dispersion of the quartz. When the

quartz Rochon is used as the rotating polarizer, one interesting effect of the dispersion is the rotation of the direction of source polarization.[41] In this case the source polarization errors are oscillatory with wavelength and quite easy to detect. Then the lamp alignment can be adjusted to eliminate these oscillations. When a Glan-Taylor or MgF_2 Rochon polarizer is used, the residual source polarization is more difficult to detect (usually through the 4ω Fourier components as noted earlier). The other critical manifestation of optical activity is that the two normal modes of the polarizer, i.e., the polarization states that form the deviated and undeviated beams, are elliptical with opposite handednesses (clockwise and counterclockwise rotation of the electric field vector).[46] In fact the transmitted handedness of the ellipse differs for the two enantiomorphs of α-quartz. The ellipticity of the normal modes, i.e., the ratio of the length of the semiminor to that of the semimajor axis, is quite small for the quartz Rochon polarizers, ~0.001 times the photon energy. In spite of this, optical activity corrections are needed in order to deduce (ψ, Δ) with an accuracy better than 0.1°.

Because of the complexity of optical activity corrections for quartz Rochons and the wider spectral range for MgF_2 Rochons, the latter are preferred. However, the advantages of MgF_2 are not gained without some costs. Rochon polarizers exhibit relatively small normal mode angular deviations, typically 2 to 3°, and as a result the beam paths between the polarizer and the compensator and the analyzer and the detector must be increased to ensure sufficient beam separation. Larger deviations are possible with larger angles of incidence at the interface between the two elements thus requiring longer crystals, however, such polarizers are more difficult to manufacture from MgF_2 and, thus, are more costly. Finally because of the smaller separation between the two beams, the extinction ratio of the MgF_2 polarizers is the poorest of all the crystal polarizers and the acceptance angle is quite narrow as well. Thus, extra care must be taken in beam collimation and alignment.

Next, the different compensators used in rotating polarizer/analyzer ellipsometers will be discussed. The most widely used adjustable compensator in compensating (or null) ellipsometry is the Babinet-Soleil type.[1] In fact, for the rotating analyzer ellipsometer of Fig. 5.4, the Babinet-Soleil compensator was also considered as an optional component for the measurement of dielectric materials or semiconductors below the lowest direct band gap.[46] The Babinet-Soleil compensator is fabricated from two quartz wedges that slide against one another to generate an adjustable thickness. The optical axes of both wedges are parallel, lying in the plane of the outer wedge surfaces, and parallel to the sliding direction. The two wedges are attached to a fixed thickness plate having an optic axis in the plane of the plate and orthogonal to the optic axes of the two wedges.

More recently, owing to the combination of errors that can occur, Babinet-Soleil compensators have been replaced by other types. First of all, the three quartz plates of the Babinet-Soleil compensator exhibit optical activity as in the case of the Rochon polarizers. This leads to very complicated corrections in calibration and data reduction. Second, if the optic axes of the wedges and the fixed plate are not precisely orthogonal, then high frequency oscillatory artifacts versus wavelength are introduced in the ellipsometric spectra.[26] Thus, the plates must be mounted in a fixture that allows the fixed plate to be rotated with respect to the two wedges in an alignment step and then clamped at the precise orthogonal position. This is straightforward for a fixed thickness biplate compensator, but more complicated for the adjustable thickness Babinet-Soleil compensator. Finally, for effective elimination of such artifacts, the wedges and the fixed plate of the Babinet-Soleil compensator must be enantiomorphs, i.e., constructed from quartz of opposite handednesses.

For compensators of fixed retardance (often called retarders) three different designs are used most widely in rotating polarizer/analyzer ellipsometers. First, a retarder constructed from two fixed MgF_2 plates, with orthogonal optic axes both lying in the plane of the plates, can operate over a very wide spectral range. This device has the key advantage over the Babinet-Soleil compensator in that optical activity corrections are not required, however, it is still necessary to ensure that the optic axes of the two fixed plates are orthogonal using the alignment fixture just described.[26] If the plates are not orthogonal, oscillatory artifacts appear versus wavelength. The two disadvantages of the biplate design include the inability to adjust the retardance, as well as the strong wavelength dependence of the retardance. This wavelength dependence is minimized, however, by selecting the thickness difference to be small such that the compensator operates in lowest order. A second compensator is of the Berek design that can be fabricated from a single plate of MgF_2 with the optic axis nearly normal to the plate but tilted at a small angle with respect to its surface. For example, a tilt angle of 9.5° yields a quarter wave wavelength of 630 nm.[39] This has the important advantage over the biplate in that the internal alignment of the compensator is not necessary. A third compensator used in rotating analyzer ellipsometry is a vitreous silica, three-reflection rhomb (see Fig. 5.6). Two internal reflections occur at silica/air interfaces at an angle of incidence of 72.25°, and one reflection occurs at a silica/MgF_2(275 Å)/air structure at an angle of incidence of 54.50°.[38,47] The three reflections yield a retardance near 90° that is nearly independent of wavelength. Thus, achromicity is an important advantage of this retarder over the MgF_2 biplate and Berek designs. Its disadvantages include its difficulty to align, its susceptibility to stress-induced birefringence that requires careful stress-free mounting, and the fact that the

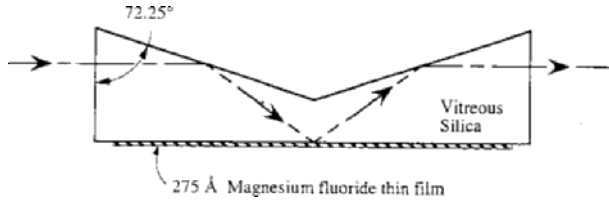


Figure 5.6 A schematic diagram of an achromatic, three-reflection compensator. In this element, two internal reflections occur at silica/air interfaces at an angle of incidence of 72.25°, and one reflection occurs at a silica/MgF_2(275 Å)/air structure at an angle of incidence of 54.50°. The three reflections yield a retardance near 90° that is nearly independent of wavelength. [After Ref. 47.]

transmission coefficient amplitudes $|t_F|$ and $|t_S|$ along the fast and slow axes of the device differ measurably. Because of the latter fact, two spectra are needed to characterize the retarder, namely $\psi_C = \tan^{-1}(|t_F|/|t_S|)$ and the retardance $\delta_C = \delta_F - \delta_S$.

Finally, a compensator with an adjustable retardance for use in rotating polarizer/analyzer ellipsometry can be designed using a single MgF_2 plate and applying the Berek configuration.[39] In this compensator, the optic axis of the plate is perpendicular to its surface, and the plate is tilted at an adjustable angle about an axis that is in the plane of the plate and orthogonal to the beam direction. As a result, a small angle is generated between the optic axis and the propagation direction. This forces orthogonal linear polarizations to travel at different speeds within the material and a quarterwave phase shift can be generated at an appropriately chosen angle (typically less than 10° for the desired wavelength). In addition to the phase shift, the two polarizations are slightly displaced yet emerge parallel, and the transmitted beam is slightly displaced from the incident beam, but parallel as well. The effect of these displacements, although small, can be eliminated by using an aperture after the compensator. Because of its adjustability, simplicity, and wide spectral range, as well as the lack of spectral artifacts associated with biplates, the Berek-type compensator is a good choice for rotating polarizer/analyzer scanning spectroscopic ellipsometers, i.e., those that collect spectra point by point.

5.3.3 Alignment

In the ellipsometer alignment process, the axes of the polarization generation and detection arms of the instrument, along which the incident and reflected beams travel, are oriented such that they intersect at a point, thus establishing the plane of incidence (see Fig. 5.1). The intersection

point must occur at the central ellipsometer axis, and the plane of incidence must be perpendicular to this axis. Sample alignment refers to the process by which the sample surface is located at this pre-established intersection point such that the central ellipsometer axis lies in the plane of the sample surface. Most of the steps required for alignment of rotating-polarizer/analyzer spectroscopic ellipsometers also apply to other ellipsometer configurations. Thus, the reader is referred to standard texts for in-depth discussions.[1,2] Here we provide an overview of alignment procedures for the most common ellipsometer orientation in which the plane of incidence is horizontal.[48]

The most powerful spectroscopic ellipsometers for ex situ sample analysis include a variable angle of incidence capability. In such ellipsometers, the polarization generation and detection arms are mounted onto a goniometer base that establishes the vertical central ellipsometer axis and allows the entire polarization detection arm to be rotated relative to the polarization generation arm about the central axis in a plane parallel to the plane of incidence. Ideally the goniometer should include both θ and 2θ motion so that as the polarization detection arm rotates by 2θ from the straight-through position toward the polarization generation arm, the sample stage rotates by θ in the same direction in order to maintain sample alignment as the angle of incidence is changed. Testing the integrity of mechanical alignment for such a configuration is not straightforward; however, once its integrity is assured then sample-to-sample alignment is simple by comparison. In addition, one can calibrate the angle of incidence in this configuration once and rely on the calibration for future measurements at different angles.

In contrast, the most versatile real time spectroscopic ellipsometers employ completely independent polarization generation and detection arms that can be mounted onto vacuum flanges in order to operate at a fixed angle of incidence in measurements of a fixed sample. In this case, the axis of the polarization generation arm and the plane of the sample surface define the plane and angle of incidence. Because one need not establish a central ellipsometer axis, alignment of this instrument to a given sample is not as time consuming as is testing the mechanical integrity of the ex situ ellipsometer configuration as noted in the previous paragraph. However, in this case, sample-to-sample realignments particularly for samples of varying substrate thickness can be very difficult and time consuming, and the angle of incidence must be determined each time a new sample is mounted.

One can turn this most difficult sample-to-sample alignment situation in real time ellipsometry into the most straightforward one if the reaction chamber can be made small enough to drop into the pre-aligned ex situ ellipsometer configuration.[8,49] In this case, when the chamber is in place, the angle of incidence is set by the goniometer base to match the angle

formed by the flanges onto which the windows through the reaction chamber wall are mounted. In addition, the sample is attached to a holder inside the reaction chamber, and the holder is connected to a precision manipulator such that adjustment access is external to the chamber. Then the angle of incidence is preserved in measurements from sample to sample, even when samples of very different substrate thickness are studied. When a complete system realignment is required, the chamber is removed, and the ellipsometer arms can be set for straight-through operation with an ex situ sample mounting stage placed at the axis. Thus, this approach to real time ellipsometry can achieve the same excellent sample-to-sample alignment reproducibility that is possible in ex situ ellipsometry.

First a traditional six-step alignment process will be reviewed and special comments added specific to rotating-element ellipsometers, as well as to real time spectroscopic ellipsometers.[48] In the following, it is assumed that the polarization generation and detection arms are mounted on a variable angle of incidence goniometer and a sample table is available that can be rotated through at least 180° about the central ellipsometer axis.

(i) The first step establishes the position of the assumed vertical central axis of the ellipsometer. First a very thin (~0.1 mm) glass plate with precisely plane parallel surfaces is coated on both sides with a reflective film such as aluminum. This plate is mounted to the sample stage of the ellipsometer in such a way that both the front and back of the sample are optically accessible. The goal is to adjust the tilt and translation of the plate so that its surface is first parallel to and then coincident with the ellipsometer central axis, i.e., the axis of sample table rotation and goniometer movement. First, a laser is set up to illuminate the front center of the plate from a direction normal to the central axis, and the location of the reflected spot is noted on a distant ruled screen. Then the plate is rotated through 180° using the sample table rotation so that the beam reflects from the back surface of the plate, and the position of this new spot is noted. Parallelism of the sample surface and the central axis is then obtained through tilt adjustment of the sample that minimizes the separation of the spots on the screen. Coincidence of the sample surface and central axis is then obtained by a translation that overlaps the two spots.

(ii) The second step establishes the optical axes of the polarization generation and detection arms. In particular, the final plate position in step (i) is used as a reference surface for alignment of all the elements on the polarization generation and polarization detection arms using laser beams directed toward the plate. First, all optical elements except the collimator pinhole (and its associated optics if they are not co-linear) are removed from the polarization generation arm. The pinhole is illuminated from the front or source side using a laser mounted on the polarization generation arm, and the alignment of its associated optics is ensured. The back side of

the pinhole is covered by a white screen with a small hole in it to detect the beam retroreflected from the plate surface, the latter being oriented normal to the beam and containing the central axis. The vertical tilt of the polarization generation arm is adjusted along with the rotation angle of the sample table to achieve the autocollimation condition. Then the sample table is rotated by ~90° so that the reflective plate is parallel to the autocollimated laser beam, and the horizontal tilt of the polarization generation arm is adjusted along with the angle of the goniometer to achieve a precise grazing condition. Iteration of the two adjustments of the polarization generation arm may be required to meet the autocollimating and grazing conditions simultaneously without further adjustment of the arm.

(iii) In the third step, the entire beam path is established using irises that can be stopped down to 1 mm or less. Once the reference beam position along the polarization generation arm is established, along with the reference sample surface position, the desired angle of incidence can be set and irises can be placed along the beam path before the position of each of the key optical components on both polarization generation and detection arms, i.e., the polarizer, compensator, analyzer, and detector, such that the iris closes down symmetrically on the beam. The polarization detection arm itself will need to be tilted and translated in general in order to fulfill this condition.

(iv) In the fourth step the rotation axes of the optical elements (polarizer, compensator analyzer) are established so as to define their optical axes. In this step, which can be performed before mounting the components on the ellipsometer arms, it is important that the key optical elements along with their tubular mounts can be tilted and translated slightly with respect to their rotators for further in-position alignment. First, the optical element must be centered within the mount so that the rotation axis is coincident with the center of the optical element at one end. For the continuously rotating element (polarizer or analyzer), an externally mounted laser beam is directed onto the surface of its centered end during rotation, and the reflected beam is viewed on a distant screen. When this spot remains stationary during rotation, the axis of rotation can be used to define the optical axis of the element. Tilt adjustments are made at the opposite end of the tubular optical element mount with respect to the rotator mount to ensure that this condition is met. Such a procedure can also be applied to the other rotatable components by using the stepping motors typically employed for calibration to rotate these elements. Due to the slower speed of the rotation in this case, it may be necessary to measure the maximum beam separation through a full rotation, and adjust the tilt to minimize this separation. For non-rotatable components, such as a compensator based on the three-reflection rhomb, such rotational alignment is not possible.

(v) In the fifth step the fully mounted optical components are added to the ellipsometer arms and their optical axes aligned. Two translational

degrees and one tilt degree of freedom for the fully mounted optical components are desired in order to center the components in the beam and ensure that the reflected laser beam from each element surface superimposes on the incident beam. This ensures that the coincident optical and rotational axes of the element are parallel to the reference beam axis. For non-rotatable components this is the only alignment step needed. For rotatable optical elements further considerations are needed. If the optical elements of this latter type are completely uniform over their surface areas, then it would not be necessary to ensure that the axis of rotation coincides with the reference beam axis passing through the ellipsometer arm. However, if non-uniformity exists then errors are likely to be introduced, and coincidence of the element rotation and beam axes is desired to minimize these errors. This alignment step is somewhat difficult and deserves an additional paragraph for discussion.

Errors due to the resulting circular movement of the beam over the surface of the optical element during its rotation can be observed by comparing the shape of the waveform at the detector over successive half-mechanical rotations (successive optical cycles) of the optical element. For the continuously rotating element, the waveforms can be superimposed on an oscilloscope by triggering the scope every optical cycle. Then the element rotation is stopped, the mount translated, and the rotation restarted. The goal is to minimize the difference in waveforms of the two successive optical cycles, in which case the beam axis is most closely coincident with the rotation axis. Rather than using an oscilloscope, the ellipsometric angles (ψ, Δ) of a reflecting surface measured from successive single optical cycles can be compared. For the other rotatable optical elements, i.e., those without continuous rotation, this alignment problem is more difficult and would require a Fourier analysis of the detector output versus the element azimuthal angle P', C', or A' as this angle is varied under stepping motor control over 360° rotation. The magnitudes of the ω Fourier coefficients, which vanish under ideal conditions, provide a measure of the rotational axis alignment error and must be minimized similarly.

(vi) In the last step, routine sample alignment is achieved. Thus, the alignment plate is removed from the sample mount and the desired sample for measurement is added. Now only the tilt and translation of the sample as applied in initial step (i) should be needed to ensure an alignment. This can be done by maximizing the output at the detector when the irises are stopped down to their minimum sizes. For routine sample alignment it is also helpful to configure the detection system with a mirror that can drop into place in order to direct the beam to a subsidiary quadrant detector. The detector position can be set using the reference beam established in step (ii). Alternatively the laser setup used in step (i) can be left in position, and a target is placed on the distant screen denoting the beam location when the

alignment plate is in the correct position. The laser should hit this target when the alignment plate is replaced by the new sample.

As described above, the alignment approach for ellipsometers used in real time measurements depends on how the ellipsometer is configured. If independent pre-aligned polarization generation and detection arms are attached to reaction chamber flanges, then it is best to fix both the polarization generation beam path and the sample and then employ a gimbal mount with sufficient degrees of freedom to align the entire polarization detection beam path. This alignment must be done on a sample-by-sample basis and can be expedited by a quadrant detector as described above. The situation is different if the reaction chamber can be dropped into a pre-aligned ellipsometer system. In this case, a sample manipulator is used to orient the sample surface so that the central ellipsometer axis lies in the surface.

Special considerations are required for the entrance and exit windows of the reaction chamber. It is important to mount the windows on adjustable bellows with tilt mechanisms so that the window surfaces can be oriented normal to the incident and reflected beams. In fact if the beam were to strike a window at non-normal incidence, a small polarization state change can occur as well as a beam displacement. In addition, when a pre-aligned instrument is adapted for real time ellipsometry, it is helpful to connect irises to the chamber in front of the entrance window and behind the exit window that are accurately positioned in an initial alignment. In this way, one can ensure that the chamber alignment is maintained from measurement to measurement, and as a result that the beams cross the same two spots on the windows. Thus any necessary window corrections due to stress induced birefringence will not vary from measurement to measurement. In contrast, when independent ellipsometer arms are used, one can use an iris on the entrance window, but the position of the beam through the exit window will vary depending on the sample substrate thickness. Thus, new exit window alignment and new window corrections may be needed every time a new sample is mounted. The best solution to this problem, however, is to employ windows that are stress-free over a relatively wide window area so that no corrections at all are needed.[50]

5.3.4 Electronic Design and Components

Rotating polarizers or analyzers can be constructed by mounting the polarizing element directly within the hollow shaft of a synchronous dc motor operating at 10 to 100 Hz.[11, 20] Frequency stability can be maintained by pulses from a crystal-controlled oscillator. The pulses trigger an audio oscillator that in turn drives the motor power amplifier. Without

stabilization, phase drift can lead to errors in the data through a shift in the apparent polarizer or analyzer phase angle as obtained in calibration. The choice of the polarizer or analyzer mechanical rotation frequency ω should be made such that ω, 2ω, and 4ω are not close to any periodic fluctuations such as source ripple, ac line pickup, or mechanical vibrations. Note that the Fourier coefficients of the irradiance waveform at 2ω (along with the dc coefficient) provide information on the sample in the rotating polarizer or analyzer configuration, whereas the ω Fourier coefficients can be used to detect mechanical misalignments, and the 4ω coefficients can be used to detect polarization of the source or polarization sensitivity of the detector.[41] An optical or magnetic encoding system is mounted on the motor shaft and designed with two outputs. The so-called Z_0 output provides one pulse per mechanical rotation of the motor shaft whereas the clock output provides $2N$ pulses per rotation, where N is typically about 50. These outputs are used in timing data acquisition for waveform analysis of the detector output as will be described next.

A PMT-based detection system is typically used for rotating polarizer and analyzer instruments in ex situ spectroscopic ellipsometry from the near infrared to the ultraviolet range. There are two different approaches for discrete data collection and detector waveform analysis. Here, these approaches will be called waveform sampling and waveform integration. In waveform sampling, the time-dependent PMT photocurrent is converted to a time-dependent voltage using a transimpedance preamplifier, and this voltage is measured by a sample-and-hold input of an analog-to-digital (a-to-d) converter.[11] In waveform integration, the PMT current, described as photon counts from the PMT, charges a capacitor during a gated time interval.[51] After the time interval the capacitor is discharged and the number of counts are read out. The waveform integration approach is also used to read out the pixels of a linear photodiode array detector used for real time spectroscopic ellipsometry.[21] In both types of waveform analysis, the overall data collection sequence is enabled by the Z_0 encoder output whereas the subsequent encoder clock pulse triggers actual a-to-d conversion or count readout. Triggering by the encoder clock continues for an integral number M of optical cycles (mechanical half-cycles), whereby NM is the total number of data values collected. For this list of data values I_q ($q = 1, 2, \ldots, NM$), where I indicates a value proportional to the irradiance of the waveform at the detector, relevant for the case of sampling, the M values with $q = n + (m - 1)N$, $m = 1, \ldots, M$, are averaged so that

$$\langle I_n \rangle = \frac{1}{M} \sum_{m=1}^{M} I_{n+(m-1)N}; \quad (n = 1,2,3,\ldots,N). \tag{5.1}$$

This yields a single set of N data values as an average over M optical cycles. The averaging is performed to improve the signal-to-noise ratio. For the case of waveform integration, the same Eq. (5.1) applies, but I is conventionally replaced by the symbol S.

In real time ellipsometry, the value of M is limited by the desired time resolution when measuring processes in real time. For example, the minimum value of $M = 1$ provides a maximum possible resolution of 50 ms when operating a motor at 10 Hz. In ex situ spectroscopic measurements in which time is not an issue, the maximum value of M is chosen such that random source fluctuations limit the precision. These fluctuations cannot be eliminated with further averaging. The number of data values per optical cycle N can be as small as 3 or as large as 100. Typical values are $N = 4$ (every 45° of mechanical rotation) for waveform integration in real time spectroscopic ellipsometry with a photodiode array detector or $N = 36$ (every 5°) for waveform sampling in ex situ ellipsometry with a PMT. When performing waveform sampling, a low-pass electronic filter is typically used to reduce the noise level on individual samples. The time constant of the preamplifier should be chosen to be ~5x the duration between successive digitized points or ~1.4 ms for $N = 36$ and $\omega/2\pi = 50$ Hz.[11]

For rotating polarizer and analyzer ellipsometers the data values collected in the experiment are expected to be consistent with the following generic waveform:[52]

$$I_{\text{exp}}(t) = a_0' + a_2' \cos(2\omega t) + b_2' \sin(2\omega t), \tag{5.2a}$$

where the subscript 'exp' indicates that this is the waveform determined experimentally, a_0' is the dc Fourier coefficient to be determined, and a_2' and b_2' are the 2ω cosine and sine Fourier coefficients also to be determined. Here, time zero is defined experimentally by the first sampling point, or the time at which the first gating interval begins.

In waveform sampling, the digitization occurs at $t = 0$, $\pi/N\omega$, $2\pi/N\omega$,…$(N - 1)\pi/N\omega$, so that the data outputs should obey the following equation:

$$<I_n> = a_0' + a_2' \cos\left[\frac{2\pi(n-1)}{N}\right] + b_2' \sin\left[\frac{2\pi(n-1)}{N}\right]; \quad (n = 1,2,3,\ldots,N). \tag{5.2b}$$

The Fourier coefficients can be extracted from the $\{\langle I_n\rangle;\ n = 1,2,3,\ldots,N\}$ averaged points according to

$$a_0' = N^{-1} \sum_{n=1}^{N} \langle I_n \rangle, \tag{5.3a}$$

$$a_2' = 2N^{-1}\sum_{n=1}^{N}\langle I_n\rangle\cos\left[\frac{2\pi(n-1)}{N}\right], \tag{5.3b}$$

$$b_2' = 2N^{-1}\sum_{n=1}^{N}\langle I_n\rangle\sin\left[\frac{2\pi(n-1)}{N}\right]. \tag{5.3c}$$

In this case the higher order even Fourier coefficients can be determined; for example, 2π in Eqs. (5.3b) and (5.3c) is replaced by 4π in order to calculate a_4' and b_4', the 4ω Fourier coefficients for the experimentally determined waveform.

In waveform integration, gating spans the N equal time intervals $(0, \pi/N\omega)$, $(\pi/N\omega, 2\pi/N\omega)$,…,$[(N-1)\pi/N\omega, \pi/\omega]$ so that the data outputs should obey the following equation:

$$<S_n> = \int_{(n-1)\pi/N\omega}^{n\pi/N\omega}\{a_0' + a_2'\cos(2\omega t) + b_2'\sin(2\omega t)\}dt; \quad (n = 1,2,3,\ldots,N). \tag{5.4}$$

In the common special case of $N = 4$, the dc and 2ω Fourier coefficients can be extracted from the $\{<S_n>; n = 1,\ldots,4\}$ averaged points according to

$$a_2' = (1/2)(<S_1> - <S_2> - <S_3> + <S_4>), \tag{5.5a}$$

$$b_2' = (1/2)(<S_1> + <S_2> - <S_3> - <S_4>), \tag{5.5b}$$

$$a_0' = (1/\pi)(<S_1> + <S_2> + <S_3> + <S_4>). \tag{5.5c}$$

In this special case, only one higher order Fourier coefficient can be determined, the 4ω sine coefficient which can be written as:

$$b_4' = (1/2)(<S_1> - <S_2> + <S_3> - <S_4>). \tag{5.5d}$$

Because waveform integration is widely used in multichannel ellipsometry data acquisition, other special cases with $N = 5$ and $N = 8$ have also been worked out and appear elsewhere in the Handbook.[53]

In a scanning or ex situ spectroscopic ellipsometry experiment, the irradiance level at the detector may change significantly as the spectrum is traversed. In fact, at the shortest and longest wavelengths of the spectrum the average irradiance at the detector may be two to three orders lower than its maximum. If one operates the detection system at constant gain and employs waveform sampling, then only a fraction of the full range of the a-to-d converter is being used at the ends of the spectrum. Typical a-to-d

converters provide a full-scale 16-bit output signal resolution; however, when the maximum signal level falls below the 12-bit range, digital noise may become a measurable source of statistical fluctuations in the data. In fact, in the early development of the rotating analyzer ellipsometer, a full-scale 12-bit a-to-d converter was used, and a triangular wave signal superimposed onto the preamplifier output signal at the input of the a-to-d was found to reduce digital noise.[54] Such a procedure, designed to gain an extra half-bit resolution, is not needed with the higher resolution a-to-d converters widely available today. Steps must still be taken, however, to adjust the gain in order to maintain a high a-to-d resolution near 16-bits throughout the full spectrum.

When the detector is a PMT, then gain regulation is relatively straightforward. In this case, the PMT cathode is powered by a high voltage supply with the capability of adjusting the voltage under external control. Thus, when the spectrum is stepped from one point to the next, the average signal output is compared to the desired average that effectively utilizes the range of the a-to-d converter. Then, the voltage at the PMT cathode is raised or lowered to attain the desired average. In addition, a limiting PMT cathode voltage level is specified in order to protect the electronic system from failure in the event of inadvertent beam interruption, e.g., due to lamp burnout during voltage regulation. Analog circuitry has also been described to achieve gain regulation in ex situ spectroscopic ellipsometry,[54] however, such circuitry should not be employed without incorporating some failure protection such as Zener diode limiters.

An important consideration in the detection system electronics of photometric ellipsometers is nonlinearity. In theory, nonlinearity corrections deduced from higher order Fourier coefficients can be applied to the 2ω Fourier coefficients; however, these corrections would not be universally applicable and would need to be determined on a measurement-by-measurement basis. Furthermore, the higher order coefficients may also include contributions from errors other than nonlinearity such as source polarization and polarization sensitive detection. This would make it difficult to extract the nonlinearity corrections with the required accuracy. Thus, eliminating nonlinearity at its source is clearly the best approach. The first consideration in this effort is to maintain the anode current at a low enough level to avoid PMT saturation. Although the operational characteristics depend on the type of PMT, typical anode currents i_a are in the range of 10 to 100 μA. This implies that the preamplifier must be designed for current-to-voltage conversion of $10^{-6} - 10^{-5}$ A/V when interfacing to an a-to-d converter with maximum input voltage of 10 V.

Circuitry that corrects for any residual nonlinearity at the proper anode current levels is based on the following equation for the overall gain G of the PMT:

$$G = \delta^s[1 + (g^s/V)(\Delta V + i_a R_b')], \tag{5.6}$$

where $R_b' \equiv R_b\{1 - [s(\delta - 1)]^{-1}\}$ [Ref. 54].

Equation (5.6) is expressed to first order in the variation ΔV of the PMT cathode voltage $V = V_a + \Delta V$ about an average value V_a. In this equation, δ is the gain of each stage (assumed to be identical for all stages), s is the number of gain stages, R_b is the value of each of the resistors in the dynode chain (assumed to be identical for all resistors), i_a is the anode current, and g is a constant characteristic of the PMT. By applying a feedback voltage to the cathode given by $\Delta V = -i_a R_b'$, the nonlinearity inherent in PMT detection caused by the non-zero dynode resistances is eliminated to first order and approximately independent of V. The best test of the operation of such circuitry is to measure the 4ω Fourier coefficients before and after adjustment of circuit element values, in order to minimize these higher order coefficients. In this way, linearity to 5×10^{-4} over three orders of magnitude of incident irradiance has been obtained for a rotating analyzer ellipsometer with data collection by waveform sampling. Similar approaches for gain regulation and linearity can also be developed for ellipsometers that use waveform integration.

Another problem associated with the data acquisition circuitry for waveform sampling involves the suppression of the 2ω Fourier component and its phase shift due to the filtering at the input to the preamplifier.[11] In considering this effect, Eq. (5.2a) provides a useful starting point. This equation can be re-expressed as:

$$I_{\text{exp}}(t) = I_0'\{1 + \alpha'\cos(2\omega t) + \beta'\sin(2\omega t)\}, \tag{5.7}$$

where $I_0' = a_0'$, $\alpha' = a_2'/a_0'$, and $\beta' = b_2'/a_0'$, the latter two representing measured dc normalized coefficients. In fact, the theoretically correct irradiance waveform at the detector, as will be derived in Sec. 5.4, exhibits the form:

$$I_{\text{th}}(t) = I_0\{1 + \alpha\cos[2A'(t)] + \beta\sin[2A'(t)]\}, \tag{5.8}$$

for the rotating analyzer configuration, or

$$I_{\text{th}}(t) = I_0\{1 + \alpha\cos[2P'(t)] + \beta\sin[2P'(t)]\}, \tag{5.9}$$

for the rotating polarizer configuration. In these equations, unprimed α and β designate the normalized 2ω Fourier coefficients that can be predicted from a Mueller matrix and Stokes vector analysis of the instrument as will be described in Section 5.4. In Eqs. (5.8) and (5.9), A' and P' are the true azimuthal angles of the analyzer and polarizer that increase linearly with time as the optical elements rotate (assuming counterclockwise rotation, looking opposite to the beam direction). These angles are measured with respect to the plane of incidence in a counterclockwise positive sense and can be written as $A'(t) = \omega t - A_S$ and $P'(t) = \omega t - P_S$ in order to make connection to the experimental result of Eq. (5.7). In these expressions, $t = 0$ is the time at which the first data sample is collected (or the time at which waveform integration is initiated) in the experiment and $-A_S$ and $-P_S$ represent the analyzer and polarizer azimuthal angles at this time.

Even with a linear relationship between the incident irradiance on the PMT and the resulting anode current, one still must be concerned with the effects of signal filtering and preamplification in the current-to-digitized voltage conversion.[11] These effects can be simulated using the following expression for the theoretical electronic signal deduced from a modification of Eqs. (5.8) and (5.9):

$$\begin{aligned} I_{\mathrm{el}}(t) = I_0[1 &+ \eta^{-1}\alpha \cos\{2[\omega t - X_S - (\phi/2)]\} \\ &+ \eta^{-1}\beta \sin\{2[\omega t - X_S - (\phi/2)]\}], \end{aligned} \tag{5.10}$$

where X_S represents either A_S for a rotating analyzer instrument or P_S for a rotating polarizer instrument. The key modifications of Eqs. (5.8) and (5.9) include the factor η^{-1} that describes the suppression of the 2ω Fourier component relative to the dc component and the phase term ϕ that describes the phase shift of the 2ω Fourier component, both as a result of the signal processing circuitry (i.e., the input filter and preamplifier). Equating Eq. (5.10) that describes the theoretical waveform, and Eq. (5.7) that describes the experimental waveform, the theoretical normalized Fourier coefficients can be expressed in terms of the experimentally determined ones through the following equations:

$$\alpha = \eta\,[\alpha' \cos\{2[X_S + (\phi/2)]\} + \beta' \sin\{2[X_S + (\phi/2)]\}], \tag{5.11a}$$

$$\beta = \eta\,[-\alpha' \sin\{2[X_S + (\phi/2)]\} + \beta' \cos\{2[X_S + (\phi/2)]\}]. \tag{5.11b}$$

One goal of instrument calibration is to determine the three parameters η, X_S, and ϕ as described in Section 5.6. From Eqs. (5.11), however, it is clear that one cannot separate X_S, the phase shift associated with the polarizer or analyzer angle due to the positioning of the Z_0 pulse (and any electronic delays in sampling) from ϕ, the phase shift due to the filtering

and preamplification. Thus, in future discussions, these phase shift terms will continue to be grouped as in Eqs. (5.11) and no attempt will be made to separate them. Finally, it should be noted that if no filtering is used as in the case of waveform integration and the circuitry is ideal, then $\eta = 1$ and $\phi = 0°$ in Eqs. (5.11). In this case, one need only determine X_S in instrument calibration.

A final consideration in the electronic circuit performance involves accounting for ambient light, PMT dark current, and zero offsets in the preamplifier and a-to-d converter. Ideally, one would like to set the electronic circuitry so that when the light on the detector is blocked, the signal level vanishes. In spectroscopic applications in which the gain of the system is changed continuously as the spectrum is traversed, however, the ambient light and dark current contributions to the dc signal level change, and this makes it impossible to zero out these contributions through electronic adjustments. As the extremes of the spectrum are approached where the Xe source output is very weak and the PMT cathode voltage is very high, the ambient light or dark current can lead to significant errors. The overall problem is best eliminated by incorporating a computer controlled shutter such that digitized or integrated data averaged over a number of optical cycles of the rotating element are collected with the shutter open, and then with the shutter closed.[11] The results for $<I_n>$ or $<S_n>$ ($n = 1, \ldots, N$) obtained with the shutter open and closed are subtracted, and these corrected average data values spanning an optical cycle are used to calculate α' and β' according to Eqs. (5.3). In this way, even if the ambient light is partially polarized, the correction procedure is still effective. Finally, without loss of accuracy, the number of optical cycles over which data are collected with the shutter closed can be smaller by a factor of about ten than the corresponding number with the shutter open. This leads to an acquisition time savings for the full data cycle.

The discussion of this section concludes with a description of the sequence of events that would make up a typical single data cycle in the collection of ellipsometric spectra with a scanning ellipsometer.

(i) *Set monochromator.* The motor driving the monochromator is energized and steps the wavelength to the desired value. The shutter is already open and the PMT cathode voltage is held while moving the stepping motor.

(ii) *Set optical elements*. Any optical element changes are made including order-sorting filter insertion into the beam or compensator adjustment based on the selected value of the wavelength.

(iii) *Regulate and hold PMT.* The experimental data are sampled over a number of optical cycles to determine the average

signal level and an adjustment is made (if needed) to the PMT cathode voltage to bring the signal level into the desired range. The PMT voltage is held at this value to prepare for measurement.

(iv) *Collect light cycle data*. The data over a specified number M of optical cycles are obtained by digitization or integration. These M sets of N data points are averaged to yield N points over a single optical cycle.

(v) *Collect dark cycle data*. The shutter is closed and data over a specified number of optical cycles, typically $\sim M/10$, are obtained by digitization or integration. The data points are averaged to yield N points over a single optical cycle.

(vi) *Compute and store normalized Fourier coefficients*. The averaged data collected in the dark cycle are subtracted from those collected in the light cycle, and the normalized Fourier coefficients α' and β' are determined from the resulting N data points. If calibration information is available, then the theoretically significant coefficients α and β can be determined using Eqs. (5.11). It is useful to save the experimental coefficients as well in case it is discovered later that the applied calibration values are in error.

After this last step, the measurement sequence is returned to step (i). Alternatively, if the properties of the sample change rapidly with wavelength, one may want to repeat the measurement at the same wavelength with new settings of the optical elements, e.g., compensator or fixed polarizer/analyzer, to achieve results with higher accuracy and precision. In the next section, the basic procedures of how to determine sample information from the theoretical coefficients α and β will be discussed.

5.4 Data Reduction for the Rotating Polarizer and Analyzer Ellipsometers

5.4.1 Ideal PXSA$_r$ Configuration

Equations (5.11) describe how one determines the theoretical 2ω Fourier coefficients α and β of Eq. (5.8) or (5.9) from the experimental coefficients of Eq. (5.7). In this part, an outline will be provided describing how one determines sample information ψ and Δ from the theoretical Fourier coefficients for the rotating analyzer ellipsometer configuration

designated PXSA$_r$, where the subscript 'r' identifies the rotating element, in this case the analyzer A. This description includes the possibility of incorporating any polarization modifying optical element (X) between the polarizer P and the sample S. The approach will be based on the following vector equation:

$$I_{\text{th}}\mathbf{s}_D = \Re(-A')\mathbf{M}_A\Re(A')\mathbf{M}_S\mathbf{M}'\Re(-P')\mathbf{M}_P\Re(P')(I_S\mathbf{s}_S), \qquad (5.12)$$

which describes the normalized Stokes vector $\mathbf{s}_D$ of the light beam at the detector as a product of several Mueller matrices times the normalized Stokes vector $\mathbf{s}_S$ of the light beam emitted by the source.[1] In this equation, I_S and I_{th} are the irradiances (scalar quantities) of the light beams exiting the source and entering the detector, respectively, where I_{th} has been defined earlier in Eq. (5.8). In Eq. (5.12), $\mathbf{M}_P$ and $\mathbf{M}_A$ are the Mueller matrices for the assumed ideal polarizer and analyzer, given by:

$$\mathbf{M}_j = (1/2)T_{jt}\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad (5.13a)$$

where the subscript 'j' is either 'P' or 'A', and T_{jt} is the transmittance of the polarizer or analyzer for light polarized along the transmission axis (subscript 't'). $\mathbf{M}_S$ is the Mueller matrix of the sample given by:

$$\mathbf{M}_S = R_u\begin{bmatrix} 1 & -\cos 2\psi & 0 & 0 \\ -\cos 2\psi & 1 & 0 & 0 \\ 0 & 0 & \sin 2\psi\cos\Delta & \sin 2\psi\sin\Delta \\ 0 & 0 & -\sin 2\psi\sin\Delta & \sin 2\psi\cos\Delta \end{bmatrix}, \qquad (5.13b)$$

where $R_u = \dfrac{R_p + R_s}{2} = \dfrac{|r_p|^2 + |r_s|^2}{2}$

is the reflectance of the sample for unpolarized light. The ellipsometric angles are defined as usual by $\psi = \tan^{-1}(|r_p|/|r_s|)$ and $\Delta = \delta_p - \delta_s$, where $r_p = |r_p|\exp(i\delta_p)$ and $r_s = |r_s|\exp(i\delta_s)$ are the complex amplitude reflection coefficients for linearly polarized light waves with electric fields along the p and s directions. The matrices of Eq. (5.13) are expressed in the coordinate system of the elements themselves. The coordinate system for the polarizer and analyzer is right-handed and is defined by its transmission axis (t, corresponding to x) and extinction axis (e, corresponding to y), whereas the z direction is along the propagation vector of the light beam. The coordinate system of the sample includes the

p direction, defined by the intersection of the surface and the plane of incidence (corresponding to x), and the s direction perpendicular to the p direction in the plane of the surface (corresponding to y), whereas the z axis is directed inward into the sample, perpendicular to its surface.

The matrix $\Re(\theta)$ appearing four times in Eq. (5.12) is a rotational transformation that converts the Stokes vector $\mathbf{S}$ or Mueller matrix $\mathbf{M}$ from an 'unprimed' to a 'primed' coordinate system through a rotation angle θ according to $\mathbf{S}' = \Re(\theta)\mathbf{S}$ and $\mathbf{M}' = \Re(-\theta)\mathbf{M}\Re(\theta)$. Thus, $\theta = P'$ and A' in Eq. (5.12) designate the angles from the sample p direction to the polarizer and analyzer transmission axes, respectively, measured in a counterclockwise positive direction looking toward the source. The general form of the rotation matrix is:

$$\Re(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\theta & \sin 2\theta & 0 \\ 0 & -\sin 2\theta & \cos 2\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{5.14}$$

Finally $\mathbf{M}'$ in Eq. (5.12) is the Mueller matrix of the general optical element X. For a compensator in the reference p-s coordinate system, this can be written as $\mathbf{M}' = T_u\Re(-C')\mathbf{m}_C\Re(C')$, where $T_u = (T_F + T_S)/2$ is the transmittance of the compensator for unpolarized incident light and C' is the orientation of the fast axis of the compensator with respect to the p direction. The matrix $\mathbf{m}_C$ is the normalized Mueller matrix for the compensator within its own frame of reference, defined by the fast axis F (corresponding to the x direction) and the slow axis S (corresponding to the y direction), whereas the z direction is in the direction of beam travel.

In general, a compensator may impose a relative F-to-S amplitude change in addition to a phase shift on the two linearly polarized orthogonal field components. In this case, $\mathbf{m}_C$ is given by

$$\mathbf{m}_C = \begin{bmatrix} 1 & -\cos 2\psi_C & 0 & 0 \\ -\cos 2\psi_C & 1 & 0 & 0 \\ 0 & 0 & \sin 2\psi_C \cos\delta_C & \sin 2\psi_C \sin\delta_C \\ 0 & 0 & -\sin 2\psi_C \sin\delta_C & \sin 2\psi_C \cos\delta_C \end{bmatrix}, \tag{5.15}$$

in analogy with the Mueller matrix of the sample. In Eq. (5.15), $\psi_C = \tan^{-1}(|t_F|/|t_S|)$ and $\delta_C = \delta_F - \delta_S$ where $t_F = |t_F| \exp(i\delta_F)$ and $t_S = |t_S|\exp(i\delta_S)$ define the complex amplitude transmission coefficients for linearly polarized light waves with electric fields along the F and S axes. In fact, for a single plate compensator, $\delta_C = 2\pi(n_S - n_F)d/\lambda$, where n_F and n_S are the indices of refraction for electric fields along the fast and slow axes. Because the phase velocities and indices of refraction are related by

$v_F = c/n_F$ and $v_S = c/n_S$, where c is the speed of light in vacuum, and because $v_F > v_S$, then $n_S > n_F$ and $\delta_C > 0$.

To determine (ψ, Δ) on the basis of Eq. (5.12), the calculation begins at the detector and progresses backwards through the ellipsometer. The Stokes vector of the beam reflected from the sample can be written $\mathbf{S}_r = I_{0r}\mathbf{s}_r = I_{0r}[1 \quad s_{2r} \quad s_{3r} \quad s_{4r}]^{\mathrm{T}}$ in the p-s reference frame, where the subscript 'r' indicates reflection. Here I_{0r} is the irradiance in the beam, and s_{2r}, s_{3r}, and s_{4r} are the three normalized Stokes vector components. The superscript 'T' indicates a transpose since the Stokes vectors are column vectors; the row vector notation is used for compactness. With this Stokes vector incident on the rotating analyzer, the irradiance at the detector follows the form:

$$I_{\mathrm{th}}(t) = (1/2)T_{At}I_{0r}\{1 + s_{2r}\cos 2A' + s_{3r}\sin 2A'\}. \tag{5.16}$$

A comparison with Eq. (5.8) shows that:

$$\alpha = s_{2r}, \tag{5.17a}$$

$$\beta = s_{3r}, \tag{5.17b}$$

$$I_0 = (1/2)T_{At}I_{0r}. \tag{5.17c}$$

Thus, by measuring the 2ω Fourier coefficients, one is actually establishing partial information on the Stokes vector of the light beam in the p-s frame just before the beam enters the rotating analyzer. This is a very simple, but important result that requires further digression.

It is helpful to recall the general expression for the Stokes vector of a partially polarized light beam, $\mathbf{S} = I[1 \quad s_2 \quad s_3 \quad s_4]^{\mathrm{T}} = I[1 \quad p\cos 2Q\cos 2\chi \quad p\sin 2Q\cos 2\chi \quad p\sin 2\chi]^{\mathrm{T}}$, in terms of the ellipse of polarization.[1] Here Q $(-90° < Q \leq 90°)$ is the azimuth angle of the semimajor axis of the polarization ellipse and χ $(-45° \leq \chi \leq 45°)$ is the ellipticity angle of the ellipse given by $\chi = \tan^{-1}e$ $(-1 \leq e \leq 1)$, where e is the ratio of the semiminor axis to the semimajor axis of the ellipse. The angle Q is measured from the plane of incidence in a counterclockwise-positive sense looking opposite to the beam direction. Finally, I and p are the irradiance and the degree of polarization in the beam. By measuring only the middle two Stokes vector components s_2 and s_3, it is not possible to extract (Q, χ, p) together. If partially polarized light of a certain nature is *assumed*, then the data from a rotating analyzer ellipsometer can be analyzed in such terms; however, the rotating analyzer polarization state detector cannot be used to *detect* partially polarized light. For example, it is clear that with the rotating analyzer one cannot distinguish between natural light with $p = 0$ and circularly polarized light with

$\chi = 45°$. In addition, because only $\cos 2\chi$ can be determined with the rotating analyzer ellipsometer, one cannot determine the sign of χ. Hence, one cannot determine the handedness of elliptically polarized light entering the rotating analyzer.

Returning to the problem at hand, the Stokes vector components at the input of the rotating analyzer can be expressed in terms of the Stokes vector at the output of unspecified component X. The latter Stokes vector is written: $\mathbf{S}_i = I_{0i}\ [1 \quad s_{2i} \quad s_{3i} \quad s_{4i}]^{\mathrm{T}}$ where the subscript 'i' indicates the beam incident on the sample. The expressions are:

$$s_{2r} = \frac{s_{2i} - \cos 2\psi}{1 - s_{2i}\cos 2\psi}, \tag{5.18a}$$

$$s_{3r} = \frac{\sin 2\psi(s_{3i}\cos\Delta + s_{4i}\sin\Delta)}{1 - s_{2i}\cos 2\psi}. \tag{5.18b}$$

Rearranging these equations in a convenient form for deduction of the sample parameters yields:

$$\tan\psi = \left|\sqrt{\frac{(1 + s_{2r})(1 - s_{2i})}{(1 - s_{2r})(1 + s_{2i})}}\right|, \tag{5.19a}$$

$$s_{3i}\cos\Delta + s_{4i}\sin\Delta = s_{3r}\left|\sqrt{\frac{1-s_{2i}^2}{1-s_{2r}^2}}\right|. \tag{5.19b}$$

In these two equations, it is assumed that the positive roots are taken since $0° \le \psi \le 90°$ and since the sign of the left-hand side of Eq. (5.19b) is carried by s_{3r} or by β, as can be seen by inspection of Eq. (5.18b). In fact, Eq. (5.19b) can be readily solved as a quadratic in $\cos\Delta$ with the substitution $\sin\Delta = \pm\sqrt{1 - \cos^2\Delta}$, yielding:

$$\cos\Delta = \frac{s_{3r}}{\left|\sqrt{(1-s_{2r}^2)(1-s_{2i}^2)}\right|}\left[s_{3i} \pm s_{4i}\left|\sqrt{\frac{1 - s_{2r}^2 - s_{3r}^2}{s_{3r}^2}}\right|\right]. \tag{5.19c}$$

The two sign choices here arise from the fact that the sign of Δ cannot be determined and, as a result, the sign of $\sin\Delta$ in Eq. (5.19b) will be unknown. This sign uncertainty must be resolved based on some knowledge of the sample characteristics; however, if s_{4i} vanishes, implying that the light incident on the sample is linearly polarized, then this ambiguity is eliminated.

Finally, to complete the analysis, one first assumes that the light from the source is unpolarized $\mathbf{s}_S = (1 \ 0 \ 0 \ 0)^{\mathrm{T}}$. Then one can express the

normalized Stokes vector of the beam incident on the sample as:

$$s_{ji} = \frac{m'_{j1} + m'_{j2}\cos 2P' + m'_{j3}\sin 2P'}{1 + m'_{12}\cos 2P' + m'_{13}\sin 2P'}; \quad (j = 2,3,4). \tag{5.20}$$

Here m'_{jk} ($k = 1,2,3$) represents the (j, k) element of the matrix $\mathbf{m'}$ given by $\mathbf{M'} = T_u\mathbf{m'}$ where $\mathbf{M'}$ is defined through Eq. (5.12). The irradiance incident on the sample is given by $I_{0i} = (1/2)I_S T_{Pt} T_u D_1$, where $D_1 = 1 + m'_{12}\cos 2P' + m'_{13}\sin 2P'$. Thus, if one knows the instrument and component characteristics η, $[A_S + (\phi/2)]$, and P', as well as the Mueller matrix elements of the component X in the p-s reference frame, and if one measures the Fourier coefficients (α', β'), then one can apply Eqs. (5.11) to extract (α, β), and then Eqs. (5.17), (5.19), and (5.20) to extract (ψ, Δ). Thus, the approach above is amenable to analysis by computer since the determination of the elements of $\mathbf{m'}$ inevitably involves three or more matrix multiplications. Next a series of cases will be considered for the matrix $\mathbf{m'}$, leading to expressions that relate the theoretical Fourier coefficients (α, β) to (ψ, Δ). These examples serve as illustrations of the wide applicability of the approach presented here.

First, consider the situation in which no optical element is in location X. In this case $m'_{jk} = \delta_{jk}$, the Kronecker delta. Then Eq. (5.20) shows that $I_{0i} = (1/2)I_S T_{Pt}$, $s_{2i} = \cos 2P'$, $s_{3i} = \sin 2P'$, and $s_{4i} = 0$. This leads to two of the most important expressions for the rotating analyzer ellipsometer in this chapter:

$$\tan\psi = \left|\sqrt{\frac{1+\alpha}{1-\alpha}}\right| |\tan P'|, \tag{5.21a}$$

$$\cos\Delta = \beta\left|\sqrt{\frac{1}{1-\alpha^2}}\right|. \tag{5.21b}$$

Here and below, it is assumed that $0° \le P' \le 90°$ so that $\sin 2P' \ge 0$, and the sign of $\cos\Delta$ is carried by β. Equations (5.21) are important because they relate the sample parameters to the Fourier coefficients for the simplest instrument design without a compensator.[11]

A second very common case with similarly simple data reduction equations occurs when an ideal compensator is aligned with its fast axis F along the p direction so that $C' = 0°$. Then $\mathbf{m'} = \mathbf{m}_C$ of Eq. (5.15) but with $\psi_C = 45°$. As a result, $I_{0i} = (1/2)I_S T_{Pt} T_u$, $s_{2i} = \cos 2P'$, $s_{3i} = \sin 2P' \cos\delta_C$, and $s_{4i} = -\sin 2P' \sin\delta_C$. For this situation, the result for ψ is the same as that of Eq. (5.21a), but the result for Δ of Eq. (5.21b) must be replaced by:

$$\cos(\Delta + \delta_C) = \beta\left|\sqrt{\frac{1}{1-\alpha^2}}\right|, \tag{5.22}$$

so that Δ is replaced by $\Delta + \delta_C$. This simply means that the phase shift of the ideal compensator and sample add when F is along p [10–12].

A third case can be derived in which $C' = 0°$ and the characteristics of the compensator (ψ_C, δ_C) in Eq. (5.15) are arbitrary. Then $I_{0i} = (1/2)I_S T_{Pt}(|t_F|^2 \cos^2 P' + |t_S|^2 \sin^2 P')$, $s_{2i} = (\tan^2\psi_C - \tan^2 P')/D_{2i}$, $s_{3i} = (2\tan P' \tan\psi_C \cos\delta_C)/D_{2i}$, and $s_{4i} = -(2\tan P' \tan\psi_C \sin\delta_C)/D_{2i}$, where $D_{2i} = \tan^2\psi_C + \tan^2 P'$. For this situation, the result for Δ in Eq. (5.22) continues to hold, but the result for ψ in Eq. (5.21a) is replaced by:

$$\tan\psi = \left|\sqrt{\frac{1+\alpha}{1-\alpha}}\right| \frac{|\tan P'|}{\tan\psi_C}. \tag{5.23}$$

Thus, $\tan\psi$ is replaced by $(\tan\psi \tan\psi_C)$, which simply means that the amplitude ratios of the compensator and sample multiply when F is along p.

The final case of an ideal compensator with $\psi_C = 45°$ at an arbitrary angle C' is more difficult to handle. In this case, only the Stokes vector of the beam incident on the sample will be presented. First, the output irradiance is given by $I_{0i} = (1/2)I_S T_{Pt} T_u$ and the normalized components are:

$$s_{2i} = \cos^2(\delta_C/2)\cos 2P' + \sin^2(\delta_C/2)\cos[2\,(2C' - P')], \tag{5.24a}$$

$$s_{3i} = \cos^2(\delta_C/2)\sin 2P' + \sin^2(\delta_C/2)\sin[2\,(2C' - P')], \tag{5.24b}$$

$$s_{4i} = \sin\delta_C \sin[2(C' - P')]. \tag{5.24c}$$

These results along with Eqs. (5.17a) and (5.17b) can be substituted into Eqs. (5.19a) and (5.19c). There are two checks on Eqs. (5.24), namely, that the sums of the squares of the three components should be unity, and that the equations must reduce to $s_{2i} = \cos 2P'$, $s_{3i} = \sin 2P' \cos\delta_C$, and $s_{4i} = -\sin 2P' \sin\delta_C$ when $C' = 0°$.

So far, the discussion has focused on the application of Eqs. (5.16) through (5.20) to extract sample information given information on the optical and electronic component settings and characteristics, including P', C', ψ_C, δ_C, η, and $[A_S + (\phi/2)]$. Another application is to use a standard sample of known (ψ, Δ) values and apply Eqs. (5.22) and (5.23) to extract compensator information, i.e., the (ψ_C, δ_C) values, when the compensator is set at $C' = 0°$. This can be dangerous owing to the sensitivity of the (ψ, Δ) values to environmental conditions such as humidity and temperature, however, an improved standard sample uses a 2θ isoceles prism, where θ is the angle of incidence, fabricated with stress-free vitreous silica and coated on the unique face with a metal film. For this sample the incident and reflected light beams enter and exit at normal incidence through the prism and the internal reflection ensures that the ellipsometric angles are not affected by ambient conditions.[55]

5.4.2 Errors in the PXSA$_r$ Configuration

A useful approach for reactor entrance window characterization also uses a standard sample. In general, one seeks to eliminate window birefringence through proper design. Oftentimes one is left with small errors that require measurement and correction. The analysis presented here employs a Mueller matrix for the entrance window based on the retardation plate model,[56] namely:

$$\mathbf{m}'_{W1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -\delta_1 \sin 2W_1 \\ 0 & 0 & 1 & \delta_1 \cos 2W_1 \\ 0 & \delta_1 \sin 2W_1 & -\delta_1 \cos 2W_1 & 1 \end{bmatrix}. \tag{5.25}$$

In fact, this derivation is based on the equation $\mathbf{M}' = T_u \mathbf{m}'_{W1} = T_u \Re(-W_1)\mathbf{m}_{W1}\Re(W_1)$, where T_u is the transmittance of the window for unpolarized light and W_1 is the azimuthal angle of the fast axis of the window, i.e., the principal axis of birefringence along which n is the lower value, i.e., $n = n_F$. In Eq. (5.25), $\delta_1 \ll 1$ (in radians) is the retardance associated with the assumed very weak stress-induced birefringence of the window. Thus, with the window as component X, one finds $I_{0i} = (1/2)I_S T_{Pt} T_u$, $s_{2i} = \cos 2P'$, $s_{3i} = \sin 2P'$, and $s_{4i} = \delta_1 \sin[2(W_1 - P')]$. The results can be substituted into Eqs. (5.19) in order to determine (ψ, Δ) as long as the window parameters (δ_1, W_1) are known. Alternatively, window characteristics can be determined using a standard sample of known (ψ, Δ) values. The following expression can be derived from Eq. (5.19b):

$$\frac{\delta_1 \sin[2(W_1 - P')]}{\sin 2P'} = \frac{\beta(1 - \alpha^2)^{-1/2} - \cos\Delta}{\sin\Delta}, \tag{5.26}$$

where (α, β) are deduced from the measurement of (α', β') with the entrance window in place, and Δ characterizes the sample surface obtained without the window. Equation (5.19a) is unchanged with the addition of the window, implying that the determination of α from (α', β') and the computation of ψ are unaffected by its presence in the beam path. Conversely, window characteristics cannot be measured through consideration of Eq. (5.19a). Returning to Eq. (5.26), this expression demonstrates the expected result that if the window is aligned so that the angle between the fast (or slow) axis and the polarizer transmission axis vanishes, then the entrance window will not generate errors in the (α', β') data. This equation also shows that in order to characterize the window, at least two measurements at different polarizer or window orientations are needed. For example if the window is rotated counterclockwise in its

coordinate frame to $W_{10} + 45°$ then the numerator of the left side of Eq. (5.26) becomes $\delta_1 \cos[2(W_{10} - P')]$. Combining measurements in this orientation with measurements at W_{10} can provide (δ_1, W_{10}).

The derivation leading to Eqs. (5.20) is based on the assumption of a perfect polarizer with Mueller matrix given by Eq. (5.13a). Two different types of errors can occur in polarizers that lead to modified Mueller matrices.[1]

First, the polarizer can partially polarize the light beam, and thus can exhibit a non-zero extinction ratio $\varepsilon_P \equiv T_{Pe}/T_{Pt}$, where T_{Pe} and T_{Pt} are the transmittances of the polarizer for linearly polarized light along the extinction and transmission axes, respectively. This imperfect polarizer leads to a Mueller matrix of the form:

$$\mathbf{M}_j = T_{Pu}\begin{bmatrix} 1 & \dfrac{1-\varepsilon_P}{1+\varepsilon_P} & 0 & 0 \\ \dfrac{1-\varepsilon_P}{1+\varepsilon_P} & 1 & 0 & 0 \\ 0 & 0 & \dfrac{2\sqrt{\varepsilon_P}}{1+\varepsilon_P} & 0 \\ 0 & 0 & 0 & \dfrac{2\sqrt{\varepsilon_P}}{1+\varepsilon_P} \end{bmatrix}, \qquad (5.27)$$

where $T_{Pu} = (T_{Pt} + T_{Pe})/2$. With this polarizer, the Stokes vector at the input of the sample becomes

$$s_{ji} = \frac{m'_{j1} + pm'_{j2}\cos 2P' + pm'_{j3}\sin 2P'}{1 + pm'_{12}\cos 2P' + pm'_{13}\sin 2P'}; \quad (j = 2, 3, 4), \qquad (5.28)$$

where $p = (1 - \varepsilon_P)/(1 + \varepsilon_P)$ is the degree of polarization of the light beam at the exit of the partial polarizer. The irradiance incident on the sample is given by $I_{0i} = I_S T_{Pu} T_u D_1$, where $D_1 = 1 + pm'_{12}\cos 2P' + pm'_{13}\sin 2P'$.

Second, some polarizers generate light with a slight ellipticity from incident unpolarized light. This is the case for Rochon polarizers fabricated from quartz, and the resulting Mueller matrix, expressed to second order in the ellipticity parameter γ_P is given by:

$$M_j = \frac{T_{Pt}}{2}\begin{bmatrix} 1 & 1-2\gamma_P^2 & 0 & 2\gamma_P \\ 1-2\gamma_P^2 & 1-4\gamma_P^2 & 0 & 2\gamma_P \\ 0 & 0 & 0 & 0 \\ 2\gamma_P & 2\gamma_P & 0 & 4\gamma_P^2 \end{bmatrix}. \qquad (5.29)$$

In fact $+\gamma_P$ is the ellipticity that results when unpolarized light enters the polarizer. Thus with the Mueller matrix of Eq. (5.29), $\gamma_P > 0$ represents a

right elliptical polarizer, e.g., right-handed crystal quartz for a Rochon polarizer. For a general elliptical polarizer, the normalized components of the Stokes vector incident on the sample are:

$$s_{ji} = \frac{m'_{j1} + (1 - 2\gamma_P^2)(m'_{j2}\cos 2P' + m'_{j3}\sin 2P') + 2\gamma_P m'_{j4}}{1 + (1 - 2\gamma_P^2)(m'_{12}\cos 2P' + m'_{13}\sin 2P') + 2\gamma_P m'_{14}}; \quad (j = 2, 3, 4), \tag{5.30}$$

and the irradiance incident on the sample is given by $I_{0i} = I_S T_{Pt} T_{iu} D_1$, where $D_1 = 1 + (1 - 2\gamma_P^2)(m'_{12}\cos 2P' + m'_{13}\sin 2P') + 2\gamma_P m'_{14}$.

By incorporating a general Mueller matrix for the element X at the position of the compensator, it is straightforward to incorporate compensator errors into the analysis. For example, some compensators can act as elliptic retarders, meaning that the polarization states that pass unchanged along the F and S axes are slightly elliptical, rather than linear.[1] In this case, it is these orthogonal elliptical modes that experience a relative retardance that changes the polarization state of the transmitted light beam. The Mueller matrix for an elliptic retarder without dichroic effects can be written in its F-S frame of reference as

$$\mathbf{m}_C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 - \xi_e^2 & 2\gamma_C \sin\delta_C & \gamma_e \\ 0 & -2\gamma_C \sin\delta_C & \cos\delta_C & (1 - 2\gamma_C^2)\sin\delta_C \\ 0 & \gamma_e & -(1 - 2\gamma_C^2)\sin\delta_C & \cos\delta_C + \xi_e^2 \end{bmatrix}, \tag{5.31}$$

to second order in γ_C, the ellipticity of the elliptical polarization mode associated with fast axis propagation. In addition, the following definitions have been made: $\xi_e^2 \equiv 4\gamma_C^2(1 - \cos\delta_C)$ and $\gamma_e \equiv 2\gamma_C(1 - \cos\delta_C)$. In fact, Eq. (5.31) can be applied specifically to the case of a properly aligned quartz biplate compensator where $\gamma_C > 0$ represents a right-handed characteristic for the equivalent single plate. Now by taking the matrix product $\mathbf{M}' = T_u\,\mathbf{m}' = T_u\,\Re(-C')\,\mathbf{m}_C\,\Re(C')$ one can deduce the following nonzero matrix elements of $\mathbf{m}'$ as:

$$m'_{11} = 1, \tag{5.32a}$$

$$m'_{22} = (1 - 2\gamma_C\gamma_e)\cos^2 2C' + \cos\delta_C \sin^2 2C', \tag{5.32b}$$

$$m'_{23} = \{[\sin^2(\delta_C/2)] - \gamma_C\gamma_e\}\sin 4C' + 2\gamma_C \sin\delta_C, \tag{5.32c}$$

$$m'_{24} = -(1 - 2\gamma_C^2)\sin\delta_C \sin 2C' + \gamma_e \cos 2C', \tag{5.32d}$$

$$m'_{32} = \{[\sin^2(\delta_C/2)] - \gamma_C\gamma_e\}\sin 4C' - 2\gamma_C \sin\delta_C, \tag{5.32e}$$

$$m'_{33} = (1 - 2\gamma_C\gamma_e)\sin^2 2C' + \cos\delta_C \cos^2 2C', \tag{5.32f}$$

$$m'_{34} = (1 - 2\gamma_C^2)\sin\delta_C \cos 2C' + \gamma_e \sin 2C', \tag{5.32g}$$

$$m'_{42} = (1 - 2\gamma_C^2)\sin\delta_C \sin 2C' + \gamma_e \cos 2C', \tag{5.32h}$$

$$m'_{43} = -(1 - 2\gamma_C^2)\sin\delta_C \cos 2C' + \gamma_e \sin 2C', \tag{5.32i}$$

$$m'_{44} = \cos\delta_C + 2\gamma_C\gamma_e, \tag{5.32j}$$

and with all other matrix elements vanishing. It is straightforward but tedious to substitute these equations into Eq. (5.20) to find the normalized Stokes vector of the beam incident on the sample.

Next, errors in the polarization detection arm that lead to different results from those of Eqs. (5.16) and (5.17) will be discussed, following the same sequence as for the errors in the polarization generation arm.

Consider first the presence of a reactor chamber exit window immediately in front of the rotating analyzer. As before, the Mueller matrix for the exit window is based on the retardation plate model:

$$\mathbf{m}_{W2} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -\delta_2 \sin 2W_2 \\ 0 & 0 & 1 & \delta_2 \cos 2W_2 \\ 0 & \delta_2 \sin 2W_2 & -\delta_2 \cos 2W_2 & 1 \end{bmatrix}, \tag{5.33}$$

where $\delta_2 \ll 1$ (in radians) is the retardance associated with assumed very weak stress-induced birefringence of the window.[56] In this analysis, recall that the Stokes vector of the light beam reflected from the sample is designated $\mathbf{S}_r = I_{0r}\ [1 \quad s_{2r} \quad s_{3r} \quad s_{4r}]^{\mathrm{T}}$ in the *p*-*s* reference frame. With this Stokes vector incident on the combined exit window and rotating analyzer, the irradiance at the detector follows the form:

$$\begin{aligned} I_{\mathrm{th}}(t) = (1/2)T_{At}T_{W2u}\, I_{0r} \{1 &+ (s_{2r} - \delta_2\, s_{4r} \sin 2W_2) \cos(2A') \\ &+ (s_{3r} + \delta_2\, s_{4r} \cos 2W_2) \sin(2A')\}, \end{aligned} \tag{5.34}$$

where T_{W2u} is the transmittance of the exit window for unpolarized light. A comparison with Eq. (5.8) shows that:

$$\alpha = s_{2r} - \delta_2\, s_{4r} \sin 2W_2, \tag{5.35a}$$

$$\beta = s_{3r} + \delta_2\, s_{4r} \cos 2W_2, \tag{5.35b}$$

$$I_0 = (1/2)T_{At}\, T_{W2u}\, I_{0r}. \tag{5.35c}$$

The first two equations can be solved to yield the two reflected beam Stokes parameters

$$s_{2r} = \alpha \pm \delta_2\, (1 - \zeta^2)^{1/2} \sin 2W_2, \tag{5.36a}$$

$$s_{3r} = \beta - [\pm\, \delta_2\, (1 - \zeta^2)^{1/2}] \cos 2W_2, \tag{5.36b}$$

where $\zeta^2 \equiv \alpha^2 + \beta^2$. Here it is assumed that only the first-order terms in the window retardance δ_2 (given in radians) are significant and that the reflected beam is completely polarized so that $s_{4r} = \pm(1 - s_{2r} - s_{3r})^{1/2}$. As long as one knows the characteristics of the window, then the sample parameters can be deduced by substitution of Eqs. (5.36) into Eqs. (5.19). The sign ambiguity is characteristic of the rotating analyzer polarization state detector. In fact, one can use a standard sample that generates a known reflected beam Stokes vector to extract the window characteristics without ambiguity by solving Eqs. (5.36) for δ_2 and W_2. Finally Eqs. (5.35) demonstrate that when the light beam reflected from the sample is linearly polarized with $s_{4r} = 0$, then the presence of the window has no effect on the measurement at least to first order in the parameter δ_2.

The derivation leading of Eqs. (5.16) and (5.17) is based on the assumption of a perfect rotating analyzer with a Mueller matrix given by Eq. (5.13a). As noted above for the polarizer, two different types of errors can occur in the analyzer that lead to modified Fourier coefficients from those in Eqs. (5.17).[1] First, the analyzer can exhibit a non-zero extinction ratio $\varepsilon_A \equiv T_{Ae}/T_{At}$, where T_{Ae} and T_{At} are the transmittances of the analyzer for linearly polarized light along the extinction and transmission axes, respectively. This imperfect analyzer leads to a Mueller matrix of identical form as Eq. (5.27) but with ε_P replaced by ε_A. Then the following waveform can be derived:

$$I_{\text{th}}(t) = T_{Au} I_{0r} \{1 + [(1 - \varepsilon_A)/(1 + \varepsilon_A)]\, s_{2r} \cos(2A') + [(1 - \varepsilon_A)/(1 + \varepsilon_A)]\, s_{3r} \sin(2A')\}, \tag{5.37}$$

so that the components of the reflected beam Stokes vector can be written as:

$$s_{2r} = [(1 + \varepsilon_A)/(1 - \varepsilon_A)]\, \alpha, \tag{5.38a}$$

$$s_{3r} = (\beta/\alpha)\, s_{2r}, \tag{5.38b}$$

$$I_{0r} = I_0/T_{Au}, \tag{5.38c}$$

where α, β, and I_0 are obtained in the experiment. Second, as described earlier, some polarizers transmit polarized light having slight ellipticity such that the semimajor axis is oriented along the transmission axis. In addition, when such a polarizer is used as an analyzer and is illuminated with a light beam polarized linearly along the *e*-axis, a weak light beam polarized linearly along the *t*-axis is transmitted. This is the case for Rochon analyzers fabricated from quartz, and the Mueller matrix of the analyzer is the same as that in Eq. (5.29), but with γ_P replaced by γ_A.

The resulting irradiance at the detector behind a rotating elliptic analyzer is given by:

$$I_{\text{th}}(t) = (1/2)T_{At}\, I_{0r}(1 + 2\gamma_A s_{4r}) \left\{1 + \frac{1 - 2\gamma_A^2}{1 + 2\gamma_A s_{4r}} s_{2r} \cos 2A' + \frac{1 - 2\gamma_A^2}{1 + 2\gamma_A s_{4r}} s_{3r} \sin 2A' \right\}. \quad (5.39)$$

In this case, the components of the reflected beam Stokes vector can be written as:

$$s_{2r} = [1 \pm 2\gamma_A \sqrt{1 - \zeta^2} - 2\gamma_A^2(1 - 2\zeta^2)]\alpha, \quad (5.40a)$$

$$s_{3r} = (\beta/\alpha)s_{2r}, \quad (5.40b)$$

$$I_{0r} = \frac{2I_0}{T_{At}(1 + 2\gamma_A s_{4r})}, \quad (5.40c)$$

where α, β, and I_0 are obtained in the experiment. With the reflected beam Stokes parameters s_{2r} and s_{3r} determined, at least to within a sign ambiguity, the sample parameters can be determined using Eqs. (5.19) and (5.20). Finally, Eqs. (5.38) and (5.40) demonstrate that when the light beam reflected from the sample is circularly polarized with $s_{2r} = s_{3r} = 0$, then the use of an analyzer with incomplete extinction or weak elliptical nature has no significant effect on the measurement.

Finally, the case of the polarization sensitive detection system will be considered.[42] In this case, the Stokes vector $\mathbf{s}_{\text{D}}$ and irradiance I_{th} at the detector can be written in terms of the following product of Mueller matrices acting on the Stokes vector of the reflected beam:

$$I_{\text{th}}\mathbf{s}_D = \mathbf{M}'_D \Re(-A')\mathbf{M}_A \Re(A')(I_{0r}\mathbf{s}_r). \quad (5.41)$$

Here $\mathbf{M}'_D$ is the Mueller matrix of the detection system components that characterizes the polarization sensitive detection system. This Mueller matrix is given in the *p-s* frame of reference of the sample. Equation (5.41) leads to the following waveform:

$$I_{\text{th}}(t) = (1/2)I_{0r}T_{At}T_{Du}\{1 + (1/2)(m'_{D12}s_{2r} + m'_{D13}s_{3r})\}\{1 + \alpha \cos 2A'(t) + \beta \sin 2A'(t) + \alpha_4 \cos 4A'(t) + \beta_4 \sin 4A'(t)\}, \quad (5.42)$$

whereby the observation of non-zero 4ω Fourier coefficients is one indication of detection system polarization sensitivity. In Eq. (5.42) T_{Du} is the transmittance of the detection system components, and $m_{Dij}{}'$ are its

normalized Mueller matrix elements. The following Fourier coefficient expressions can be derived from Eq. (5.41):

$$\alpha = \frac{s_{2r} + m'_{D12}}{1 + (1/2)(m'_{D12}s_{2r} + m'_{D13}s_{3r})}, \tag{5.43a}$$

$$\beta = \frac{s_{3r} + m'_{D13}}{1 + (1/2)(m'_{D12}s_{2r} + m'_{D13}s_{3r})}, \tag{5.43b}$$

$$\alpha_4 = \frac{(1/2)(m'_{D12}s_{2r} - m'_{D13}s_{3r})}{1 + (1/2)(m'_{D12}s_{2r} + m'_{D13}s_{3r})}, \tag{5.43c}$$

$$\beta_4 = \frac{(1/2)(m'_{D13}s_{2r} + m'_{D12}s_{3r})}{1 + (1/2)(m'_{D12}s_{2r} + m'_{D13}s_{3r})}. \tag{5.43d}$$

Equations (5.43a) and (5.43b) can be inverted to obtain the Stokes vector components of the reflected beam from which the sample properties are evaluated. The results are:

$$s_{2r} = \frac{(\alpha - m'_{D12}) + (1/2)m'_{D13}(m'_{D12}\beta - m'_{D13}\alpha)}{1 - (1/2)(m'_{D12}\alpha + m'_{D13}\beta)}, \tag{5.43e}$$

$$s_{3r} = \frac{(\beta - m'_{D13}) + (1/2)m'_{D12}(m'_{D13}\alpha - m'_{D12}\beta)}{1 - (1/2)(m'_{D12}\alpha + m'_{D13}\beta)}. \tag{5.43f}$$

In applying these equations to deduce sample properties, one must have previously characterized the detection system Mueller matrix, or at least the matrix elements m'_{D12} and m'_{D13} . By using a standard sample that generates known s_{2r} and s_{3r}, then one can extract these elements from either of Eqs. (5.43a) and (5.43b) or Eqs. (5.43c) and (5.43d). The latter two are the most useful since the 4ω Fourier coefficients are small and the s_{2r} and s_{3r} values can be approximated as those obtained from the 2ω coefficients, neglecting the polarization sensitivity. The following expressions are obtained by inversion of Eqs. (5.43c) and (5.43d):

$$m'_{D12} = \frac{-(\alpha_4 s_{2r} + \beta_4 s_{3r})}{\beta_4 s_{2r}s_{3r} + (1/2)\{\alpha_4(s_{2r}^2 - s_{3r}^2) - (s_{2r}^2 + s_{3r}^2)\}}, \tag{5.44a}$$

$$m'_{D13} = \frac{\alpha_4 s_{3r} - \beta_4 s_{2r}}{\beta_4 s_{2r}s_{3r} + (1/2)\{\alpha_4(s_{2r}^2 - s_{3r}^2) - (s_{2r}^2 + s_{3r}^2)\}}. \tag{5.44b}$$

Other ellipsometer errors in addition to polarization sensitivity in the detection system can give rise to non-zero 4ω Fourier coefficients, an

example being detector nonlinearity.[54] One test to isolate polarization sensitivity is to rotate the detector to different azimuthal angles relative to the p-s coordinate system. The detector angle dependence of the Mueller matrix elements is extracted from the transformation $\mathbf{m}'_D = \Re(-D')\ \mathbf{m}_D\Re(D')$, that leads to the equations

$$m'_{D12} = m_{D12}\cos 2D' - m_{D13}\sin 2D', \tag{5.45a}$$

$$m'_{D13} = m_{D12}\sin 2D' + m_{D13}\cos 2D', \tag{5.45b}$$

where D' denotes the detector angle relative to the *p-s* coordinate system. These equations are valid only for the purpose of computing the irradiance at the detector, in which case the transformation matrix $\Re(-D')$ has no influence. Upon rotating the detector m'_{D12} and m'_{D13} should show oscillatory behavior with D' whereas m_{D12} and m_{D13} are given in the frame of reference of the detector and therefore should not.

5.4.3 P_rXSA Configuration

Next, the rotating polarizer ellipsometer configuration will be described. For this configuration, the only possible variation to be considered is that in which the general element X, typically representing a compensator, is placed immediately before the sample, yielding the P_rXSA sequence. This avoids changes in the compensator (or other optical element) characteristics due to beam deviations caused by sample alignment irreproducibilities. One might expect to be able to apply the principle of time-reversal symmetry to the mathematical description of the error-free rotating analyzer ellipsometer and thereby extract the corresponding description of the error-free rotating polarizer ellipsometer. However, because the time-reversal analog of the $PXSA_r$ configuration is the P_rSXA (with the compensator after the sample), this principle is not applicable under all circumstances. It is clear that time-reversal symmetry holds if there is no optical element in position X or if the Mueller matrix $\mathbf{M}'$ representing optical element X commutes with $\mathbf{M}_S$ representing the sample. In such symmetric cases, one can select the corresponding equation for the rotating analyzer instrument, replace time t by $-t$, such that the optical fields oscillate according to $\exp(-i\omega t)$ rather than $\exp(i\omega t)$ (assuming that the optics convention is used throughout). This sign change reverses the signs of the polarization state ellipticities. In addition, all variables A' are replaced by P' and vice versa. Upon making these changes, the equations for the rotating polarizer ellipsometer result.

Thus, the principle of time reversal symmetry can apply to an ellipsometer configuration incorporating certain errors, as long as the error-free

ellipsometer configuration is similarly symmetric. For example, in converting the equations for a symmetric rotating analyzer ellipsometer in which the polarizer and analyzer are both elliptic in nature to those for the corresponding rotating polarizer ellipsometer, the ellipticity parameters γ_A and γ_P that characterize the analyzer and polarizer are changed to $-\gamma_P$ and $-\gamma_A$, respectively; in addition, the $A' \leftrightarrow P'$ reversal is applied as noted in the previous paragraph. Furthermore, for the symmetric ellipsometer configuration including both reaction chamber windows, the window retardances and angles (δ_1, W_1) and (δ_2, W_2) in the equations for the rotating analyzer system are replaced by (δ_2, W_2) and (δ_1, W_1), respectively, to obtain the equations for the corresponding rotating polarizer configuration. Because of the possibility of such symmetric interchanges, many of the results of the previous section may also be applicable for the rotating polarizer ellipsometer. As a consequence, this section will be far less comprehensive than the previous one.

In this section, a general solution to data analysis for the ideal rotating polarizer ellipsometer will be presented without appealing to time reversal symmetry. In this case, the same Stokes vector equation must be applied as in Eq. (5.12):

$$I_{\text{th}}\,\mathbf{s}_D = \Re(-A')\,\mathbf{M}_A\,\Re(A')\,\mathbf{M}_S\,\mathbf{M}'\,\Re(-P')\,\mathbf{M}_P\,\Re(P')\,(I_S\mathbf{s}_S), \quad (5.46)$$

where all the matrices are defined as before, and it is assumed that the light from the source is unpolarized so that $\mathbf{s}_S = [1 \quad 0 \quad 0 \quad 0]^{\mathrm{T}}$.[1] In this case, however, the matrix product is handled all at once in order to obtain I_{th} as a Fourier expansion in the time-dependent angle $2P'$. The simplest expression of the detector irradiance from this full matrix product can be given as in Eq. (5.9):

$$I_{\text{th}}(t) = I_0\{1 + \alpha \cos 2P'(t) + \beta \sin 2P'(t)\}, \quad (5.47)$$

with the following expressions

$$I_0 = (1/4)\,T_{At}\,T_u\,\mu_1 T_{Pt}\,I_S\{|r_p|^2 \cos^2 A' + |r_s|^2 \sin^2 A'\}, \quad (5.48a)$$

$$\alpha = \mu_2/\mu_1, \quad (5.48b)$$

$$\beta = \mu_3/\mu_1. \quad (5.48c)$$

In these equations,

$$\mu_j = \sum_{i=1}^{4} \tau_i m'_{ij}; \quad (j = 1,2,3,4), \quad (5.49)$$

where m'_{ij} are the elements of the normalized Mueller matrix $\mathbf{m}'$ given by $\mathbf{m}' = \mathbf{M}'/\mathrm{T}_u$, so that $m'_{11} = 1$, and τ_i are the components of a vector τ that satisfies the characteristics of a normalized Stokes vector. The components are defined by $\tau_1 = 1$ and

$$\tau_2 = \frac{\cos 2A' - \cos 2\psi}{1 - \cos 2A' \cos 2\psi}, \tag{5.50a}$$

$$\tau_3 = \frac{\sin 2A' \sin 2\psi \cos \Delta}{1 - \cos 2A' \cos 2\psi}, \tag{5.50b}$$

$$\tau_4 = \frac{\sin 2A' \sin 2\psi \sin \Delta}{1 - \cos 2A' \cos 2\psi}. \tag{5.50c}$$

In fact, if A' were replaced by P' in Eqs. (5.50), then τ_i would be the normalized Stokes parameters obtained at the entrance of the rotating analyzer in the $\mathrm{PSA_r}$ ellipsometer configuration (i.e., without optical element X). This result is a manifestation of the time reversal symmetry discussed in previous paragraphs.

From Eqs. (5.50a) and (5.50b), the sample parameters ψ and Δ can be deduced by inversion, yielding the familiar forms:

$$\tan \psi = \left| \sqrt{\frac{1 + \tau_2}{1 - \tau_2}} \right| |\tan A'|, \tag{5.51a}$$

$$\cos \Delta = \tau_3 \left| \sqrt{\frac{1}{1 - \tau_2^2}} \right|. \tag{5.51b}$$

As before, it is assumed that $0° \leq A' \leq 90°$ so that $\sin 2A' \geq 0$, and the sign of $\cos\Delta$ is carried by τ_3. These equations exhibit the same form as Eqs. (5.21) due to the Stokes vector characteristics of Eqs. (5.50). Only two of the three Eqs. (5.50) are employed since the third equation is not independent; in fact,

$$\tau_4 = \pm\sqrt{1 - \tau_2^2 - \tau_3^2}. \tag{5.52a}$$

An alternative useful expression for τ_4 derivable from Eqs. (5.50) is given by:

$$\tau_4 = |\sqrt{1 - \tau_2^2}| \sin \Delta. \tag{5.52b}$$

It is important to note that only two parameters are obtained in the experiment (α and β) so only two independent quantities can be determined. As

a result the rotating polarizer ellipsometer exhibits the same characteristic of the rotating analyzer ellipsometer in that it cannot be used to detect depolarization by the sample or any other optical element in the path. In fact, depolarization would lead to violations of Eqs. (5.52) and then three parameters must be extracted.

The strategy for data analysis in the P_rXSA configuration involves first noting that Eqs. (5.48b), (5.48c), and (5.49) represent two equations that provide the measurable quantities α and β in terms of the components of the hypothetical Stokes vector τ and the elements of the Mueller matrix $\mathbf{m}'$. A more accessible form of these equations is:

$$\alpha = \frac{m'_{12} + \tau_2 m'_{22} + \tau_3 m'_{32} + \tau_4 m'_{42}}{1 + \tau_2 m'_{21} + \tau_3 m'_{31} + \tau_4 m'_{41}}, \tag{5.53a}$$

$$\beta = \frac{m'_{13} + \tau_2 m'_{23} + \tau_3 m'_{33} + \tau_4 m'_{43}}{1 + \tau_2 m'_{21} + \tau_3 m'_{31} + \tau_4 m'_{41}}, \tag{5.53b}$$

which are the analogs of Eqs. (5.20) for the rotating analyzer ellipsometer. First τ_4 is eliminated using $\tau_4 = \pm\sqrt{1 - \tau_2^2 - \tau_3^2}$ and the equations are inverted in order to express the unknown quantities τ_2 and τ_3 in terms of the known quantities α, β, and m'_{ij}. Then the resulting expressions for τ_2 and τ_3 are substituted into Eqs. (5.51) to generate the sample parameters (ψ, Δ). Rather than attacking Eqs. (5.53) for the most general case, which involves considerable algebra, it is useful to deal with the simplest situations first. In this approach to be applied next, the values of m'_{ij} are substituted into Eqs. (5.53) first, which is then solved for τ_2 and τ_3. Clearly, the inversion problem for these simple cases will be much easier than the most general case.

First, consider the situation in which no optical element is in location X. In this case $T_u = 1$ and $m'_{ij} = \delta_{ij}$, the Kronecker delta. Then Eqs. (5.53) become $\alpha = \tau_2$ and $\beta = \tau_3$ which do not require inversion. Substituting $\tau_2 = \alpha$ and $\tau_3 = \beta$ into Eqs. (5.51) yields equations identical to Eqs. (5.21) of the rotating analyzer ellipsometer, but with P' replaced by A':

$$\tan\psi = \left|\sqrt{\frac{1+\alpha}{1-\alpha}}\right| |\tan A'|, \tag{5.54a}$$

$$\cos\Delta = \beta \left|\sqrt{\frac{1}{1-\alpha^2}}\right|. \tag{5.54b}$$

As before, these equations are important because they relate the sample parameters to the Fourier coefficients for a simple rotating polarizer ellipsometer design without a compensator.

The second very common case assumes an ideal compensator that is aligned with its fast axis F along the p direction so that $C' = 0°$. Then $\mathbf{m}' = \mathbf{m}_C$ of Eq. (5.15) but with $\psi_C = 45°$. As a result, $m'_{11} = m'_{22} = 1$, $m'_{33} = m'_{44} = \cos\delta_C$, $m'_{34} = -m'_{43} = \sin\delta_C$, and all other elements vanish. With these substitutions, Eqs. (5.53) become:

$$\alpha = \tau_2, \tag{5.55a}$$

$$\beta = \tau_3 \cos\delta_C - \tau_4 \sin\delta_C. \tag{5.55b}$$

In the inversion of Eq. (5.55b), it is convenient to apply Eq. (5.52b) to eliminate τ_4. Such a step is not useful under all circumstances, since it introduces $\sin\Delta$, which is unknown. In this circumstance, the substitution followed by inversion leads to

$$\tau_3 = \frac{\beta + |\sqrt{1-\alpha^2}|\sin\delta_C \sin\Delta}{\cos\delta_C}. \tag{5.56}$$

From here, it is easy to apply Eqs. (5.55a) and (5.56) in order to eliminate τ_2 and τ_3 from Eqs. (5.51). The final result for ψ is identical to that of Eq. (5.54a) for the P_rSA configuration; however, Δ in Eq. (5.54b) for the P_rSA is replaced by $\Delta + \delta_C$ when the ideal compensator is added at $C' = 0°$:

$$\cos(\Delta + \delta_C) = \beta \left| \sqrt{\frac{1}{1-\alpha^2}} \right|. \tag{5.57}$$

As in the case of the $PXSA_r$ configuration, the phase shift of the ideal compensator and sample add when F is along p. Furthermore, Eqs. (5.54a) and (5.57) are consistent with Eqs. (5.21a) and (5.22) if one makes the substitutions $P' \to A'$ in the latter, as required by time reversal symmetry, which is applicable because $\mathbf{M}_S$ and $\mathbf{M}'$ commute in this case.

A third case can be derived in which $C' = 0°$ and the characteristics of the compensator (ψ_C, δ_C) are arbitrary. In this case, T_u becomes $\{|t_F|^2 \cos^2 P' + |t_S|^2 \sin^2 P'\}$, and $m'_{11} = m'_{22} = 1$, $m'_{12} = m'_{21} = -\cos 2\psi_C$, $m'_{33} = m'_{44} = \sin 2\psi_C \cos\delta_C$, $m'_{34} = -m'_{43} = \sin 2\psi_C \sin\delta_C$, and all other elements vanish. With these substitutions, Eqs. (5.53) become:

$$\alpha = \frac{\tau_2 - \cos 2\psi_C}{1 - \tau_2 \cos 2\psi_C}, \tag{5.58a}$$

$$\beta = \frac{\sin 2\psi_C (\tau_3 \cos\delta_C - \tau_4 \sin\delta_C)}{1 - \tau_2 \cos 2\psi_C}. \tag{5.58b}$$

Eliminating τ_4 using Eq. (5.52b) and inverting yields:

$$\tau_2 = \frac{\alpha + \cos 2\psi_C}{1 + \alpha \cos 2\psi_C}, \tag{5.58c}$$

$$\tau_3 = \frac{\sin 2\psi_C \, (\beta + |\sqrt{1 - \alpha^2}| \sin\delta_C \sin\Delta)}{\cos\delta_C \, (1 + \alpha \cos 2\psi_C)}. \tag{5.58d}$$

Finally, substitution of these results into Eqs. (5.51) yields the sample parameters. In fact, the result for Δ is identical to that of Eq. (5.57), whereas the result for ψ is:

$$\tan\psi = \left| \sqrt{\frac{1 + \alpha}{1 - \alpha}} \right| \frac{|\tan A'|}{\tan \psi_C}, \tag{5.59}$$

so that $\tan\psi$ in Eq. (5.54a) is replaced by $\tan\psi \tan\psi_C$. Finally, Eq. (5.59) is consistent with Eq. (5.23) as long as $P' \rightarrow A'$ in the latter, again as required by time reversal symmetry.

The case of an ideal compensator with $\psi_C = 45°$ at an arbitrary angle C' is more difficult to handle. In this case, the Mueller matrix elements for substitution into Eqs. (5.53) are:

$$m'_{11} = 1, \tag{5.60a}$$

$$m'_{22} = \cos^2 2C' + \cos\delta_C \sin^2 2C', \tag{5.60b}$$

$$m'_{23} = m'_{32} = \sin 2C' \cos 2C' \, (1 - \cos\delta_C), \tag{5.60c}$$

$$m'_{24} = -m'_{42} = -\sin 2C' \sin\delta_C, \tag{5.60d}$$

$$m'_{33} = \sin^2 2C' + \cos \delta_C \cos^2 2C', \tag{5.60e}$$

$$m'_{34} = -m'_{43} = \cos 2C' \sin \delta_C, \tag{5.60f}$$

$$m'_{44} = \cos \delta_C, \tag{5.60g}$$

whereas all other elements vanish. These matrix elements are substituted into Eqs. (5.53). Considering cases such as this one in which ($m'_{i1} = 0$; $m'_{1j} = 0$; $\quad i, j = 2,3,4$), then Eqs. (5.54) become

$$\alpha = \tau_2 m'_{22} + \tau_3 m'_{32} + \tau_4 m'_{42}, \tag{5.61a}$$

$$\beta = \tau_2 m'_{23} + \tau_3 m'_{33} + \tau_4 m'_{43}. \tag{5.61b}$$

Eliminating τ_4 from these equations leads to a linear relationship between τ_2 and τ_3, e.g., $\tau_3 = c_1\tau_2 + c_0$, where

$$c_1 = \frac{m'_{23}m'_{42} - m'_{22}m'_{43}}{m'_{32}m'_{43} - m'_{33}m'_{42}}, \tag{5.62a}$$

$$c_0 = \frac{m'_{43}\alpha - m'_{42}\beta}{m'_{32}m'_{43} - m'_{33}m'_{42}}. \tag{5.62b}$$

Then by selecting Eq. (5.61a) and applying Eq. (5.52a) along with the linear relation to eliminate τ_4 and τ_3, respectively, one generates the following quadratic equation in τ_2:

$$[(c_1 m'_{32} + m'_{22})\tau_2 + (c_0 m'_{32} - \alpha)]^2 = {m'_{42}}^2[1 - \tau_2^2 - (c_1\tau_2 + c_0)^2], \tag{5.62c}$$

that is straightforward to solve. Of course there will be a sign ambiguity in this solution that also propagates through to τ_3. This ambiguity is characteristic of the rotating polarizer system and must be resolved in data reduction using prior knowledge of the sample. Finally, Eqs. (5.53) can be inverted by applying a similar approach as that leading to Eqs. (5.62) even for the most general Mueller matrix, $\mathbf{m}'$, for which no matrix elements are assumed to vanish. In this case the algebraic manipulations are quite extensive leading to a lengthy analytical expression, and so will not be provided here.

For the rotating polarizer system, the only error to be described in detail is source polarization.[41] In this case, the Stokes vector for the source is given in general by $\mathbf{s}_S = I_S\,[1 \quad s_{2s} \quad s_{3s} \quad s_{4s}]^{\mathrm{T}}$, where s_{2s}, s_{3s}, and s_{4s} are the normalized Stokes vector components that describe the source polarization. Substituting this result into Eq. (5.46) leads to the following waveform:

$$\begin{aligned} I_{th}(t) = (1/4)\, & T_{At} T_u \mu_1 T_{Pt} I_S \{|r_p|^2 \cos^2 A' + |r_s|^2 \sin^2 A'\} \\ & \times \{1 + (1/2)\,[(\mu_2/\mu_1)\, s_{2s} + (\mu_3/\mu_1) s_{3s}]\} \\ & \times \{1 + \alpha \cos 2P'(t) + \beta \sin 2P'(t) \\ & \quad + \alpha_4 \cos 4P'(t) + \beta_4 \sin 4P'(t)\}. \end{aligned} \tag{5.63}$$

Thus, detection of non-zero 4ω Fourier coefficients is one indication of source polarization. In Eq. (5.63) μ_1, μ_2, and μ_3 are defined in Eq. (5.49). The following Fourier coefficient expressions can be derived from Eq. (5.42):

$$\alpha = \frac{(\mu_2/\mu_1) + s_{2s}}{1 + (1/2)\,[(\mu_2/\mu_1) s_{2s} + (\mu_3/\mu_1) s_{3s}]}, \tag{5.64a}$$

$$\beta = \frac{(\mu_3/\mu_1) + s_{3s}}{1 + (1/2)\,[(\mu_2/\mu_1) s_{2s} + (\mu_3/\mu_1) s_{3s}]}, \tag{5.64b}$$

$$\alpha_4 = \frac{(1/2)\,[(\mu_2/\mu_1)s_{2s} - (\mu_3/\mu_1)s_{3s}]}{1 + (1/2)[(\mu_2/\mu_1)s_{2s} + (\mu_3/\mu_1)s_{3s}]}, \tag{5.64c}$$

$$\beta_4 = \frac{(1/2)\,[(\mu_2/\mu_1)s_{3s} + (\mu_3/\mu_1)s_{2s}]}{1 + (1/2)[(\mu_2/\mu_1)s_{2s} + (\mu_3/\mu_1)s_{3s}]}. \tag{5.64d}$$

It is interesting to note that if $(s_{jr}, m'_{D1j}; j = 2,3)$ in Eqs. (5.43) of the detector waveform for a PXSA$_r$ system are replaced by $\{(\mu_j/\mu_1), s_{js}; j=2,3\}$ then Eqs. (5.64) result. Thus, source polarization in a rotating polarizer system is the time-reversal analog of polarization sensitivity in the detection system of a rotating analyzer system. Equations (5.64a) and (5.64b) can be inverted to obtain the Stokes vector components of the reflected beam from which the sample properties are evaluated. The results are:

$$\mu_2/\mu_1 = \frac{(\alpha - s_{2s}) + (1/2)s_{3s}(s_{2s}\beta - s_{3s}\alpha)}{1 - (1/2)(s_{2s}\alpha + s_{3s}\beta)} \equiv \alpha_{\text{eff}}, \tag{5.65a}$$

$$\mu_3/\mu_1 = \frac{(\beta - s_{3s}) + (1/2)s_{2s}(s_{3s}\alpha - s_{2s}\beta)}{1 - (1/2)(s_{2s}\alpha + s_{3s}\beta)} \equiv \beta_{\text{eff}}. \tag{5.65b}$$

In applying these equations, one must have previously characterized the normalized Stokes vector of the source, or at least the two components s_{2s} and s_{3s}. Then the center expressions of Eqs. (5.65), defined as the quantities α_{eff} and β_{eff}, are known. In order to extract sample properties, one must invert the following two equations:

$$\alpha_{\text{eff}} = \frac{m'_{12} + \tau_2 m'_{22} + \tau_3 m'_{32} + \tau_4 m'_{42}}{1 + \tau_2 m'_{21} + \tau_3 m'_{31} + \tau_4 m'_{41}}, \tag{5.66a}$$

$$\beta_{\text{eff}} = \frac{m'_{13} + \tau_2 m'_{23} + \tau_3 m'_{33} + \tau_4 m'_{43}}{1 + \tau_2 m'_{21} + \tau_3 m'_{31} + \tau_4 m'_{41}}, \tag{5.66b}$$

obtained by substituting the expressions of Eq. (5.49) for μ_j (j = 1, 2, 3) into Eqs. (5.65). Finally, the resulting τ_2 and τ_3 are substituted into Eq. (5.51) in order to extract ψ and Δ.

If one uses a standard sample that generates known τ_2, τ_3 and τ_4 values via Eqs. (5.50), from which (μ_2/μ_1) and (μ_3/μ_1) can be determined via Eq. (5.49), then one can extract the two components s_{2s} and s_{3s} from either of Eqs. (5.64a) and (5.64b) or Eqs. (5.64c) and (5.64d). The latter two are the most useful since the 4ω Fourier coefficients are small and the (μ_2/μ_1) and (μ_3/μ_1) values can be approximated as those obtained from the 2ω

coefficients, neglecting the polarization sensitivity. The following expressions are obtained by inversion of Eqs. (5.64c) and (5.64d):

$$s_{2s} = \frac{-[\alpha_4(\mu_2/\mu_1) + \beta_4(\mu_3/\mu_1)]}{\beta_4(\mu_2/\mu_1)(\mu_3/\mu_1) + (1/2)\{\alpha_4[(\mu_2/\mu_1)^2 - (\mu_3/\mu_1)^2] - [(\mu_2/\mu_1)^2 + (\mu_3/\mu_1)^2]\}}, \tag{5.67a}$$

$$s_{3s} = \frac{\alpha_4(\mu_3/\mu_1) - \beta_4(\mu_2/\mu_1)}{\beta_4(\mu_2/\mu_1)(\mu_3/\mu_1) + (1/2)\{\alpha_4[(\mu_2/\mu_1)^2 - (\mu_3/\mu_1)^2] - [(\mu_2/\mu_1)^2 + (\mu_3/\mu_1)^2]\}}. \tag{5.67b}$$

As was noted earlier when polarization sensitivity of the detection system was discussed for rotating analyzer ellipsometers, errors in addition to polarization sensitivity in the detection system can give rise to non-zero 4ω Fourier coefficients. Source polarization can be isolated by rotating the source to different azimuthal angles relative to the *p*-*s* coordinate system. The source angle (S') dependence of the Mueller matrix elements is extracted from the transformation $\mathbf{s}_S = \Re(-S')\mathbf{s}_{S'}$, that leads to the equations

$$s_{2s} = s'_{2s}\cos 2S' - s'_{3s}\sin 2S', \tag{5.68a}$$

$$s_{3s} = s'_{2s}\sin 2S' + s'_{3s}\cos 2S'. \tag{5.68b}$$

Thus, upon rotating the source in this case s_{2s} and s_{3s} should show oscillatory behavior with S' whereas s'_{2s} and s'_{3s} should not, since they are given in the frame of reference of the source.

5.5 Precision Considerations

Errors in the measurement of the ellipsometric angles (ψ, Δ) due to source and/or detector noise establish the detection limits when the data are analyzed for weak features in the deduced optical properties or for small changes in the deduced film thickness or other structural parameters. Generally, the limits on the precision of the classical null ellipsometer are due to detector noise, owing to very low light levels encountered in the vicinity of the null.[22] In contrast, the limits on the precision of the rotating polarizer or analyzer ellipsometer are due to the inherent fluctuations in light flux, so-called shot noise, owing to the much higher light levels of operation.[24] Prior to the development of data acquisition routines in spectroscopic ellipsometry or in real time ellipsometry at a fixed wavelength, it is useful to be able to estimate — using the experimental characteristics of

the instrument — the limiting precision in ρ as well as the precision accessible based on the time available for signal averaging. In ex situ spectroscopic ellipsometry, the limiting precision is approachable owing to the long time available for individual measurements. In real time spectroscopic ellipsometry, the accessible precision is established by the time resolution desired in studying the dynamic processes of interest. A detailed theoretical analysis of the precision of the $PCSA_r$ system based on waveform sampling was first published almost 30 years ago, and it remains a key source of useful information, including the formalism to assess other system designs.[24] The analysis of the P_rCSA proceeds in an identical fashion, and in fact the conclusions are identical as well, as long as one replaces P' in the rotating analyzer treatment by A' for the rotating polarizer treatment.

For the $PCSA_r$ configuration, the uncertainties in ψ and Δ due to uncertainties in the sampled data will depend on the values of ψ and Δ and on the optical settings, P' and δ_C. For an ideal ellipsometer, no gain in precision can be achieved though an independent adjustment of C'. Thus, the choice $C' = 0°$ that decouples P' and δ_C is made to simplify the treatment.[24] One can define an effective amplitude reflection ratio from Eqs. (5.21a) and (5.22) as $\rho_{\text{eff}} = \{|\cot P'| \exp(i\delta_C)\}\rho = \tan\psi\, |\cot P'| \exp\{i(\Delta + \delta_C)\}$, where ρ is defined as usual by $\rho = r_p/r_s = \tan\psi \exp(i\Delta)$. Because ρ_{eff} is a function of α and β alone, it is clear that for any given surface, optimum precision in ψ and Δ can be achieved through independent adjustment of P' and δ_C, a situation only possible with $C' = 0°$. The following equations can be derived that relate uncertainties in the Fourier coefficients to uncertainties in ψ and Δ:

$$\delta(\tan\psi)_{\text{rms}} = \frac{\tan\psi}{1 - \alpha^2}\,\delta\alpha_{\text{rms}}, \tag{5.69a}$$

$$\delta\Delta_{\text{rms}} = \sqrt{\frac{(\delta\beta_{\text{rms}})^2 + \dfrac{\alpha^2\beta^2(\delta\alpha_{\text{rms}})^2}{(1 - \alpha^2)^2}}{1 - \alpha^2 - \beta^2}}, \tag{5.69b}$$

where δX_{rms} designates the root-mean-square uncertainty in X due to the experimental precision limitations. The values α, β, $\delta\alpha_{\text{rms}}$, and $\delta\beta_{\text{rms}}$ are related to the directly measurable results α', β', $\delta\alpha'_{\text{rms}}$, and $\delta\beta'_{\text{rms}}$ through the standard $2[A_S + (\phi/2)]$ rotational transformation [Eqs. (5.11)].

Three different situations may arise.[24] (1) If the detector can be considered ideal such as when the uncertainties in α and β are due to fluctuations in the source output, then $\delta\alpha_{\text{rms}}$ and $\delta\beta_{\text{rms}}$ are independent of α, β, and I_0, the latter defined as the time-averaged irradiance incident on the detector. (2) If the uncertainties are due to shot noise, then $\delta\alpha_{\text{rms}}$ and $\delta\beta_{\text{rms}}$ depend on α and β and are proportional to ${I_{0i}}^{-1/2}$. (3) If the uncertainties

are due to detector noise, the proportionality is to I_{0i}^{-1}. The explicit expressions for $\delta\alpha_{\text{rms}}$ for the three cases are:

$$\text{(Case 1)} \quad \delta\alpha_{\text{rms}} = \text{const.}, \tag{5.70a}$$

$$\text{(Case 2)} \quad \delta\alpha_{\text{rms}} = [(2 + \alpha^2)/I_0]^{1/2} [\eta_{\text{eff}}\, h\nu/\Delta t_{\text{tot}}]^{1/2}, \tag{5.70b}$$

$$\text{(Case 3)} \quad \delta\alpha_{\text{rms}} = [(2 + \alpha^2)/I_0^2]^{1/2} [I_{\text{nep}}\eta_{\text{eff}}\, h\nu/\Delta t_{\text{tot}}]^{1/2}, \tag{5.70c}$$

where I_0 is given by:

$$I_0 = (1/4)\, I_S\, T_{Pt}\, T_u\, T_{At}\, \{|r_p|^2 \cos^2 P' + |r_s|^2 \sin^2 P'\}. \tag{5.71}$$

In these equations, Δt_{tot}, is the total data acquisition time for all data samples, i.e., $\Delta t_{\text{tot}} = MN\Delta t$, where M is the number of optical cycles averaged, N is the number of sampled points per optical cycle, and Δt is the sampling interval for a single data point. In addition, Eqs. (5.70b) and (5.70c) are based on the noise characteristics of a PMT detection system, where η_{eff} is the PMT quantum efficiency, I_{nep} is the PMT noise equivalent power, and $h\nu$ is the photon energy. In Eqs. (5.70), the corresponding expressions for $\delta\beta_{\text{rms}}$ are obtained by substituting β for α. In the first case of source fluctuations the two constants are identical since $\delta\alpha_{\text{rms}} = \delta\beta_{\text{rms}}$.

Equations (5.69) through (5.71) are very important in that they allow assessment of the optimum operating parameters $(P')_{\text{opt}}$ and $(\delta_C)_{\text{opt}}$.[24] For case (1), given by Eq. (5.70a), optimum measurement conditions occur when $\alpha = \beta = 0$, so that circularly polarized light is generated upon reflection. In order to achieve this condition, Eqs. (5.21a) and (5.22) show that one must have $|\tan(P')_{\text{opt}}| = \tan\psi$ and $(\delta_C)_{\text{opt}} = (\pm 90° - \Delta)$. It should also be noted that several instrument errors that reduce the accuracy of the measurement are also eliminated when circularly polarized light is reflected from the sample. These include detector nonlinearity, signal processing phase shift and suppression for the 2ω Fourier components of the detector waveform (which are no longer present), and ellipticity and partial polarization analysis associated with the analyzer. An especially interesting case occurs for a rotating analyzer ellipsometer in the absence of a compensator, in which case $\delta_C = 0°$. The optimum now corresponds simply to $\alpha = 0$, so that $|\tan(P')_{\text{opt}}| = \tan\psi$ and $\cos\Delta = \beta$. The optimization criteria imply that tracking the polarizer angle with ψ as a function of photon energy in spectroscopic measurements or as a function of time in dynamic measurements is a simple way to maintain a higher precision. Tracking can use the same stepping motor mounted on the polarizer that would normally be for calibration purposes.

As an aside, it should be noted that high precision and accuracy rotating analyzer ellipsometers in the PSA_r configuration have been developed in the past for both spectroscopic[57] and real time single-wavelength[58]

ellipsometry applications, using the principal-angle-of-incidence concept. In these ellipsometers, the true polarizer angle P' and the angle of incidence θ of the probe beam at the sample surface (the latter taking the place of the adjustable compensator retardance δ_C in the $PCSA_r$ configuration) are both set to achieve the optimum condition of circularly polarized light upon reflection. Automatic balancing can be performed to yield $\alpha' = \beta' = 0$ in a time as short as 0.5 s, allowing real time observations with this type of system. For measurements of processes in vacuum, however, the requirement of a variable angle of incidence in general implies that the optical elements must be mounted within the vacuum chamber. In fact, the first principal-angle-of-incidence ellipsometer for use in real time applications adopted this very difficult approach.[58] With the trends in the 1980s and 1990s toward variable angle of incidence spectroscopic ellipsometry with auto-compensation[25,59] and real time spectroscopic ellipsometry,[21] the principal-angle-of-incidence designs have not attracted sustained interest.

Continuing with the ideal detector case under the optimum settings of $|(P')_{opt}| = \psi$ and $(\delta_C)_{opt} = (\pm 90° - \Delta)$ for the $PCSA_r$ configuration and $|(P')_{opt}| = \psi$ for the PSA_r configuration, the results for the precision in the measurement of $|\delta\rho|$ under these optimum conditions are given by:[24]

$$(PCSA_r) \qquad |\delta\rho|_{rms} = (\sqrt{2}\tan\psi)\delta\alpha_{rms}, \tag{5.72a}$$

$$(PSA_r) \qquad |\delta\rho|_{rms} = \sqrt{\frac{2-\beta^2}{1-\beta^2}}(\tan\psi)\delta\alpha_{rms}, \tag{5.72b}$$

where

$$|\delta\rho|_{rms} = \sqrt{[\delta(\tan\psi)_{rms}]^2 + \tan^2\psi[\delta\Delta_{rms}]^2}. \tag{5.73}$$

Thus in both instrument configurations, precision improves with an ideal detector as $\tan\psi$ is minimized; this occurs when the angle of incidence is equal to the principal angle. When $\beta = \cos\Delta$, Eq. (5.72b) demonstrates that it is impossible to maintain precision in Δ for $\Delta \cong 0°$ or 180° without the compensator. In addition, with the compensator in place, the same problem occurs if $|\cos(\Delta + \delta_C)| = 1$. [Equation (5.72a) assumes an adjustable compensator is available, however, and has been set such that $\cos(\Delta + \delta_C) = 0$.] As a result of the above considerations, complete measurements of the complex optical functions $\varepsilon = \varepsilon_1 - i\varepsilon_2$ of bulk dielectric materials are very difficult without a compensator, although $\tan\psi$ and the real part of the dielectric function ε_1 can still be obtained with accuracy. Thus, for bulk dielectric materials in which the imaginary part of the dielectric function ε_2 is known to vanish, the rotating analyzer ellipsometer

can be used without a compensator, however, the difficult situation arises when one seeks to determine low ε_2 values.

The ideal detector case may be relevant when laser sources are used as in process monitoring applications. However, when continuum sources (which tend to exhibit a higher stability by an order of magnitude or more) are used, either shot noise or detector noise determines the uncertainties. The former tends to be relevant for PMT detectors, and the latter for solid state detectors, particularly those used in infrared ellipsometry.

Because I_0 is independent of δ_C, the optimum criteria for δ_C for the cases of (2) shot noise and (3) detector noise remain unchanged, namely $\cos(\Delta + \delta_C) = 0$. The expressions corresponding to those of Eqs. (5.72) for cases (2) and (3) are quite complicated and the optimum ψ must be determined numerically. Some of the main conclusions will be summarized here assuming $P' > 0°$ (see Fig. 5.7).[24] For both cases, $(P')_{\text{opt}}$ deviates from ψ such that $(P')_{\text{opt}} > \psi$ for $\psi < 45°$ and $(P')_{\text{opt}} < \psi$ for $\psi > 45°$.

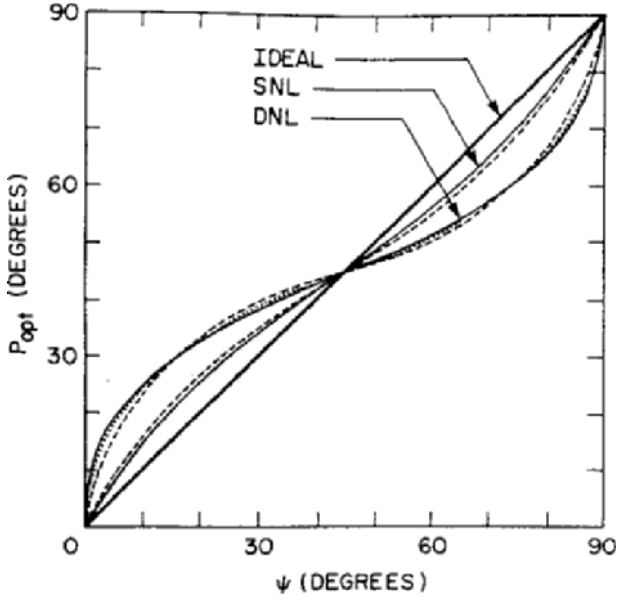


Figure 5.7 Optimum value of the polarizer angle P′ for a rotating analyzer ellipsometer as a function of ψ for three different situations: operation assuming an ideal detector, shot noise limited operation (SNL), and detector noise limited operation (DNL) assuming $(\Delta + \delta_C) = 90°$ (solid lines). Also shown are the corresponding curves calculated for $(\Delta + \delta_C) = 50°$ (dotted lines) and $(\Delta + \delta_C) = 10°$ (dashed lines). In fact, the curves are identical when calculated for $(\Delta + \delta_C)$ values for which the values of the function $\cos^2(\Delta + \delta_C)$ are also identical. [After Ref. 24.]

The deviation is largest for the detector noise limitation [case (3)]. In contrast to case (1) in Eqs. (5.72), for case (2) of shot noise limitation, the uncertainty, $|\delta\rho|_{rms}$, does not vanish as $\psi \rightarrow 0$, but gradually approaches a constant, whereas for case (3) of detector noise limitation, $|\delta\rho|_{rms}$ increases according to $\tan^{-1/2}\psi$ as $\psi \rightarrow 0$. Such behavior indicates that in case (2), one need not be at the principal angle for acceptably low uncertainty. In fact, in case (3), it is desirable to avoid the principal angle by performing measurements at higher angles of incidence. In both cases, the uncertainty is near its minimum and a very weak function of $(\Delta + \delta_C)$ for $(\Delta + \delta_C) > 20°$, but rises dramatically as $(\Delta + \delta_C)$ approaches 0°. This indicates that a compensator is not a requirement except for samples with Δ within 20° of 0° and 180°. Finally, in these two cases, the minima in the uncertainty in $|\delta\rho|_{rms}$ as a function of P' that defines $(P')_{opt}$ are relatively broad (as long as P' is within the range of 20°–70°) and deviations from the optimum by as much as 10° are easily tolerable.

When the reflecting surface is uniform and isotropic with a complex dielectric function, ε, and it forms an atomically smooth interface to the ambient of dielectric function ε_a, then

$$\varepsilon = \varepsilon_a \sin^2\theta\left\{1 + \left(\frac{\rho - 1}{\rho + 1}\right)^2 \tan^2\theta\right\}, \tag{5.74}$$

where θ is the angle of incidence.[7] From Eq. (5.74) the uncertainty in the measured dielectric function can be determined as

$$|\delta\varepsilon|_{rms} = \left|\frac{4(\varepsilon - \sin^2\theta)}{1 - \rho^2}\right| |\delta\rho|_{rms}. \tag{5.75}$$

The uncertainty $|\delta\varepsilon|_{rms}$ is minimized as a function of ρ when ρ is purely imaginary, and as a function of θ when θ is near the pseudo-Brewster angle where $\tan\psi$ is a minimum. As noted earlier, the latter provides a minimum in $|\delta\rho|_{rms}$ for cases (1) and (2) corresponding to the ideal detector and shot noise precision limitation, respectively. The uncertainty in ε vs. photon energy has been calculated for Si and Au samples for a theoretical PSA_r system without any form of tracking for optimization.[24] It was also assumed that the system was limited by shot noise, with $\eta_{eff} = 0.2$, $\Delta t_{tot} = 5$ s, and $I_0 = 100$ nW. For Au with P' set in the range of 30° to 40°, the theoretical precision in ε, i.e., $|\delta\varepsilon|_{rms}$, was determined to be $< 2 \times 10^{-5}$ from 2 eV to 5 eV, and was essentially the same as that possible assuming optimization at each photon energy by variation of P' and δ_C (with a compensator). For Si with $P' = 30°$, the precision was found to be $\sim 5 \times 10^{-5}$ at 5 eV and $\sim 2 \times 10^{-4}$ at 2 eV, the former value being the same as the optimum and the latter more than a factor of 5 higher because

of the absence of the compensator. Thus, for a material such as Si in which Δ varies significantly, from near 0° or 180° to near 90° as the spectrum is traversed, automatic adjustment of the compensator for optimum precision during measurement provides significant advantages. In general, for measurements of semiconductors over a spectral range that spans from below to above the lowest direct gap, or for measurements of film/substrate structures with strongly modulated interference patterns, such automatic compensator adjustment is very helpful. In contrast, polarizer tracking is of lesser importance. For example, for both Au and Si a fixed polarizer angle of $P' = 30°$ was found to be suitable for spectroscopic measurements.

Performance measurements on a single crystal of Ni at 3.1 eV using an optimized spectroscopic ellipsometer in the PSA_r configuration, have revealed a precision of 2×10^{-5}, on the same order as that predicted theoretically. With an average flux on the PMT of 60 nW and 1000 optical cycle signal averaging, the factor limiting the precision was found to be source fluctuation. This quoted precision translates to remarkable values of $\pm 0.0005°$ in ψ and $\pm 0.001°$ in Δ, or a sensitivity in terms of film thickness on the order of 0.01 Å. Similarly good performance has been reproduced in a second laboratory.[60]

5.6 Calibration Procedures

In the discussion of rotating polarizer and analyzer ellipsometers so far, the angles describing the true orientation of the polarizer and analyzer transmission axes P' and A', and the true orientation of the compensator fast axis C' relative to the p-direction have been assumed to be accurately known values. In general, however, only nominal values are available and these are identified by the unprimed variables P, C, and A. Ellipsometer calibration involves determining the corrections P_S, C_S, and A_S to the nominal values that yield the true values in accordance with the equations:

$$P' = P - P_S, \tag{5.76a}$$

$$A' = A - A_S, \tag{5.76b}$$

$$C' = C - C_S. \tag{5.76c}$$

For the fixed elements, the nominal values may come from an actual physical reading on a rotator mount, as in the early null ellipsometers. More likely, the nominal values are those values stored in the memory of a control computer that identify the angular position of a stepping motor.

In either case, it is assumed that P_S, C_S, and A_S are relatively small, ~1° or less, for the fixed elements so that the nominal values are approximately equal to the true values. For the rotating element, however, the value of X_S ($X = A$ or P) can lie within the full range of $-90° < X_S \leq 90°$ and is determined by the analyzer position when the data acquisition begins, as discussed earlier.

It is important to design the calibration methods for the offset and phase angles, i.e., for P_S, C_S, and A_S, so that they can be performed just prior to sample measurement, with the sample in place, using exactly the same configuration and alignment as is used in the measurement. Two different approaches have been described for the determination of P_S, C_S, and A_S for rotating polarizer and analyzer ellipsometers. The first approach is a small angle method that involves determining specific functions of the measured Fourier coefficients at a given wavelength versus a fixed element angular reading in the neighborhood of the p and/or s directions [i.e., $X \approx X_S$ and/or $X \approx X_S + (\pi/2)$ radians, respectively, with $X = C$ or A for the rotating-polarizer instrument and $X = P$ or C for the rotating-analyzer instrument].[46,61] In the small angle method, second-order expansions of $\sin(X - X_S)$ and $\cos(X - X_S)$ are applied. As a result, the functions of the Fourier coefficients can be fit with linear or quadratic expressions in X to obtain C_S and A_S or P_S and C_S. No other information on the sample or the compensator is required in the fitting. The alternative approach is a large angle method in which the measured Fourier coefficients (α', β') at a given wavelength are obtained at numerous closely spaced values of the fixed element angular reading $X = P$, C, or A all the way from the p to the s-direction.[62] The results are complicated functions of the parameters (ψ, Δ), (ψ_C, δ_C), P_S, C_S, and A_S, and these functions can be fitted to deduce all the parameters (or certain functions of the parameters). Thus, in this case, sample and compensator information is determined along with the calibration angles.

In both calibration approaches, specific error parameters have also been extracted in addition to the desired calibration angles. Thus, these calibrations can serve the dual purposes of fixed element angular offset and rotating element phase angle determination as well as a detailed error assessment. For example, in one of the first demonstrations of the small angle method, the ellipticity parameters γ_P, γ_C, and γ_A of the quartz Rochon polarizer, compensator, and analyzer were incorporated into the calibration equations for a wavelength scanning rotating analyzer ellipsometer.[46] In the first demonstration of the large angle method, no compensator was used, which simplified overall calibration considerably; however, the parameters m'_{D12} and m'_{D13} that describe the polarization sensitivity of the detection system were determined. For both instruments, the ac/dc gain ratio of the detection system electronics η^{-1} was also determined.[62]

5.6.1 Ideal Rotating Polarizer and Analyzer Ellipsometers

In this section, the primary focus will be on small angle methods of calibration. In certain cases, however, large angle approaches will be described where their implementation is most useful. Generally, the large angle approaches can apply the expressions directly from Section 5.4 and need little further elaboration. In contrast, the small angle methods require derivation of additional formulas that are second-order expansions in the small angle. The overall instrument calibration procedure first involves the calibration of the compensator in straight-through during the initial setup of the instrument prior to addition of the sample. This is a useful first step since addition of the sample into the P_rCA and PCA_r configurations leads to a second wavelength dependent pair of parameters, and it is often difficult to extract the compensator characteristics (ψ_C, δ_C) and sample characteristics (ψ, Δ) independently. Once possible compensator calibrations have been described, then different small angle methods will be presented for the rotating polarizer and analyzer ellipsometers. These approaches include the residual/phase function calibration,[46] the zone-difference-phase/phase function calibration,[63] and the irradiance/phase function calibration.[64] At the end, a general in situ method of compensator retardance calibration will be described. Throughout this section, the ellipsometer will be considered ideal to the extent that no optical errors will be included; however, unavoidable errors due to the electronic circuitry will be incorporated.

5.6.1.1 PCA_r and P_rCA Configurations: Initial Compensator Calibration

In the initial compensator calibration steps, the ellipsometer is assumed to be set for straight-through with the sample removed (i.e., in the P_rCA or PCA_r configuration). Here, for definiteness, the PCA_r or rotating analyzer configuration is considered. The compensator is set at the scale reading, $C = 0°$, which is near the most common setting of $C' = 0°$ (or $C = C_S$) used in measurements of samples (in which case the compensator fast axis is along the p-direction). Under the aforementioned conditions, the residual function R is given by:

$$R(P) = 1 - (\alpha'^2 + \beta'^2) \tag{5.77a}$$

$$= 1 - \frac{1}{\eta^2}(\alpha^2 + \beta^2) \tag{5.77b}$$

$$R(P) \cong \left(1 - \frac{1}{\eta^2}\right) + \frac{4}{\eta^2}(\cot^2 \psi_C \sin^2 \delta_C)(P - P_C)^2; \; P \approx P_C = P_S - C_S. \tag{5.77c}$$

In these equations, P_C describes the azimuthal angle of the fast axis of the compensator in the polarizer frame of reference, given in general by $P_C = C + (P_S - C_S)$.[46] Thus, $P - P_C = (P - C) + (C_S - P_S)$ is the azimuthal angle of the transmission axis of the polarizer in the compensator frame of reference. As long as both P_S and C_S are small, experimental Fourier coefficients can be determined at a number of closely spaced polarizer scale readings in the vicinity of $P \cong C = 0°$, and a minimum in the experimental residual function will be observed. This minimum is predicted to occur at $P_C = P_S - C_S$. Thus, by fitting the experimental results for $R(P)$ at a fixed wavelength to a second-order polynomial, e.g., $R(P) = c_0 + c_1P + c_2P^2$, then P_C is determined as the value of P at which the minimum in the best fit polynomial occurs, e.g., $P_C = -c_1/2c_2$. In addition, η can be determined from the minimum value $R_{\min} = R(P_C)$ of the best fit function:

$$\eta = \frac{1}{\sqrt{1 - R_{\min}}}. \tag{5.77d}$$

Finally, Eq. (5.77c) shows that the calibration can be performed most precisely when the curvature of the parabola in $P - P_C$ is maximized. This occurs at the quarter wave point of the compensator where $\delta_C = 90°$.

A similar development is needed to determine the phase of the analyzer. In this case, the phase function is defined by:

$$\Theta(P) = \frac{\tan^{-1}(\beta'/\alpha')}{2} \tag{5.78a}$$

$$= (A_C + \phi/2) + \frac{\tan^{-1}(\beta/\alpha)}{2} \tag{5.78b}$$

$$\cong (A_C + \phi/2) + (\cot \psi_C \cos \delta_C)(P - P_C); \quad P \approx P_C. \tag{5.78c}$$

In these equations, A_C describes the azimuthal angle of the fast axis of the compensator in the analyzer frame of reference, given in general by $A_C = C + (A_S - C_S)$. Thus, $A - A_C = (A - C) + (C_S - A_S)$ is the azimuthal angle of the transmission axis of the analyzer in the compensator frame of reference. The experimental Fourier coefficients, determined at fixed wavelength and at a number of closely spaced polarizer scale readings in the vicinity of $P \cong C = 0°$ as described in the previous paragraph, are now substituted into Eq. (5.78a). Calibration proceeds by

fitting the experimental results to a linear function, i.e., $\Theta(P) = c_0 + c_1 P$; then $(A_C + \phi/2)$ is determined as the value of $\Theta(P_C)$, e.g., $(A_C + \phi/2) = c_0 + c_1 P_C$. It should be noted that the contributions to the analyzer phase angle from the arbitrary placement of the analyzer in its mounting (which defines A_S and A_C) and from the phase shift of the electronic circuitry (which defines $\phi/2$) cannot be separated. Finally, Eq. (5.78c) shows that the phase angle calibration can be performed most precisely when the slope of the linear relationship in $P - P_C$ is minimized. This optimum also occurs at the quarter wave point of the compensator where $\delta_C = 90°$.

Lastly, it should be noted that if the inverse tangent in Eq. (5.78a) is confined by computation to the range $-\pi/2 < 2\Theta(P) \leq \pi/2$, then $(A_C + \phi/2)$ similarly must lie within the range $-\pi/4 < (A_C + \phi/2) \leq \pi/4$ for the equations to be valid. If the phase angle lies outside this range, then quadrant corrections of $\pm\pi/2$ will be needed on the right-hand side of Eq. (5.78c).[46] In order to avoid troublesome quadrant corrections, it is helpful to pre-align the transmission axis of the analyzer with respect to the motor shaft in the straight-through configuration. In this procedure, $A' - \phi/2$ is set to a value near zero at the onset of data collection ($t = 0$), thus avoiding any possibly ambiguous quadrant corrections.[53]

For the rotating polarizer ellipsometer, a similar approach to that of the previous three paragraphs is applied with the exceptions that the Fourier coefficients are determined at a number of closely spaced analyzer readings in the vicinity of $A \cong C = 0°$ and the minimum in Eq. (5.77c) occurs at $A_C = C + (A_S - C_S) = A_S - C_S$. Equation (5.78c) then becomes a linear relationship in $A - A_C$ which is evaluated at A_C to deduce $P_C + \phi/2$. In general, Eqs. (5.77) and (5.78) are valid for the rotating polarizer ellipsometer as well, but with P replaced everywhere by A and vice versa.

Next, two possible approaches for determining the spectrum in δ_C are summarized. The first uses a rotating polarizer or analyzer ellipsometer in the P_rCA or PCA_r configuration in which the compensator is treated as the sample and the second uses the fixed PCA configuration with a large angle rotation of the compensator. Both approaches are performed in straight-through with $C = 0°$, and both rely on the above calibrations to determine either $(P_C, A_C + \phi/2)$ for the rotating analyzer or $(P_C + \phi/2, A_C)$ for the rotating polarizer ellipsometer.

In the first approach, the description is based on the rotating analyzer ellipsometer configuration.[38] The compensator is fixed at the scale reading $C = 0°$, and the assumed fixed polarizer is first rotated so that $P = 45° + P_C$ and then fixed, and spectra in (α', β') are collected. These are interpreted using the same equations as for rotating-analyzer ellipsometry in the PSA_r reflection geometry, but with (P_S, A_S) replaced by (P_C, A_C), and (ψ, Δ) replaced by (ψ_C, δ_C). Thus, the following expressions from

Eqs. (5.11) and (5.21) are applied:

$$\alpha = \eta\,[\alpha' \cos\{2[A_C + (\phi/2)]\} + \beta' \sin\{2[A_C + (\phi/2)]\}], \quad (5.79a)$$

$$\beta = \eta\,[-\alpha' \sin\{2[A_C + (\phi/2)]\} + \beta' \cos\{2[A_C + (\phi/2)]\}], \quad (5.79b)$$

$$\tan \psi_C = \sqrt{\frac{1+\alpha}{1-\alpha}}, \quad (5.80a)$$

$$\cos \delta_C = \sqrt{\frac{\beta^2}{1-\alpha^2}}, \quad (5.80b)$$

where the positive square roots are to be taken. In this case also, the compensator parameters in the rotating polarizer configuration can be determined from the same procedure and set of equations, but replacing A_C by P_C and vice versa.

In the second approach for compensator retardance calibration, which is a large angle method,[62] the rotating-analyzer PCA_r ellipsometer configuration is assumed. In addition, for simplicity, the compensator is assumed to be ideal so that $\psi_C = 45°$. Initially, the compensator is removed from the bench, and the analyzer rotation is stopped at an arbitrary angle and fixed in position. Then an irradiance function calibration is performed with $P - P_A = (P - A) + (A_S - P_S) \cong 0°$ where P_A describes the azimuthal angle of the transmission axis of the analyzer in the polarizer frame of reference, given by $P_A = A + (P_S - A_S)$. This involves measuring the dc irradiance level at the detector $I'_0(P)$ for several closely spaced polarizer angular settings P such that the transmission axes of the polarizer and analyzer are nearly parallel. In the absence of errors, $I'_0(P)$ exhibits a minimum when $P = P_A$ (via Malus's Law). The value of P_A can be obtained accurately by fitting $I'_0(P)$ to a parabolic function of P. Next, the polarizer is fixed at a known value relative to the analyzer angle, i.e., a known value of $P - P_A$. Finally, the rotatable compensator is added while maintaining the straight-through configuration. In this configuration with an arbitrary starting compensator setting C, the irradiance I_n at the detector is measured at a number N of equally spaced settings $C_n = C + (n\pi/2N)$ $(n = 0, 1 \ldots N - 1)$ (in radians) spanning a full 90° range in C. These N irradiance values are Fourier analyzed according to the approach of Eqs. (5.3). The $2C$ ac Fourier coefficients vanish, whereas the $4C$ coefficients (α'_4, β'_4) yield the compensator retardance in accordance with the following expression:[53]

$$\delta_C = 2\cos^{-1}\sqrt{\frac{1 - \sqrt{\alpha'^2_4 + \beta'^2_4}}{1 + \sqrt{\alpha'^2_4 + \beta'^2_4}\cos 2(P - P_A)}}. \quad (5.81)$$

In the case of a rotating polarizer ellipsometer, the same overall procedure can be applied. In this case, however, in the first steps, the polarizer rotation is halted and $A_P = P + (A_S - P_S)$ is determined by stepping the analyzer. In the final step, $P - P_A$ in Eq. (5.81) must be replaced by $A - A_P$.

5.6.1.2 PSA_r and P_rSA Configurations: Ellipsometer Calibration

As described earlier, the rotating polarizer and analyzer ellipsometer configurations without a compensator in the optical path are important for the study of material systems in which the ellipsometry angle Δ is not near 0° or 180°, e.g., strongly absorbing semiconductors (for photon energies above the lowest direct band gap) and metals. Here, three procedures for the calibration of the polarizer angular offset and analyzer phase will be described for the rotating analyzer ellipsometer. The corresponding equations for the rotating polarizer ellipsometer are obtained as usual by an interchange of all angular variables P and A.

Residual-Function/Phase-Function Calibration

The residual function calibration approach for rotating analyzer ellipsometers in the PSA_r configuration is a small angle method relying on the fact that a light beam reflected from the surface of an isotropic sample will be linearly polarized if the incident wave is linearly polarized along the orthogonal *p*- or *s*-directions.[46] For any sample in which $\Delta \neq 0°$ or 180°, the polarization state of the reflected light beam gradually increases in ellipticity as the linear polarization state of the incident beam is rotated away from these two directions. In the absence of errors then, the experimentally determined modulation amplitude function $\alpha'^2 + \beta'^2$ should reach its maximum value of unity for $P' = 0$ or $\pi/2$ radians, i.e., for P' along the orthogonal *p*- and *s*-directions. The residual function $R(P)$ is defined as the deviation of the amplitude function from its maximum, which is minimized in the absence of errors when $P' = 0$ or $\pi/2$, or alternatively, when $P = P_S$ or $P_S + (\pi/2)$.

The following expressions for the residual function form the basis for the calibration to determine P_S:

$$R(P) = 1 - (\alpha'^2 + \beta'^2) \tag{5.82a}$$

$$= 1 - \frac{(\alpha^2 + \beta^2)}{\eta^2} \tag{5.82b}$$

$$R(P) = \left(1 - \frac{1}{\eta^2}\right) + 4\,\frac{(P - P_S)^2 \cot^2\psi \sin^2\Delta}{\eta^2}; \quad P \approx P_S, \tag{5.82c}$$

$$\cong \left(1 - \frac{1}{\eta^2}\right) + 4\frac{\{P - [P_S + (\pi/2)]\}^2 \tan^2\psi \sin^2\Delta}{\eta^2}; \quad P \approx P_S + (\pi/2). \tag{5.82d}$$

Equation (5.82a) provides an expression into which the experimental Fourier coefficients (α', β') can be substituted. Equation (5.82b) is derived by substituting the inverted form of the transformation of Eqs. (5.11) into Eq. (5.82a). Equations (5.82c) and (5.82d) can be derived from Eq. (5.82b) using the methods of Section 5.4.1; these two derivations are valid to second order in the small quantities $(P - P_S)$ and $\{P - [P_S + (\pi/2)]\}$, respectively. The Fourier coefficients (α', β') are determined at fixed wavelength and at a number of closely spaced polarizer scale readings P in the vicinity of either P_S or $P_S + (\pi/2)$, and a minimum in the experimental residual function will be observed in each case. By fitting these experimental results for $R(P)$ to a second-order polynomial, then P_S or $P_S + (\pi/2)$ is determined as the value of P at which the minimum in the best fit polynomial occurs. In this way, P_S can be determined either from Eq. (5.82c) or from Eq. (5.82d). In addition, η is determined from the best fit polynomial evaluated either at P_S or at $P_S + (\pi/2)$ according to Eq. (5.77d).

A similar development is needed to determine the rotating analyzer phase angle, A_S.[46] In this case, the phase function is defined by:

$$\Theta(P) = \frac{\tan^{-1}(\beta'/\alpha')}{2} \tag{5.83a}$$

$$= [A_S + (\phi/2)] + \frac{\tan^{-1}(\beta/\alpha)}{2} \tag{5.83b}$$

$$\cong [A_S + (\phi/2)] + (P - P_S)\cot\psi \cos\Delta; \quad P \approx P_S, \tag{5.83c}$$

$$\cong [A_S + (\phi/2)] + \{P - [P_S + (\pi/2)]\}\tan\psi \cos\Delta; \quad P \approx P_S + (\pi/2). \tag{5.83d}$$

Equation (5.83a) provides an expression into which one substitutes the experimental Fourier coefficients (α', β') collected at different polarizer angles. In addition, Eqs. (5.83c) and (5.83d) are derived from Eq. (5.83b) using the methods of Section 5.4.1. These two derivations are valid to second order in the small quantities $(P - P_S)$ and $\{P - [P_S + (\pi/2)]\}$,

respectively; in fact, the quadratic terms vanish. Thus, calibration proceeds from an analysis of the experimental results of Eq. (5.83a) obtained at a number of closely spaced polarizer scale readings in the vicinity of either P_S or $P_S + (\pi/2)$. These results are fitted to a linear function; then $[A_S + (\phi/2)]$ is determined in two possible ways as the value of $\Theta(P_S)$ or $\Theta[P_S + (\pi/2)]$. Finally, if the analyzer is fixed within its mounting such that $A' - \phi/2$ is near zero at the onset of data collection ($t = 0$), possible quadrant ambiguities in Eqs. (5.83) are eliminated as described in Section 5.6.1.1.

Limitations on the residual-function/phase-function calibration arise because there are certain ranges in (ψ, Δ) over which the calibration fails. Failure of the method occurs when Δ is near 0° or ±180° because the second-order terms in the polynomials of Eqs. (5.82c) and (5.82d) nearly vanish, and the minima that define P_S essentially disappear. Since P_S cannot be determined under these conditions, $[A_S + (\phi/2)]$ cannot be determined either. Thus, P_S is found most accurately via Eqs. (5.82c) and (5.82d) when the curvature of the parabola is maximized; $[A_S + (\phi/2)]$ is found most accurately via Eqs. (5.83c) and (5.83d) when the slope of the linear function is minimized. These optimum conditions occur for Eqs. (5.82c) and (5.83d) when $|\Delta| \sim 90°$ and $\psi \sim 0°$, and for Eqs. (5.82d) and (5.83c) when $|\Delta| \sim 90°$ and $\psi \sim 90°$. Thus, if Δ is sufficiently far away from 0° or ±180° for the residual-function/phase-function approach to be effective, i.e., if $30° < |\Delta| < 150°$, then Eqs. (5.82c) and (5.83d) are preferable when $\psi < 45°$ and Eqs. (5.82d) and (5.83c) are preferable when $\psi > 45°$. In practice, however, if $|\Delta| < 30°$ or $|\Delta| > 150°$, an alternative approach is needed for accurate offset and phase angle calibration. For many samples, the wavelength of calibration can be varied and a value selected such that the P_S and $[A_S + (\phi/2)]$ values from the residual-function/phase-function calibration will be accurate. For samples such as transparent glasses, however, an alternate approach is unavoidable.

Zone-Difference-Phase-Function/Phase-Function Calibration

The zone-difference phase function provides an alternative method to obtain P_S:[63]

$$\Phi(P) = \Theta(P) - \Theta[P + (\pi/2)] \tag{5.84a}$$

$$= \frac{\tan^{-1}(\beta'/\alpha')|_P - \tan^{-1}(\beta'/\alpha')|_{P+(\pi/2)}}{2} \tag{5.84b}$$

$$\cong 2(P - P_S)\cot 2\psi \cos\Delta; \quad P \approx P_S. \tag{5.84c}$$

Equation (5.84b) provides an expression into which the experimental spectra (α', β') collected at different pairs of polarizer angles $[P, P + (\pi/2)]$

can be substituted. The derivation leading to Eq. (5.84c) is valid to second order in the small quantity $(P - P_S)$. A quadrant correction may be required in Eq. (5.84b) in order to apply Eq. (5.84c). Thus, calibration proceeds by fitting the experimental results at a given wavelength to a linear function $\Phi(P) = c_0 + c_1 P$; then P_S is determined as the value of P for which $\Phi = 0$, i.e., $P_S = -c_0/c_1$. Once P_S is determined by this method, $[A_S + (\phi/2)]$ can be determined from the phase function as described either by Eq. (5.83c) or Eq. (5.83d).

The offset angle P_S can be determined most accurately from Eq. (5.84c) when the slope of the linear relationship is maximized. This occurs when $|\Delta| = 0°$ or 180° and $\psi = 0°$ or 90°. This calibration procedure fails when the slope of the relationship in Eq. (5.84c) vanishes, i.e., when $|\Delta| = 90°$ or $\psi = 45°$. Thus, the zone-difference phase function method of calibration is complementary to the residual function method. As a general rule, the residual-function/phase-function calibration is suitable for metals and for semiconductors excited at photon energies above the lowest direct bandgap, whereas the zone-difference-phase-function/ phase-function calibration is suitable for dielectrics and for semiconductors excited at photon energies below the lowest direct bandgap.

Irradiance-Function/Phase-Function Calibration

The previously described calibration approaches in rotating analyzer ellipsometry involve acquiring spectra in the dc-normalized 2ω Fourier coefficients (α', β') of the irradiance waveform at the detector as a function of the polarizer angular reading P. In these approaches, the angular dependence of the dc Fourier coefficient, i.e., the average irradiance in the reflected beam, is used only for normalization of the 2ω Fourier coefficients. Thus, potentially useful information in I'_0 of Eq. (5.7) is discarded. Here it is shown how this information can be employed in instrument calibration. The basic equations from which P_S can be determined from I'_0 are as follows:

$$I'_0(P) = I_{00}\{|r_p|^2 + (|r_s|^2 - |r_p|^2)(P - P_S)^2\};\ P \approx P_S, \tag{5.85a}$$

$$= I_{00}\{|r_s|^2 - (|r_s|^2 - |r_p|^2)(P - [P_S + (\pi/2)])^2\};\ P \approx P_S + (\pi/2), \tag{5.85b}$$

where I_{00} is an experimental instrument response function that does not depend on the sample properties.[64] These expressions are valid to second order in the small quantities $(P - P_S)$ and $\{P - [P_S + (\pi/2)]\}$. Thus, by fitting the experimental results for I'_0 at a given wavelength to a second-order polynomial, then P_S or $P_S + (\pi/2)$ is determined as the value of P at which the extremum in the polynomial occurs. Both Eqs. (5.85a) and (5.85b) provide the same sensitivity for the determination

of P_S. The highest sensitivity occurs when the difference between the reflectances for linearly polarized light in the p and s directions are maximized, the procedure fails when $|r_p|^2 = |r_s|^2$ or when $\psi = 45°$. The irradiance function calibration has an advantage over the more widely used residual function and the zone-difference phase function approaches because there are no constraints on the value of Δ for accurate calibration. Once P_S is determined according to Eq. (5.85a) or (5.85b), A_S can be deduced as usual from the phase functions in Eq. (5.83c) or (5.83d).

5.6.1.3 PCSA$_r$ and P$_r$CSA Configuration: Compensator Calibration

It is natural to ask whether it is possible to perform the rotating analyzer ellipsometer calibrations of Section 5.6.1.2 with the compensator in place. In fact, such calibrations will work correctly only when the compensator fast axis is precisely aligned along the p or s direction so that $C' = 0°$ or 90°. Only under these conditions is the incident polarization state unchanged upon transmission through the compensator. One of these conditions can be set, however, only after the compensator angular offset is determined. In fact, it is possible to iterate two separate procedures to calibrate the ellipsometer with the compensator in place. This involves (i) setting $C = 0°$ and performing one of the three calibrations of Section 5.6.1.2 to obtain P_S, (ii) setting $P = P_S$ and performing the calibration to determine C_S as described in the next paragraph, (iii) setting $C = C_S$ and repeating one of the three calibrations of Section 5.6.1.2, and (iv) iterating the last two steps until convergence to a single pair of (P_S, C_S) is observed. Convergence generally occurs quite quickly as long as C_S is small ($<1°$). Under these conditions, with $C = C_S$, the values of $[A_S + (\phi/2)]$ and η obtained with one of the three procedures of Section 5.6.1.2 are then correct. Furthermore, if a compensator is incorporated and the correct setting $C = C_S$ is made, then in all equations of Section 5.6.1.2, Δ is replaced by $\Delta + \delta_C$ and ψ is unchanged (assuming an ideal compensator with $\psi_C = 45°$). Finally, this overall procedure will also work for the rotating polarizer ellipsometer with incorporated compensator, however, the principle of time-reversal symmetry described in Section 5.4.3 is not valid when the compensator is moved from $C = C_S$.

First a small angle approach will be described for the determination of C_S in the PCSA$_r$ configuration. When $P = P_S$, the following expressions for the residual function form the basis for the calibration to determine C_S:

$$R(C) = 1 - (\alpha'^2 + \beta'^2) \tag{5.86a}$$

$$R(C) = 1 - \frac{\alpha^2 + \beta^2}{\eta^2} \tag{5.86b}$$

$$\cong \left(1 - \frac{1}{\eta^2}\right) + 4\frac{\cot^2\psi\{4\sin^2(\delta_C/2) - [\cos\Delta - \cos(\Delta + \delta_C)]^2\}(C - C_S)^2}{\eta^2};$$

$$C \approx C_S, \tag{5.86c}$$

$$\cong \left(1 - \frac{1}{\eta^2}\right) + 4\frac{\cot^2\psi\{4\sin^2(\delta_C/2) - [\cos\Delta - \cos(\Delta - \delta_C)]^2\}\{C - [C_S + (\pi/2)]\}^2}{\eta^2};$$

$$C \approx C_S + (\pi/2). \tag{5.86d}$$

Equation (5.86a) provides an expression into which one can substitute the experimental Fourier coefficients (α', β'). Equations (5.86c) and (5.86d) can be derived from Eq. (5.86b) using the methods of Secton 5.4.1; these two derivations are valid to second order in the small quantities $(C - C_S)$ and $\{C - [C_S + (\pi/2)]\}$, respectively. The Fourier coefficients (α', β') are determined at fixed wavelength and at a number of closely spaced compensator scale readings C in the vicinity of either C_S or $C_S + (\pi/2)$. By fitting these experimental results for $R(C)$ to a second-order polynomial, then C_S or $C_S + (\pi/2)$ is determined as the value of C at which the minimum in the best fit polynomial occurs. In this way, C_S can be determined either from Eq. (5.86c) or from Eq. (5.86d). This approach fails if $\psi = 90°$ or if $\delta_C = 0°$.

To conclude this section, a general large angle approach will be described for determination of all calibration information in the $PCSA_r$ configuration. In this calibration, the Fourier coefficients (α', β') are measured at a number of compensator and/or polarizer angular settings over the 180° range from −90° to 90°. The functional variations $[\alpha'(P), \beta'(P)]$ and/or $[\alpha'(C), \beta'(C)]$ can be fit by least squares regression analysis using Eqs. (5.11), (5.17), (5.18), and (5.24) presented in Sections 5.3 and 5.4. In their most useful form, these equations are given as follows:

$$\alpha' = \frac{1}{\eta}[\alpha\cos\{2[A_S + (\phi/2)]\} - \beta\sin\{2[A_S + (\phi/2)]\}], \tag{5.87a}$$

$$\beta' = \frac{1}{\eta}[\alpha\sin\{2[A_S + (\phi/2)]\} + \beta\cos\{2[A_S + (\phi/2)]\}], \tag{5.87b}$$

where

$$\alpha = \frac{s_{2i} - \cos 2\psi}{1 - s_{2i}\cos 2\psi}, \tag{5.87c}$$

$$\beta = \frac{\sin 2\psi\,(s_{3i}\cos\Delta + s_{4i}\sin\Delta)}{1 - s_{2i}\cos 2\psi}, \tag{5.87d}$$

and where

$$s_{2i} = \cos^2(\delta_C/2)\,\cos 2P' + \sin^2(\delta_C/2)\,\cos[2(2C' - P')], \tag{5.87e}$$

$$s_{3i} = \cos^2(\delta_C/2)\,\sin 2P' + \sin^2(\delta_C/2)\,\sin[2(2C' - P')], \tag{5.87f}$$

$$s_{4i} = \sin\delta_C\,\sin[2(C' - P')]. \tag{5.87g}$$

In this analysis, trial values are assigned to the seven calibration parameters $\{\eta, [A_S + (\phi/2)], \psi, \Delta, P_S, \delta_C, C_S\}$, and Eqs. (5.87) then provide theoretical expressions for $[\alpha'(P), \beta'(P)]$ and/or $[\alpha'(C), \beta'(C)]$ from the selected values of P and C. The theoretical and experimental results are compared by calculating a weighted mean square deviation, which is a measure of the fitting error.[65] Then adjustments to the trial values are made iteratively in order to minimize the fitting error. For the best fit set of parameters, correlation coefficients and confidence limits must be computed to assess the meaningfulness of the values. Typical confidence limits of $\sim\pm 0.02°$ are sought for the six angular variables.

Two strategies can be employed if confidence limits are too large. First, one can collect and analyze a larger data set covering the two dimensional (P, C) parameter space. Second, one can reduce the size of the free parameter set by determining some of the values in alternative calibrations, e.g., those based on the small angle approach. For example, if $\{\eta, [A_S + (\phi/2)], P_S, C_S\}$ can be obtained in advance, then one can set $P = P_S$, and $\{\psi, \Delta, \delta_C\}$ can be deduced by fitting experimental data $[\alpha'(C), \beta'(C)]$ to a simpler set of equations. Such equations are given by Eqs. (5.87a) and (5.87b), but with Eqs. (5.87c) through (5.87g) replaced by:

$$\alpha = \frac{[\cos^2(\delta_C/2) + \sin^2(\delta_C/2)\cos 4C'] - \cos 2\psi}{1 - \cos 2\psi\,[\cos^2(\delta_C/2) + \sin^2(\delta_C/2)\cos 4C']}, \tag{5.88a}$$

$$\beta = \frac{\sin 2\psi\,[\cos\Delta\sin^2(\delta_C/2)\,\sin 4C' + \sin\Delta\sin\delta_C\sin 2C']}{1 - \cos 2\psi\,[\cos^2(\delta_C/2) + \sin^2(\delta_C/2)\,\cos 4C']}. \tag{5.88b}$$

This large angle calibration approach is advantageous for the $PCSA_r$ instrument since it can provide a means for checking the compensator

calibration in situ, i.e., without removing the sample from the ellipsometer axis. In order to perform a full calibration of the compensator, however, as described in Section 5.6.1.1, the large angle calibration must be performed at numerous closely spaced values of the wavelength. Finally, it should be noted that Eqs. (5.86) through (5.88) are valid only for the $PCSA_r$ optical configuration. It is straightforward to derive the corresponding results for the P_rCSA configuration, and these will be given next.

When $A = A_S$, the following expressions for the residual function in the P_rCSA configuration form the basis for the calibration to determine C_S:

$$R(C) = 1 - (\alpha'^2 + \beta'^2) \tag{5.89a}$$

$$= 1 - \frac{(\alpha^2 + \beta^2)}{\eta^2} \tag{5.89b}$$

$$= \left(1 - \frac{1}{\eta^2}\right) + \frac{1}{2}\frac{\sin^2\delta_C\{1 - \cos 4(C - C_S)\}}{\eta^2} \tag{5.89c}$$

$$\cong \left(1 - \frac{1}{\eta^2}\right) + 4\frac{\sin^2\delta_C(C - C_S)^2}{\eta^2}; \quad C \approx C_S, \tag{5.89d}$$

$$\cong \left(1 - \frac{\cos^2\delta_C}{\eta^2}\right) - 4\frac{\sin^2\delta_C\{C - [C_S + (\pi/4)]\}^2}{\eta^2}; \quad C \approx C_S + (\pi/4). \tag{5.89e}$$

As usual, the Fourier coefficients (α', β') are determined at fixed wavelength and at a number of closely spaced compensator scale readings C in the vicinity of either C_S or $C_S + (\pi/4)$. By fitting these experimental results for $R(C)$ to a second-order polynomial, then C_S or $C_S + (\pi/4)$ is determined as the value of C at which the extrema in the best fit polynomial occurs. In this way, C_S can be determined either from Eq. (5.89d) or from Eq. (5.89e).

Next, the general large-angle approach will be described for determination of all calibration information in the P_rCSA configuration. As described before, the Fourier coefficients (α', β') are measured at a number of compensator and/or analyzer angular settings from $-90°$ to $90°$. The functional variations $[\alpha'(C), \beta'(C)]$ and/or $[\alpha'(A), \beta'(A)]$ can be fit by least squares regression analysis using Eqs. (5.11), (5.48), (5.50), (5.60), and (5.61) presented in Sections 5.3 and 5.4. The useful form of these equations is given as follows:

$$\alpha' = \frac{1}{\eta}[\alpha\cos 2\{P_S + (\phi/2)\} - \beta\sin 2\{P_S + (\phi/2)\}], \tag{5.90a}$$

$$\beta' = \frac{1}{\eta}\left[\alpha \sin 2\{P_S + (\phi/2)\} + \beta \cos 2\{P_S + (\phi/2)\}\right], \quad (5.90b)$$

where

$$\alpha = \tau_2 m'_{22} + \tau_3 m'_{23} + \tau_4 m'_{42}, \quad (5.90c)$$

$$\beta = \tau_2 m'_{23} + \tau_3 m'_{33} + \tau_4 m'_{43}, \quad (5.90d)$$

and where

$$\tau_2 = \frac{\cos 2A' - \cos 2\psi}{1 - \cos 2A' \cos 2\psi}, \quad (5.90e)$$

$$\tau_3 = \frac{\sin 2A' \sin 2\psi \cos \Delta}{1 - \cos 2A' \cos 2\psi}, \quad (5.90f)$$

$$\tau_4 = \frac{\sin 2A' \sin 2\psi \sin \Delta}{1 - \cos 2A' \cos 2\psi}, \quad (5.90g)$$

$$m'_{22} = \cos^2 2C' + \cos\delta_C \sin^2 2C', \quad (5.90h)$$

$$m'_{23} = \sin 2C' \cos 2C' (1 - \cos\delta_C), \quad (5.90i)$$

$$m'_{42} = \sin 2C' \sin \delta_C, \quad (5.90j)$$

$$m'_{33} = \sin^2 2C' + \cos\delta_C \cos^2 2C', \quad (5.90k)$$

$$m'_{43} = -\cos 2C' \sin\delta_C. \quad (5.90l)$$

In this analysis, the seven calibration parameters $\{\eta, [P_S + (\phi/2)], \psi, \Delta, A_S, \delta_C, C_S\}$ are determined by fitting the data $[\alpha'(C), \beta'(C)]$ and/or $[\alpha'(A), \beta'(A)]$ with the theoretical expressions of Eqs. (5.90). Finally, if $\{\eta, [P_S + (\phi/2)], A_S, C_S\}$ can be obtained in advance, then one can set $A = A_S$, and δ_C can be deduced by fitting experimental data $[\alpha'(C), \beta'(C)]$ to a much simpler set of equations. Such equations are given by Eqs. (5.90a) and (5.90b), but with Eqs. (5.90c) through (5.90l) replaced by:

$$\alpha = m'_{22} = \cos^2(\delta_C/2) + \sin^2(\delta_C/2)\cos 4C', \quad (5.91a)$$

$$\beta = m'_{23} = \sin^2(\delta_C/2)\sin 4C' \quad (5.91b)$$

It is clear that the P_rCSA configuration is easier to calibrate in this way because the measured Fourier coefficients (α', β') do not depend on the sample properties (ψ, Δ) when the fixed analyzer is set at $A = A_S$.

5.6.2 Detecting and Correcting Errors in Calibration

Next, modifications of the small angle calibration methods will be discussed that allow one to extract the polarizer and analyzer offset and phase angles for the PSA_r (and P_rSA) configuration(s) even in the presence of errors due to the source or detector, or due to imperfect optical components.[21,46,61] In the modified small angle calibration methods, the strategy is to incorporate first-order corrections into the residual, phase, zone-difference phase, and irradiance functions that account for the effects of the errors or imperfections. Then expansions of these functions in the neighborhood of the *p* and/or *s* directions allow one to express the predicted P_S and A_S values in terms of (i) minima or intersection points obtained experimentally by fitting these functions and (ii) correction terms that are linear in the error parameters and depend on (ψ, Δ), as well. In fact, if the error parameters are known in advance, then the expressions for P_S and A_S can be evaluated using the zero-order values of (ψ, Δ), designated (ψ_0, Δ_0), i.e., the values deduced from calibration and data reduction equations in which all error parameters are set to zero. If the error parameters are not known in advance, sometimes they can be determined by performing the calibrations in both the *p* and *s* directions (or by performing multiple calibration methods) and solving the multiple equations for P_S, A_S, and the error parameters. This approach will be illustrated later in this section, but only the final results of such derivations will be provided.

First, the influence of reaction chamber windows on calibration for the PSA_r system will be discussed. Stress-induced birefringence generated either in the window fabrication process or by improper mounting (or both) leads to changes in the polarization state of the light beam as it enters and exits the chamber. By far the most desirable approach is to minimize stress and the resulting birefringence as much as possible by specially designed fabrication and mounting procedures. It should be emphasized that standard, off-the-shelf windows mounted on flanges for optical access to ultrahigh vacuum (uhv) environments may exhibit such severe birefringence that the first-order Mueller matrix representation of Eq. (5.33) may be inappropriate. Specialized uhv windows[50] in which the birefringence is nearly eliminated are now widely available, but costly. Because electrochemical cells are often fabricated from Teflon, stress-free fused silica optical flats can be mounted with o-rings; as a result, window problems are not nearly as severe.

The quantitative details of the calibration procedure incorporating the window errors are given in the following paragraphs. If such errors exist and are not taken into account, calibration in the $PW_1SW_2A_r$ configuration will be in error and will distort subsequent data reduction. As in Sections 5.4.1 and 5.4.2, imperfect entrance and exit windows can be represented

as waveplates with relative retardances, δ_1 and δ_2 given in radians (such that $\delta_1 \approx 0$ and $\delta_2 \approx 0$), and azimuthal angles, W_1 and W_2, respectively.[56] The required algebraic manipulations to determine the effects of windows are best undertaken within the 4×4 Mueller matrix framework as in Eq. (5.33). Here, the following definitions are made to simplify the notation:

$$B_{Wi} \equiv -M_{i24} = M_{i42} = \delta_i \sin 2W_i, \tag{5.92a}$$

$$A_{Wi} \equiv M_{i34} = -M_{i43} = \delta_i \cos 2W_i, \tag{5.92b}$$

where the subscripts $i = 1$ and 2 designate the entrance and exit windows, respectively.

It is convenient to define P_1 as the polarizer angle near $P = P_S$ at which the minimum occurs in the parabolic best fit to the measured residual function [as in Eq. (5.82c)]. Furthermore, A_1 is defined by evaluating the linear best fit to the phase function, $\{\tan^{-1}(\beta'/\alpha')\}/2$, at P_1 [as in Eq. (5.83c). With these two definitions, the following expressions are valid to first order in all the Mueller matrix elements of the two windows:[21]

$$P_S = P_1 - \frac{B_{W2}\tan\psi_0 + B_{W1}\cos\Delta_0}{2\sin\Delta_0}, \tag{5.93a}$$

$$A_S = A_1 - \frac{B_{W1}\cot\psi_0 + B_{W2}\cos\Delta_0}{2\sin\Delta_0}. \tag{5.93b}$$

Evidently, the calibration angles are independent of the Mueller matrix elements A_{W1} and A_{W2} to first order. In Eqs. (5.93), ψ_0 and Δ_0 are determined as zero-order approximations using the calibration procedures of Section 5.6.1 without the window corrections, i.e., with $B_{W1} = 0$ and $B_{W2} = 0$. Furthermore, if P_S is small, then it can also be neglected in this zero-order approximation. It is clear that for proper correction of P_1 and A_1 to give P_S and A_S, the sign of Δ_0 must be known. In the measurement of bulk media, this requirement causes no problems; however, for layered materials with strongly modulated Δ spectra versus wavelength, the sign ambiguity must be resolved as part of the overall data analysis problem. Note also that as Δ approaches 0° or ±180°, the first-order corrections of Eqs. (5.93) become very large. This problem compounds the failure of the residual function calibration, and samples (or wavelengths) must be chosen to avoid these conditions.

If the window parameters B_{Wi} have been determined in advance, e.g., by the methods described in Section 5.4.2, then P_S and A_S can be determined directly from Eqs. (5.93). Otherwise the calibration can be performed in the vicinity of the minimum in the residual function $R(P)$, not

only near $P = P_S$ but also near $P = P_S + (\pi/2)$, and both B_{W1} and B_{W2} can be determined explicitly. If the minimum in the best parabolic fit to the experimental $R(P)$ near $\pi/2$ occcurs at $P_2 + \pi/2$, and if the linear fit to the phase function, $\{\tan^{-1}(\beta'/\alpha')\}/2$, evaluated at $P_2 + \pi/2$ is A_2, then B_{W1} and B_{W2} can be expressed solely in terms of the known calibration values and (ψ_0, Δ_0):

$$B_{W1} = (A_1 - A_2)\tan 2\psi_0 \sin\Delta_0, \tag{5.94a}$$

$$B_{W2} = -(P_1 - P_2)\tan 2\psi_0 \sin\Delta_0. \tag{5.94b}$$

For useful results from these expressions, ψ_0 cannot be near 0° or 90° and Δ_0 cannot be near 0° or ±180°. With B_{W1} and B_{W2} determined from Eqs. (5.94), P_S and A_S can now be determined via Eqs. (5.93).

The calibration angles determined from the zone-difference phase function method are given to first order in the Mueller matrix elements of the windows as:[21]

$$P_S = P_0 + \frac{B_{W1}\tan\Delta_0}{2}, \tag{5.95a}$$

$$A_S = A_0. \tag{5.95b}$$

Here P_0 is the P-axis intercept of the best fit to the experimentally determined zone-difference phase function, and A_0 is obtained by evaluating the best linear fit to the phase function at P_0. Note also that in the range where this calibration is useful, i.e., for $|\Delta_0| < 30°$ and $|\Delta_0| > 150°$, the window correction is small.

It should be noted that this first-order treatment of window errors in calibration has neglected the effects of the transfer characteristics of the detection circuitry. As in Section 5.6.1, η is determined from the residual function minimum near $P = P_S$ or $P_S + (\pi/2)$ via Eq. (5.77d), i.e., $\eta = (1 - R_{\min})^{-1/2}$. Again, the correction $\phi/2$ is included automatically in all the measured analyzer angles A_S, A_0, A_1, and A_2 since it is inseparable from these angles. To include the phase lag explicitly, A_S, A_0, A_1, and A_2 must be replaced by $A_S + (\phi/2)$, $A_0 + (\phi/2)$, $A_1 + (\phi/2)$, and $A_2 + (\phi/2)$, respectively. In fact, throughout this section these detection system errors will be ignored. Finally, for application to the rotating polarizer system in the configuration P_rSA, the quantities P_S, A_S, P_1, A_1, B_{W1}, B_{W2}, P_2, A_2, P_0, and A_0 in Eqs. (5.93) through (5.95) must be replaced by A_S, P_S, A_1, P_1, B_{W2}, B_{W1}, A_2, P_2, A_0, and P_0, respectively, adopting the principle of time reversal symmetry developed in Section 5.4.3.

It is clear from Eqs. (5.94) that one cannot fully characterize the two windows in a conventional two-zone residual/phase function calibration

procedure. If one is interested in performing such an analysis on the initial state of the windows, i.e., prior to mounting on the chamber, then a separate optical arrangement in the straight-through $PCWA_r$ configuration is convenient. Initially, the system is set up without the window for circularly polarized light at the detector, for example, with $\delta_C = \pi/2$ and $P - P_C = -\pi/4$ radians. Then, upon insertion of the window, which is centered and aligned with respect to the beam, the amplitude of the 2ω component provides δ_i and a phase function provides $P_{Wi} = W_i + P_S$, which describes the azimuthal angle of the fast axis of the window in the polarizer frame of reference. The final results are given by:

$$\delta_i = \sqrt{\alpha'^2 + \beta'^2}, \tag{5.96a}$$

$$P_{Wi} = \frac{\tan^{-1}(-\alpha'/\beta')}{2} + (P_C - A_C) + \frac{\pi \,\mathrm{sgn}(-\alpha')\mathrm{u}(-\beta')}{2}, \tag{5.96b}$$

where the third term in Eq. (5.96b) is a quadrant correction, defined such that $\mathrm{u}(x) = 0$ for $x < 0$ and $\mathrm{u}(x) = 1$ for $x \geq 0$ and $\mathrm{sgn}(x) = -1$ for $x < 0$ and $\mathrm{sgn}(x) = +1$ for $x \geq 0$. An alternative form of these equations is:

$$\delta_i \cos 2P_{Wi} = \alpha' \sin 2(P_C - A_C) + \beta' \cos 2(P_C - A_C), \tag{5.96c}$$

$$\delta_i \sin 2P_{Wi} = -\alpha' \cos 2(P_C - A_C) + \beta' \sin 2(P_C - A_C). \tag{5.96d}$$

When the two window flanges are then attached to the deposition system, centered and aligned (e.g., making use of bellows connections), their retardation values may not be altered significantly if the flanges are tightened uniformly and the beam impinges on the same spot. However, the window corrections become uncertain because the window azimuths relative to the plane of incidence are now known only approximately. By a judicious combination of PCW_1A_r and PCW_2A_r measurement procedures and two-zone calibration, the full imperfection matrices for the mounted windows can be calculated in principle.

Next, the case of imperfect polarizer and analyzer elements will be considered, assuming that the normal modes of these elements are elliptical polarization states with orthogonal tilt angles, but very small ellipticities.[46] When the polarizer and analyzer are Rochon elements fabricated from optically active quartz, then these normal modes exhibit ellipticities of

$$\gamma_i = \pm 0.0010\, E\ (\mathrm{eV}^{-1}), \tag{5.97}$$

where the subscript i designates either the polarizer (P) or analyzer (A), the + and − signs denote either right-handed or left-handed quartz, respectively, and E designates the photon energy in electron-volt units.

To determine the polarizer offset and analyzer phase angles for a rotating analyzer ellipsometer with elliptically polarizing elements, one starts as usual by defining P_1 as the polarizer angle near $P = P_S$ at which the minimum occurs in the parabolic best fit to the measured residual function. Furthermore, A_1 is again defined by evaluating the linear best fit to the phase function, $\{\tan^{-1}(\beta'/\alpha')\}/2$, at P_1. With these two definitions, the following expressions are valid to first order in the polarizer and analyzer ellipticity parameters:

$$P_S = P_1 - \frac{\gamma_A \tan\psi_0 + \gamma_P \cos\Delta_0}{\sin\Delta_0}, \tag{5.98a}$$

$$A_S = A_1 - \frac{\gamma_P \cot\psi_0 + \gamma_A \cos\Delta_0}{\sin\Delta_0}. \tag{5.98b}$$

The subscripts on the ellipsometry angles indicate zero-order approximations obtained by neglecting the polarizer and analyzer ellipticity corrections (i.e., with $\gamma_A = \gamma_P = 0$). In Eqs. (5.98), the corrections become large as Δ_0 approaches 0° or 180°, where the overall calibration procedure fails as well. Eqs. (5.98) are equivalent in form to Eqs. (5.93) but with γ_A replaced by $B_{W2}/2$ and γ_P replaced by $B_{W1}/2$.

For the quartz Rochon polarizer and analyzer, the signs of the ellipticity parameters, γ_P and γ_A, will be unknown in general and must be determined before the corrections of Eq. (5.98) can be applied. In fact, to determine γ_P and γ_A explicitly, new values P_2 and A_2 can be obtained experimentally in the same way as P_1 and A_1, but when the residual and phase function calibrations are performed instead for P near $P_S + (\pi/2)$. Then γ_P and γ_A can be expressed solely in terms of the known calibration quantities and (ψ_0, Δ_0):[46]

$$\gamma_P = \frac{(P_1 - P_2)\cos\Delta_0 - (A_1 - A_2)\csc 2\psi_0}{D_1}, \tag{5.99a}$$

$$\gamma_A = \frac{(A_1 - A_2)\cos\Delta_0 - (P_1 - P_2)\csc 2\psi_0}{D_1}, \tag{5.99b}$$

$$D_1 = 2\frac{\cos^2\Delta_0 - \csc^2 2\psi_0}{\sin\Delta_0}. \tag{5.99c}$$

Thus, by performing the calibration for P near P_S and $P_S + (\pi/2)$ at different photon energies, the relationship of Eq. (5.97) can be verified. Experimental results for such calibrations will be presented later in this section.

The effects of ellipticity on the polarizer offset and analyzer phase angle obtained by the zone-difference phase function method are given to first order by:[20]

$$P_S = P_0 + \frac{\gamma_P \tan\Delta_0}{\cos 2\psi_0}, \tag{5.100a}$$

$$A_S = A_0 + \gamma_P \sin\Delta_0 \tan 2\psi_0. \tag{5.100b}$$

Here P_0 is the P-axis intercept of the linear fit to the experimentally determined zone-difference phase function, and A_0 is obtained by evaluating the linear fit to the corresponding phase function at P_0. Thus, the terms associated with ellipticity of the analyzer are eliminated in first order by the zone average. Note that within the (ψ_0, Δ_0) regimes for which this calibration procedure is preferred over the residual function procedure, the first order approximation remains valid. Again, the present treatment on imperfect optical elements has neglected the effects of the transfer characteristics of the detection circuitry. As before, the gain ratio η is determined from the residual function minima and the electronic phase shift $\phi/2$ is included as a component of each of the analyzer angles. Finally, symmetric sets of equations for both the residual and zone-difference phase function are obtained for the P_rSA configuration. For application to the rotating polarizer system, the quantities P_S, A_S, P_1, A_1, γ_A, γ_P, P_2, A_2, P_0, and A_0 in Eqs. (5.98)–(5.100) must be replaced by A_S, P_S, A_1, P_1, $-\gamma_P$, $-\gamma_A$, A_2, P_2, A_0, and P_0, respectively.

As an example, Fig. 5.8 shows the results of (a) residual/phase function and (zone-difference-phase)/phase function calibrations for a PSA_r instrument with a quartz Rochon polarizer and analyzer.[21] Three different reflecting surfaces are used in the calibrations: thin film polycrystalline gold, bulk single crystal silicon (c-Si), and thin film amorphous silicon (a-Si). A photon energy of 3.5 eV was used for the Au and c-Si in the residual function calibration, and 2.75 eV and 3.5 eV were used for the c-Si and a-Si, respectively, in the zone-difference phase function calibration. All samples were opaque at these energies. The broken lines label the positions of the residual function minima P_1 and the zone-difference phase function intercepts P_0, and have been extended upward to intersect the phase functions. These values must be corrected for the effects of polarizer and analyzer ellipticity according to Eqs. (5.98) and (5.100). The ellipticity parameters (γ_P, γ_A) are obtained by first performing residual/phase function calibrations near $P = P_S$ and $P = P_S + (\pi/2)$ for a Au surface, then applying Eqs. (5.99) to the resulting calibration values (P_1, A_1) and (P_2, A_2). In fact, the values of (γ_P, γ_A) are given in Fig. 5.9 for calibrations performed at seven photon energies between 2 and 4.5 eV, and the broken

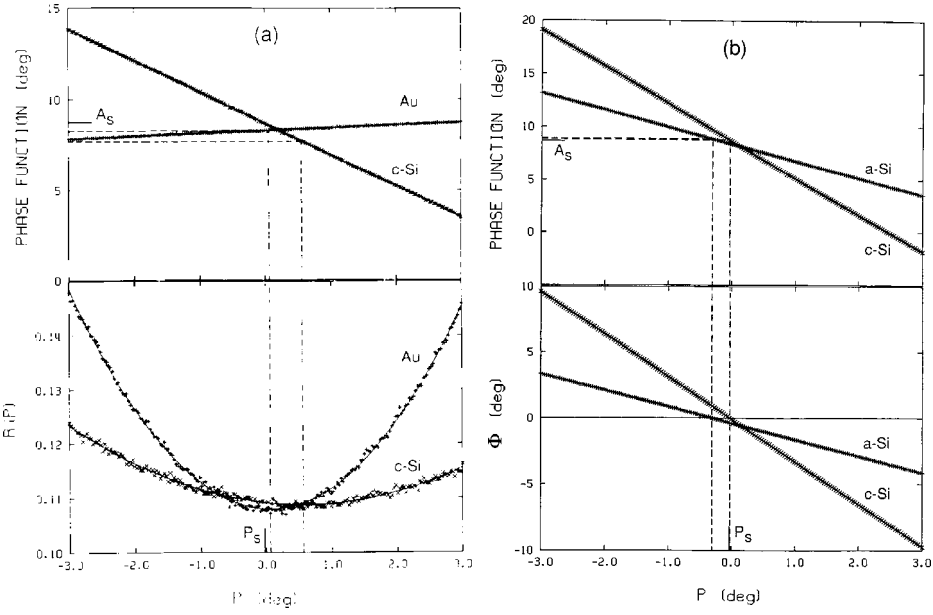


Figure 5.8 Calibration data for gold, crystalline silicon (c-Si), and amorphous silicon (a-Si) samples obtained by applying the (a) residual function and (b) zone-difference phase function methods with a PSA_r ellipsometer having quartz Rochon elements. The broken lines show the positions of the minima in the residual function, $R(P)$, in (a) and the P-axis intercepts of the zone-difference phase function, $\Phi(P)$, in (b). The solid segments show the values of P_S and A_S obtained after correction for the effects of optical activity in the polarizer and analyzer. The photon energy was 3.5 eV in (a) and 2.75 eV and 3.5 eV in (b) for c-Si and a-Si, respectively. [After Ref. 21.]

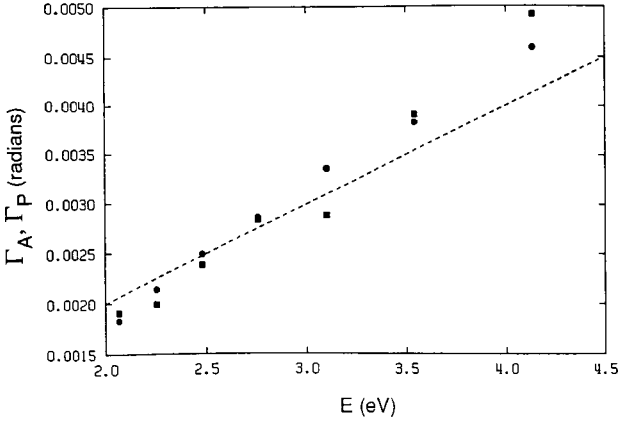


Figure 5.9 Optical activity coefficients, γ_P (circles) and γ_A (squares), for the quartz Rochon polarizer and analyzer elements in a PSA_r ellipsometer obtained from residual and phase function calibrations near $P = P_S$ and $P = P_S + (\pi/2)$. The broken line is determined according to Eq. (5.97). [After Ref. 21.]

line is the prediction of Eq. (5.97). The solid line segments along the ordinate and abscissa of Fig. 5.8 label the positions of the proper calibration angles, P_S and A_S, obtained after correcting the minima and intercepts for the polarizer and analyzer ellipticity. Upon applying these corrections, the values of P_S and A_S for three of the four calibrations are brought into agreement to within 0.02° and 0.01°, respectively. The discrepancy in P_S and A_S for the fourth calibration, c-Si by the residual method, is somewhat larger (within 0.1°) owing to the uncertainty in the position of the broad minimum in the residual function.

When two types of optical errors are incorporated into the calibration description, the corrections to the observables P_1 and A_1 are additive in a first-order analysis. Applying such considerations to the rotating analyzer ellipsometer with elliptic polarizers and imperfect reactor windows is straightforward using the results of Eqs. (5.93) and (5.98). The window imperfection parameters can be determined if the ellipticity parameters are known. Thus, Eqs. (5.94) can now be written:

$$B_{W1} = (A_1 - A_2)\tan 2\psi_0 \sin\Delta_0 - 2\gamma_P \sec 2\psi_0 - 2\gamma_A \tan 2\psi_0 \cos\Delta_0, \quad (5.101a)$$

$$B_{W2} = -(P_1 - P_2)\tan 2\psi_0 \sin\Delta_0 + 2\gamma_A \sec 2\psi_0 + 2\gamma_P \tan 2\psi_0 \cos\Delta_0. \quad (5.101b)$$

Again, by time-reversal symmetry, Eqs. (5.101) become valid for the $P_rW_1SW_2A$ system when P_1, P_2, A_1, A_2, B_{W1}, B_{W2}, γ_A, and γ_P are replaced by A_1, A_2, P_1, P_2, B_{W2}, B_{W1}, $-\gamma_P$, and $-\gamma_A$, respectively.

A third error to be considered within the framework of the small-angle calibration is polarization sensitivity of the detection system.[42] As noted earlier, this is a problem that occurs in the PSA_r and $PCSA_r$ ellipsometer configurations in which case the rotating analyzer is directly in front of the detector. The effect can arise from stress-induced birefringence of the input window of the detector, typically a photomultiplier tube (PMT). It can be shown, however, that P_S and A_S determined in calibration are not influenced in first order by the waveplate parameters A_{WD} and B_{WD} that describe the Mueller matrix of a birefringent detector window. One can extend beyond the waveplate description of the detector window, and employ instead the general Mueller matrix elements of the entire detector in correcting the calibration angles. For this error, more detail will be provided since it is a commonly encountered unavoidable error in scanning PSA_r systems. In the discussion below, the effects of this error on the irradiance calibration will also be provided.

For the PSA_r ellipsometer configuration, the effect of polarization sensitivity on the calibration parameters obtained by the residual and phase function approach is described by the expressions:[21]

$$P_S = P_1 - \frac{m'_{D13} \tan\psi_0 \cos\Delta_0}{4\sin^2\Delta_0}, \tag{5.102a}$$

$$A_S = A_1 - m'_{D13} \frac{1 + \frac{\cot^2\Delta_0}{2}}{2}, \tag{5.102b}$$

where m'_{D13} is the (1,3) element of the normalized Mueller matrix of the detector, defined in Eqs. (5.41) and (5.45). As usual, P_1 is the polarizer angle defining the minimum in the residual function fit and A_1 is the phase function fit evaluated at P_1, both measured in calibration for P near P_S. Again, ψ_0 and Δ_0 are the ellipsometry parameters obtained in zero order. The corresponding equations for P near $P_S + (\pi/2)$ are given by:

$$P_S = P_2 + \frac{m'_{D13} \cot\psi_0 \cos\Delta_0}{4\sin^2\Delta_0}, \tag{5.102c}$$

$$A_S = A_2 + m'_{D13} \frac{1 + \frac{\cot^2\Delta_0}{2}}{2}. \tag{5.102d}$$

If P_S and A_S are eliminated from the previous two pairs of equations, then m'_{D13} is in fact overdetermined:

$$m'_{D13} = \frac{2\sin^2\Delta_0 \sin 2\psi_0}{\cos\Delta_0}(P_1 - P_2), \tag{5.103a}$$

$$m'_{D13} = \frac{1}{1 + \frac{\cot^2\Delta_0}{2}}(A_1 - A_2). \tag{5.103b}$$

The normalized Mueller matrix element m'_{D12} of the detector is also required, however, in order to characterize polarization sensitivity completely and to calculate the exact ψ and Δ values including this effect [see Eqs. (5.43e) and (5.43f) above]. In the presence of the previously described errors, including imperfect windows and an elliptical polarizer and analyzer, the minima $R(P_1)$ and $R(P_2 + \pi/2)$ of the fit to the measured residual function R are equal (at least to first order in the error parameters) and provide η according to Eq. (5.77d), i.e.,

$$\eta = \frac{1}{\sqrt{1 - R(P_1)}} = \frac{1}{\sqrt{1 - R(P_2 + \pi/2)}}.$$

Polarization sensitivity introduces a difference in these minima from which m'_{D12} can be calculated:[21]

$$\eta = \frac{1}{\sqrt{1 - \dfrac{R(P_1) + R(P_2 + \pi/2)}{2}}}, \tag{5.104a}$$

$$m'_{D12} = \eta^2 \frac{R(P_2 + \pi/2) - R(P_1)}{2}. \tag{5.104b}$$

When a quartz Rochon polarizer and analyzer are used, the ellipticity corrections to P_1, A_1, P_2, and A_2, given to first order in γ_P and γ_A, add to the polarization sensitivity corrections, given to first order in m'_{D13}. Then a two-zone calibration giving four equations is insufficient to determine the five unknowns P_S, A_S, γ_P, γ_A, and m'_{D13}. If the ellipticity corrections are larger, then an analysis neglecting polarization sensitivity can at least provide the handedness of the quartz elements. Then Eq. (5.97) can be employed with the appropriate signs to eliminate two of the five unknowns and determine m'_{D13}. Alternatively, in theory it is also possible to obtain m'_{D13} by applying Eqs. (5.104) with m'_{D12} replaced by m'_{D13} after having rotated the detector by $\pi/4$ radians.

The polarization sensitivity correction to the zone difference phase function, performed when $|\Delta|$ is near 0 or 180°, yields the expressions:[21]

$$P_S = P_0 + \frac{m'_{D13} \tan 2\psi_0}{2\cos \Delta_0}, \tag{5.105a}$$

$$A_S = A_0 + \frac{m'_{D13}}{2\cos 2\psi_0}, \tag{5.105b}$$

where P_0 is the P-axis intercept of the best linear fit to Φ [see Eq. (5.84a)] and A_0 is the best fit to the phase function evaluated at P_0.

The corresponding corrections to the irradiance function calibration for P near P_S yields the expressions:[53]

$$P_S = P_{I,1} + \frac{m'_{D13} \tan 2\psi_0 \cos \Delta_0}{4}, \tag{5.106a}$$

$$A_S = A_{I,1} - m'_{D13} \frac{1 - \dfrac{\cos^2 \psi_0 \cos^2 \Delta_0}{\cos 2\psi_0}}{2}. \tag{5.106b}$$

Here, $P_{I,1}$ is the polarizer angle defining the extremum near P_S in the fit to the irradiance function [see Eqs. (5.85)] and $A_{I,1}$ is the phase function fit

evaluated at $P_{I,1}$. In contrast, for P near $P_S + \pi/2$, the following expressions apply:

$$P_S = P_{I,2} + \frac{m'_{D13} \tan 2\psi_0 \cos \Delta_0}{4}, \tag{5.106c}$$

$$A_S = A_{I,2} + m'_{D13} \frac{1 + \dfrac{\sin^2\psi_0 \cos^2\Delta_0}{\cos 2\psi_0}}{2}. \tag{5.106d}$$

In principle, P_S and A_S can be eliminated from Eqs. (5.106) to obtain two different expressions for m'_{D13} in a procedure similar to that using the two zones of the residual and phase function calibration [see Eqs. (5.103)]. The simplest such expression is derived from Eqs. (5.106a) and (5.106c) and is given by:

$$m'_{D13} = \frac{P_{I,1} - P_{I,2}}{1 - \dfrac{1}{2}\cos^2\Delta_0}. \tag{5.107}$$

This approach is useful for determining m'_{D13} when Δ is near 0° or 180° in which case the two zone residual/phase function approaches based on Eqs. (5.103) fail.

The equivalent problem that plagues measurements in the P_rSA configuration is residual polarization of the light source, which is most often a Xe lamp in spectroscopic applications.[41,53] Ellipsometers in the P_rSA configuration that incorporate a quartz Rochon polarizer as the rotating element often show oscillations as a function of photon energy that arise from the coupling of residual source polarization errors with polarizer ellipticity errors. By time-reversal symmetry, Eqs. (5.102) through (5.107) for the correction of polarization sensitivity in the calibration of PSA_r ellipsometers are also applicable for correction of residual source polarization in the calibration of P_rSA ellipsometers, with an interchange of all polarizer and analyzer azimuths. In other words, P_S, A_S, P_1, A_1, P_2, A_2, P_0, A_0, $P_{I,1}$, $A_{I,1}$, $P_{I,2}$, and $A_{I,2}$, in Eqs. (5.103) through (5.107) must be replaced by A_S, P_S, A_1, P_1, A_2, P_2, A_0, P_0, $A_{I,1}$, $P_{I,1}$, $A_{I,2}$, and $P_{I,2}$, respectively. In addition, the Mueller matrix elements m'_{D12} and m'_{D13} are replaced by the components s_{2s} and s_{3s} first defined above in Eq. (5.63). In fact, these components derive from the Stokes vector of the beam entering the polarizer, $\mathbf{s}_S = I_S[1 \;\; s_{2s} \;\; s_{3s} \;\; s_{4s}]^T$, expressed in the p-s frame of reference of the sample. Because of the difficulty of such corrections and their variation with time as the lamp ages, it is helpful to use depolarizing optics, for example opal glass or a sufficient length of optical fiber, after the source to reduce the necessary corrections.

Calibration of a rotating analyzer ellipsometer in the presence of polarization sensitive detection will be illustrated through the large angle calibration procedure.[66] In this procedure, the experimental Fourier coefficients (α', β') measured at a selected photon energy as a function of the angular setting P of the fixed polarizer are fit using the following series of equations:

$$\alpha' = \frac{\alpha \cos 2A_S - \beta \sin 2A_S}{\eta}, \tag{5.108a}$$

$$\beta' = \frac{\alpha \sin 2A_S + \beta \cos 2A_S}{\eta}, \tag{5.108b}$$

$$\alpha = \frac{s_{2r} + m'_{D12}}{1 + (1/2)(m'_{D12}s_{2r} + m'_{D13}s_{3r})}, \tag{5.108c}$$

$$\beta = \frac{s_{3r} + m'_{D13}}{1 + (1/2)(m'_{D12}s_{2r} + m'_{D13}s_{3r})}, \tag{5.108d}$$

$$s_{2r} = \frac{\cos 2P' - \cos 2\psi}{1 - \cos 2P' \cos 2\psi}, \tag{5.108e}$$

$$s_{3r} = \frac{\sin 2P' \sin 2\psi \cos \Delta}{1 - \cos 2P' \cos 2\psi}. \tag{5.108f}$$

These equations are either reproduced or derived from Eqs. (5.11), (5.18), (5.20), and (5.43), and the contribution of the electronic phase shift to Eqs. (5.108a) and (5.108b) has been ignored without loss of generality. In the fitting procedure, there are seven free parameters, in order of appearance in Eqs. (5.108): η, A_S, m'_{D12}, m'_{D13}, P_S, ψ, and Δ.

Figure 5.10 shows the experimental results for $\{\alpha'(P), \beta'(P)\}$ (open symbols) over the range $-90° \leq P \leq 90°$ along with the best fits (solid lines) for a sample consisting of a 1615 Å thermally grown SiO_2 layer on a crystalline silicon wafer illuminated at 3.5 eV.[66] The following parameters were obtained in the fit: $\eta = 1.0583 \pm 0.0007$, $A_S = 1.717° \pm 0.054°$, $m'_{D12} = (-2.26 \pm 0.84) \times 10^{-3}$, $m'_{D13} = (2.74 \pm 1.24) \times 10^{-3}$, $P_S = -1.826° \pm 0.055°$, $\psi = 27.39° \lesseqgtr 0.04°$, and $\Delta = 142.57° \pm 0.06°$. Also shown in Fig. 5.10 is the residual function $R(P) = 1 - \{[\alpha'(P)]^2 + [\beta'(P)]^2\}$ along with its best fit over the full range of P. Parabolic minima are evident in this function for P near P_S and $P_S + \pi/2$ that can be used in the small angle calibrations as described above. The same calibration and fitting procedure as in Fig. 5.9 can be performed at several different photon energies, and the results for the fitting parameters versus energy are shown in Figs. 5.11 through 5.13.

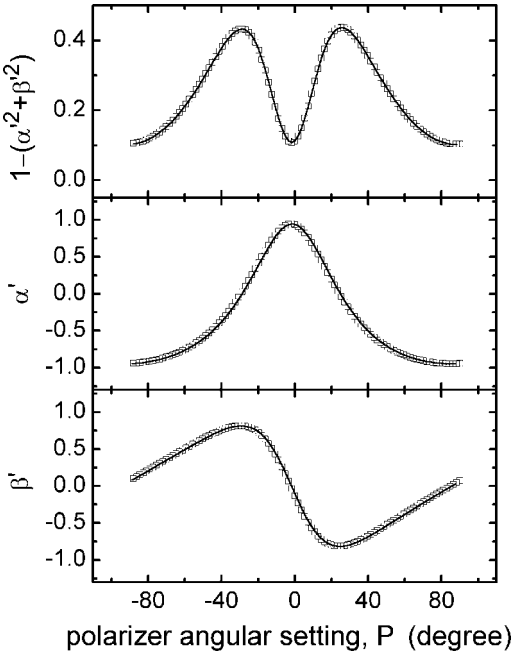


Figure 5.10 Experimental results for $\{\alpha'(P), \beta'(P)\}$ (open symbols) over the range $-90° \leq P \leq 90°$ along with the best fit (solid line) for a sample consisting of a 1615 Å thermally grown SiO_2 layer on a crystalline silicon wafer illuminated at 3.5 eV. The residual function $R(P) = 1 - \{[\alpha'(P)]^2 + [\beta'(P)]^2\}$ is shown in the top panel along with the simulation over the full range of P. [After Ref. 66.]

In Fig. 5.11, the calibration angles are given and average values of $P_S = -1.829 \pm 0.036°$ and $A_S = 1.715 \pm 0.025°$ have been obtained, where the standard deviations are entered as the confidence limits.[66] The top panel of Fig. 5.12 shows the detection system Mueller matrix elements m'_{D12} and m'_{D13}. These two error parameters exhibit strong variations with photon energy and are clearly parallel with an offset of 0.005. In the bottom panel of Fig. 5.12, η is presented as a function of the magnitude of the cathode voltage of the PMT. A consistent and reproducible trend is observed apparently because of the coupling between the PMT signal processing electronics and the preamplifier circuitry. Such an effect may also explain fluctuations in A_S versus photon energy since a possible variation in the electronic phase shift ϕ in Eq. (5.11) with PMT voltage has been ignored. In fact, it may be reasonable to test for a PMT voltage dependent contribution to A_S, and possibly include this effect in the calibration analysis. Finally, Fig. 5.13 shows the excellent agreement between the (ψ, Δ)

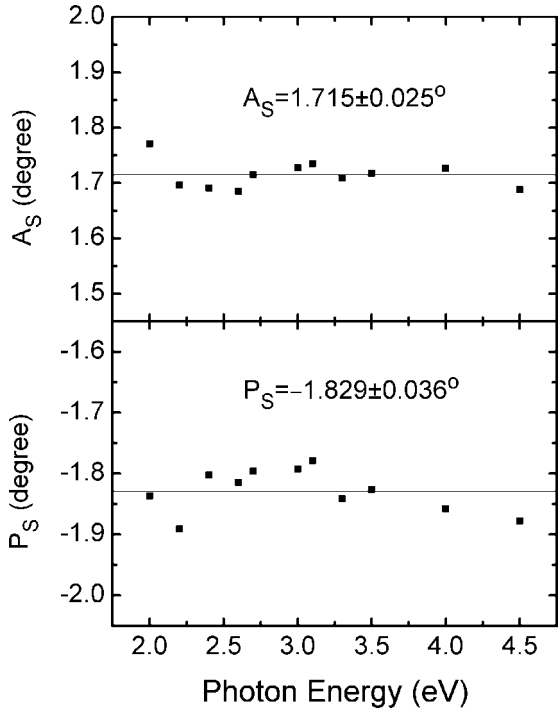


Figure 5.11 Polarizer and analyzer calibration angles obtained in the best fits to calibration data sets at different photon energies, including the one depicted in Fig. 5.10 at 3.5 eV, for a sample consisting of a 1615 Å thermally grown SiO_2 layer on a crystalline silicon wafer. Average values of $A_S = 1.715 \pm 0.025°$ and $P_S = -1.829 \pm 0.036°$ have been determined, whereby the confidence limits indicate the spectral standard deviations. [After Ref. 66.]

values obtained in calibration with those obtained in the measurement. These results suggest an alternative approach for calibration that reduces the number of parameters required to fit the (α', β') variations in Fig. 5.9 from seven to five. This would be expected to narrow the confidence limits. An initial calibration can be performed in which the polarization sensitivity of the detection system is neglected. Measurement of the sample with these calibration values provides the zero-order values (ψ_0, Δ_0) that can be fixed in a refined calibration procedure that provides the other five parameters.

This section will conclude with a brief discussion of the effect of detector nonlinearity on calibration and data reduction. It has already been noted that weak nonlinearities inherent in PMT detectors result from the non-zero resistances along the dynode resistor chain.[54] To

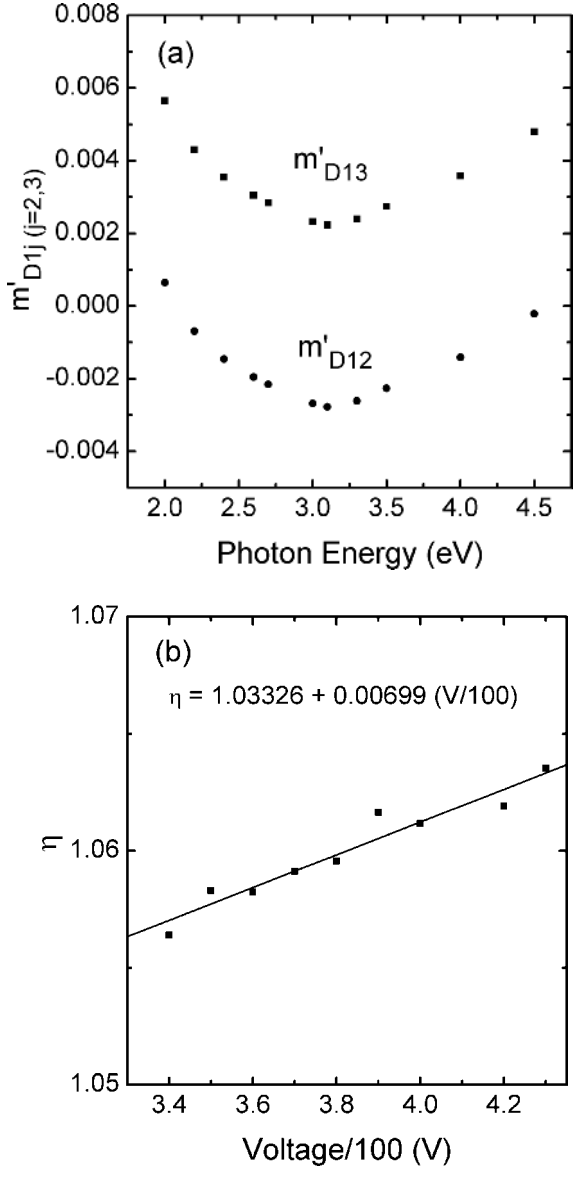


Figure 5.12 (a) The detection system Mueller matrix elements m'_{D12} and m'_{D13} as a function of photon energy that characterize the polarization sensitivity of the detection system and (b) the gain ratio η as a function of the magnitude of the PMT cathode voltage. Both sets of results were obtained in the best fits to calibration data sets at different photon energies, including the one at 3.5 eV depicted in Fig. 5.10, for a sample consisting of a 1615 Å thermally grown SiO_2 layer on a crystalline silicon wafer. [After Ref. 66.]

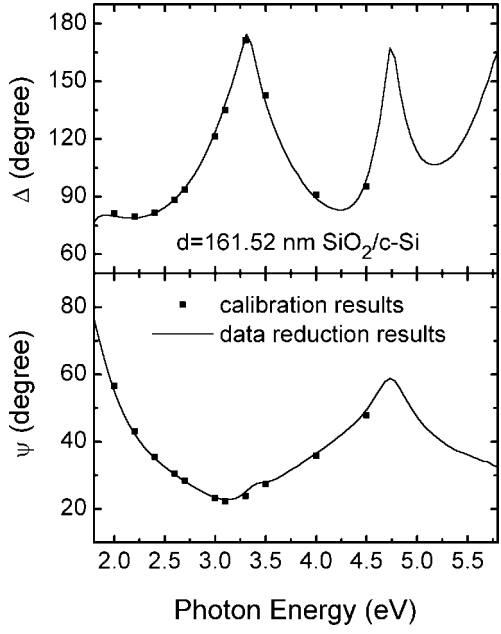


Figure 5.13 A comparison of the continuous (ψ, Δ) spectrum for crystalline silicon coated with a 1615 Å thick layer of thermally grown oxide (solid line) with the discrete values obtained in the best fits to calibration data sets at different photon energies, including the one at 3.5 eV depicted in Fig. 5.10. [After Ref. 66.]

assess the effect of nonlinearity on the calibration parameters, it is first assumed that the PMT output signal at the input of the processing circuitry can be expressed as:

$$I_{\text{exp}}(t) \propto I_{\text{th}}(t) + \sum_{m=2}^{\infty} k_m [I_{\text{th}}(t)]^m, \tag{5.109}$$

where $I_{\text{th}}(t)$ is the irradiance at the detector predicted theoretically, assuming no optical system errors.[42,67] Thus far, all equations connecting experiment to theory have employed the coefficients, (α', β') measured using an assumed linear detector. Equations can be derived that express these coefficients in terms of the corresponding ones, α'_{eff} and β'_{eff}, measured with the nonlinear detector. If only the first term is considered in the series in Eq. (5.109) and only first order terms in k_2 are retained, then:

$$\alpha' = \alpha'_{\text{eff}} \left[1 - k_2 \frac{2 - (\alpha'^{\,2}_{\text{eff}} + \beta'^{\,2}_{\text{eff}})}{2} \right], \tag{5.110}$$

$$\beta' = \beta'_{\text{eff}}\left[1 - k_2\frac{2 - (\alpha'^{\,2}_{\text{eff}} + \beta'^{\,2}_{\text{eff}})}{2}\right], \tag{5.111}$$

for the P(C)SA$_r$ or P$_r$(C)SA systems. The corrected coefficients α' and β' can be used in all previously derived equations. As might be expected, detector nonlinearity of this form does not influence the determination of the calibration angles P_S and A_S by the residual, phase, and zone difference phase function methods. However, nonlinearity does generate an additional contribution to the zero offset of the residual function minima occurring near 0 and $\pi/2$:

$$R(P_1) = R(P_2 + \pi/2) = 1 - \eta^{-2}\{1 + k_2(2 - \eta^{-2})\}, \tag{5.112}$$

valid to first order in k_2. As expected, there is no way to separate the effects of detector nonlinearity and signal processing circuit characteristics, at least within the calibration procedures described so far. For this reason detector nonlinearity is a problem that is best corrected at the source rather than at the calibration and data reduction stages.

5.6.3 Detecting and Correcting Compensator Errors

As demonstrated in Section 5.6.2, both small angle and large angle calibration methods can be effective in extracting ellipsometer error parameters simultaneously with calibration data. Figures 5.8 and 5.9 have demonstrated the application of the small angle calibration method to extract P_S and A_S and the ellipticity parameters γ_A and γ_P of the elliptical polarizer and analyzer versus photon energy for an ellipsometer in the PSA$_r$ configuration. In contrast, Figs. 5.10 through 5.13 have demonstrated the application of the large angle calibration method to extract P_S and A_S and the Mueller matrix parameters m'_{D12} and m'_{D13} of the detector versus photon energy for an ellipsometer in the PSA$_r$ configuration. The large angle method applies the same equations as those developed in Section 5.4, and so requires no further explanation for the determination of compensator errors. The small angle method, however, requires additional derivations, namely second-order expansions versus the fixed polarizer, compensator, or analyzer angles in the vicinity of the p and s directions. As a result, a final example of this method will be described for an ellipsometer in the PCSA$_r$ configuration in which the polarizer, compensator, and analyzer are all elliptic in nature so that the normal transmission modes are elliptical polarization states with small ellipticities γ_j; j = P, C, A.[46]

The three required calibration steps of this procedure and the final outcome will be described. Again the fixed elements are assumed to be mounted such that the scale readings are near 0° when the transmission and fast axes are in the plane of incidence.

(1) First the system is aligned for straight-through operation without the sample or compensator (which is to be used at the scale reading C). The relative scale shift, $A_S - P_S$, between the polarizer and rotating analyzer is given by:

$$A_{P0} \equiv A_S - P_S = \frac{\tan^{-1}(\beta'/\alpha') + \pi\, \mathrm{u}(-\alpha')\mathrm{sgn}(\beta')}{2} - P, \quad (5.113)$$

where $\mathrm{u}(x) = 0$ for $x < 0$ and $\mathrm{u}(x) = 1$ for $x \geq 0$ and $\mathrm{sgn}(x) = -1$ for $x < 0$ and $\mathrm{sgn}(x) = 1$ for $x \geq 0$. The best linear fit to measurements of A_{P0} is evaluated for P near $P_C = C + (P_S - C_S)$ and for P near $P_C + \pi/2$, labelled A_{P01} and A_{P02}, respectively. In theory, the slopes of the best fit linear relationships should vanish and A_{P01} and A_{P02} should be identical. In practice, however, various optical and electronic errors cause weak variations with P and small differences in A_{P01} and A_{P02}.

(2) In the next step, the compensator is added to the system, and the residual function $R(P)$ is determined in the vicinity of its minima versus P near $P = P_C$ and $P = P_C + \pi/2$. Fits to the resulting data provide the polarizer readings at the minima in $R(P)$, and these readings are designated $P = P_{C1}$ and $P = P_{C2} + \pi/2$, respectively. The following phase functions are evaluated at these minima:

$$P_{C3} \equiv \left.\frac{\tan^{-1}(\beta'/\alpha')}{2}\right|_{P=P_{C1}} - A_{P01}, \quad (5.114a)$$

$$P_{C4} \equiv \left.\frac{\tan^{-1}(\beta'/\alpha')}{2}\right|_{P=P_{C2}+\pi/2} - A_{P02}. \quad (5.114b)$$

The compensator calibration angle and optical activity coefficients to first order are given by:

$$C + (P_S - C_S) = (P_{C1} + P_{C2} + P_{C3} + P_{C4})/4 \quad (5.115a)$$

$$\gamma_C = \{\cot(\delta_C/2)\}\{P_{C1} + P_{C2} - P_{C3} - P_{C4}\}/4 \quad (5.115b)$$

$$\gamma_A = \{(P_{C4} - P_{C3})\cot\delta_C - (P_{C2} - P_{C1})\csc\delta_C\}/2 \quad (5.115c)$$

$$\gamma_P = \{(P_{C2} - P_{C1})\cot\delta_C - (P_{C4} - P_{C3})\csc\delta_C\}/2 \quad (5.115d)$$

where γ_C is the ellipticity parameter for the compensator, using a Mueller matrix of the form of Eq. (5.31).

(3) Finally, the angle of incidence is set, the sample is mounted and aligned, the compensator is removed temporarily, and the calibration described by Eqs. (5.98) for the PSA_r is performed to determine P_S and A_S. Then, the compensator is returned to system and its calibration angle, C_S, is given by Eq. (5.115a). Determination of the ellipsometric angles (ψ, Δ) from the measurement using the fixed parameters P, C, P_S, C_S, A_S, γ_P, γ_C, γ_A, and δ_C is described above [see Eqs. (5.11), (5.17), (5.19), (5.20), and (5.32)].

5.7 Summary: Recent and Future Directions

Automatic rotating polarizer and analyzer ellipsometers in the P_rSA and PSA_r configurations were first developed over the time period from about 1965 to 1980, and so the basic foundations of this chapter have been established for many years. For example, the descriptions of instrument design, precision, and calibration in Sections 5.3, 5.5, and 5.6, respectively, derive for the most part from publications of the 1970s. (In contrast, the comprehensive description of data reduction in Section 5.4 based on Mueller matrices is new to this review.) The simplicity of the design of the automatic rotating analyzer ellipsometers, as first reported by Cahan and Spanier[10] in 1969 and later by Aspnes[68] in 1973, accounted for their popularity in succeeding years. The first automatic spectroscopic instruments were applied for the characterization of the optical properties of strongly absorbing solids and thin films. In fact, the most widely used dielectric function set for semiconductors, spanning a wide photon energy range above their lowest direct gaps, was obtained with a simple rotating analyzer ellipsometer operating without a compensator.[28]

Over time, however, the limitations of this instrument design were addressed with numerous modifications while retaining the simple photometric measurement principle based on waveform analysis of the detector irradiance. Recent advances have involved:

(i) incorporation of improved compensators;
(ii) development of variable angle of incidence and generalized spectroscopic ellipsometry;
(iii) extensions to rotating compensator and dual rotating compensator configurations:
(iv) increases in spectral range;
(v) development of microellipsometry and imaging ellipsometry; and
(vi) improvement of spectroscopic ellipsometry for real time applications.

The basic rotating analyzer ellipsometer has established the foundation for these advances, and improvements in all these areas of instrumentation development are expected in the future.

In the rotating polarizer and analyzer P_rSA and PSA_r configurations, a compensator must be inserted for accurate characterization of weakly absorbing thin-film/substrate systems when linearly (or nearly linearly) polarized light is detected in reflection from the sample, i.e., when Δ is near 0° or ±180°. This requirement has been the driving force for the development of many useful optical elements in ellipsometry. One of the first descriptions of the calibration of the rotating analyzer spectroscopic ellipsometer considered the possible incorporation of a traditional quartz Babinet-Soleil compensator.[46] Due to the optical activity of quartz, this device is essentially a weakly elliptic compensator. No demonstrations of the experimental application of the Babinet-Soleil were reported. Difficulties arise in part due to spectroscopic artifacts, for example, high frequency oscillations versus wavelength due to misalignment of the fast and slow axes of two of the plates.[26] Greater success has been achieved more recently with an achromatic compensator, consisting of a vitreous silica three-reflection rhomb; however, this element suffers from a lack of adjustability.[38] Thus, in rotating polarizer and analyzer ellipsometry, the low sensitivity regions in Δ are simply shifted from near (0°, ±180°) to near ±90°. State-of-the-art commercial rotating analyzer ellipsometers employ adjustable Berek (or tilting plate) compensators for maximum versatility in the measurement of reflecting surfaces with different properties.[25] Future improvements in adjustable compensator design and calibration are expected. In addition, continuously rotatable achromatic compensators are needed for highest sensitivity in rotating compensator ellipsometry, a more advanced version of rotating element ellipsometry, as described in greater detail below.

Rotating analyzer spectroscopic ellipsometers of exceptional power and versatility are now available that combine the features of an automatic compensator, a variable angle of incidence, and a variable incident polarization state.[25] First, adjustment of the settings of the fixed polarizer and automatic compensator as the wavelength is varied during sample measurement ensures operation at the optimum instrument precision for all wavelengths. Second, collection and analysis of multiple (ψ, Δ) spectra at different angles of incidence can be used to narrow the confidence limits in the determination of thin film structural parameters and optical properties. Third, characterization of anisotropic sample structures with a rotating analyzer ellipsometer is possible by performing measurements with different incident polarization states in a procedure called generalized ellipsometry.[69] Generalized ellipsometry is also possible in the rotating polarizer configuration by performing measurements with different fixed

compensator and analyzer settings. In this general area, extensive recent progress has been made, as described elsewhere in the Handbook, and further progress is expected in the future in terms of advanced data acquisition modes and data reduction/analysis procedures.

Rotating compensator ellipsometer configurations provide pathways alternative to generalized ellipsometry that also enhance data acquisition and analysis capabilities for complex sample systems, including those exhibiting inhomogeneity and anisotropy.[70–72] Once electromechanical and optical systems for ellipsometry in the P_rSA and PSA_r configurations have been established, they can be adapted readily to more powerful configurations through the incorporation of one or two rotating compensators. For example, the incorporation of a single rotating compensator after the sample in the PSC_rA configuration allows one to determine all four components of the Stokes vector of the light beam reflected from the sample surface.[53,70] Methodologies have been demonstrated that use the reflected irradiance and degree of polarization to characterize sample heterogeneities and non-uniformities over a wide range of scales. By incorporating two compensators in the $PC_{1r}SC_{2r}A$ configuration, the full 16-element Mueller matrix of the sample can be determined for specific ratios of the rotation frequency of the two compensators, e.g., for ratios n:5 where $n = 1, \ldots, 4$.[72] Establishing the full Mueller matrix allows one to extract information on the properties of samples that are both inhomogeneous and anisotropic. Analysis procedures for such complex samples are in an early stage of development and significant advances are expected in the future. In fact, the most difficult instrumentation issue to be addressed in converting from a rotating polarizer/analyzer to a rotating compensator configuration is the development of a rotatable compensator element that is free of spectroscopic artifacts. Furthermore, as noted earlier in this section, such an instrument operates best near the quarter-wave point of the compensator; thus, a rotating achromatic compensator is ideal. This is another subject of recent activity directed toward the development of powerful new instruments.

Rotating polarizer and analyzer designs have been used widely in extensions of spectroscopic ellipsometry into the vacuum ultraviolet[73,74] and infrared[75] ranges. In fact, the well-established operational principles of the conventional rotating polarizer and analyzer spectroscopic ellipsometers for the near-infrared to ultraviolet range can be readily extended into the vacuum ultraviolet. Because the polarizer, compensator, and analyzer can all be fabricated from MgF_2, which provides the required spectral range, the remaining concerns then become the selection of sources and detection systems from commercial vendors of ultraviolet spectrometers, as well as the elimination of absorption along the beam path with N_2 purging. In contrast, infrared ellipsometers based on Fourier transform

spectral scanning and rotating element polarization detection differ from their conventional counterparts for the near-infrared to ultraviolet range. In the conventional instruments, of course, the optical elements are rotated continuously at a number of wavelength settings. For the Fourier transform infrared ellipsometers, however, the frequency of wavelength scanning is comparable with the frequency of the rotating polarizer or analyzer (both 10–100 Hz). As a result, the waveform at the detector is simply reconstructed from wave number scans performed at a number (three or more) of polarizer/analyzer (or compensator, if used) angular settings.[76] The overall simplicity of the rotating polarizer/analyzer vacuum ultraviolet and infrared ellipsometers account for their rapid recent development and subsequent popularity. In contrast, phase modulation instruments, when operated in the infrared spectral range, employ more complicated double-modulation techniques for data acquisition.[77]

Spatially resolved ellipsometry, including the techniques of microellipsometry with spectroscopic capability and imaging ellipsometry at a single wavelength, have also relied on the simplicity of rotating polarizer and analyzer principles.[78–81] In fact in microellipsometry, the rotating analyzer configuration is preferred in order to eliminate beam movement on the sample surface due to the beam deviation errors associated with a rotating polarizer. The first rotating analyzer ellipsometer for microellipsometry achieved a resolution of 10 μm at the sample.[78] An ellipsometric image could be achieved by translating the surface, and an ellipsometric spectrum could be collected for selected points on the surface. The combination of spatial and spectral resolution have made this design best suited for semiconductor wafer mapping. The disadvantage of this approach, however, is the inordinate time required for serial mapping combined with serial spectroscopic data collection using a single PMT. More recently, an instrument for microellipsometry has been developed that incorporates a multichannel detector that allows the spectra to be collected in parallel.[81] Thus, with this improvement the spectroscopic image collection time can be reduced by two orders of magnitude. A different approach to spatially resolved ellipsometry involves mapping in parallel by imaging the surface through the analyzer onto a two-dimensional CCD.[79] An ellipsometric image can be obtained from images obtained from at least three azimuthal angles of the rotating polarizer. The disadvantage of this approach lies in the single wavelength operation. Future advances in rotating polarizer/analyzer ellipsometry are expected through high speed spectroscopic imaging of surfaces.

In multichannel spectroscopic ellipsometry for real time analysis of thin films and surfaces, the most popular instrument configurations are also based on the rotating polarizer and analyzer designs. Such designs have been conceived and developed over the past 15 years, and the goal

is to reach theoretical precision limits and improve accuracy to the levels approaching ex situ instruments.[82] In the rotating element designs for multichannel ellipsometry, the readout time of the multichannel detector can be selected as an integral fraction of the optical period of the rotating element. Thus, the detector can be operated so as to acquire spectra in parallel over the wavelength range, and the irradiance waveform at each pixel of the detector can be reconstructed from an integral number of detector readouts. In this way, high quality spectra can be obtained even from a single optical cycle of ~15 ms duration, the highest speed reported to date for real time spectroscopic ellipsometry.[34] Recent and future advances in this area involve incorporating many of the capabilities described above into instruments that can operate in real time, including the adoption of rotating compensator designs, extending the spectral ranges into the infrared and ultraviolet, and imaging surfaces in real time. Although the rotating polarizer and analyzer ellipsometers are now decades old, their principles are expected to be incorporated into the advanced instrument designs of the future.

5.8 References

1. R.M.A. Azzam and N.M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, Amsterdam (1977).
2. H.G. Tompkins, *A User's Guide to Ellipsometry*, Academic Press, New York (1992).
3. E. Passaglia, R.R. Stromberg, and J. Kruger, eds., *Ellipsometry in the Measurement of Surfaces and Thin Films*, NBS, Washington DC (1964), published as National Bureau of Standards Miscellaneous Publication 256.
4. N.M. Bashara, A.B. Buckman, and A.C. Hall, eds., *Proceedings of the Symposium on Recent Developments in Ellipsometry*, North-Holland, Amsterdam (1969); published as *Surf. Sci*., **16** (1969).
5. N.M. Bashara and R.M.A. Azzam, eds., *Proceedings of the Third International Conference on Ellipsometry*, North-Holland, Amsterdam (1976); published as *Surf. Sci*., **56** (1976).
6. R.H. Muller, R.M.A. Azzam, and D.E. Aspnes, eds., *Proceedings of the Fourth International Conference on Ellipsometry*, North-Holland, Amsterdam (1980); published as *Surf. Sci*., **96** (1980).
7. D.E. Aspnes in: *Optical Properties of Solids: New Developments*, edited by B.O. Seraphin, North-Holland, Amsterdam (1976), p. 799.
8. E.A. Irene and Y.J. van der Meulen, *J. Electrochem Soc*., **123**, 1380 (1976).
9. J.B. Theeten, F. Hottier, and J. Hallais, *J. Cryst. Growth*, **46**, 245 (1979).
10. B.D. Cahan and R.F. Spanier, *Surf. Sci*., **16**, 166 (1969).
11. D.E. Aspnes and A.A. Studna, *Appl. Opt*., **14**, 220 (1975).
12. J.B. Theeten, F. Simondet, M. Erman, and J. Pernas, *Proceedings of the Fourth International Conference on Solid Surfaces*, Cannes, 1980, in: *Le Vide, Les Couches Minces Suppl*., **201**, 1071 (1980).

13. S.N. Jasperson and S.E. Schnatterly, *Rev. Sci. Instrum.*, **40**, 761 (1969).
14. V.M. Bermudez and V.H. Ritz, *Appl. Opt.*, **17**, 542 (1978).
15. J.L. Ord, *Surf. Sci.*, **16**, 155 (1969).
16. H.J. Mathieu, D.E. McClure, and R.H. Muller, *Rev. Sci. Instrum.*, **45**, 798 (1974).
17. R.H. Muller and J.C. Farmer, *Rev. Sci. Instrum.*, **55**, 371 (1984).
18. B. Drevillon, J. Perrin, R. Marbot, A. Violet, and J.L. Dalby, *Rev. Sci. Instrum.*, **53**, 969 (1982).
19. W.M. Duncan and S.A. Henck, *Appl. Surf. Sci.*, **63**, 9 (1993).
20. P.S. Hauge and F.H. Dill, *IBM J. Res. Develop.*, **17**, 472 (1973).
21. R.W. Collins, *Rev. Sci. Instrum.*, **61**, 2029 (1990).
22. D.E. Aspnes, *Appl. Opt.*, **14**, 1131 (1975).
23. G.E. Jellison, Jr., and F.A. Modine, *Appl. Opt.*, **29**, 959 (1990).
24. D.E. Aspnes, *J. Opt. Soc. Am.*, **64**, 639 (1974).
25. J.A. Woollam, B. Johs, C.M. Herzinger, J.N. Hilfiker, R. Synowicki, and C. Bungay, *Proc. Soc. Photo-Opt. Instrum. Eng., Crit. Rev.*, **72**, 3 (1999); B. Johs, J.A. Woollam, C.M. Herzinger, J.N. Hilfiker, R. Synowicki, and C. Bungay, *Proc. Soc. Photo-Opt. Instrum. Eng., Crit. Rev.*, **72**, 29 (1999).
26. J. Lee, P.I. Rovira, I. An, and R.W. Collins, *J. Opt. Soc. Am. A*, **18**, 1980 (2001).
27. O. Archer, E. Bigan, and B. Drevillon, *Rev. Sci. Instrum.*, **60**, 65 (1989).
28. D.E. Aspnes and A.A. Studna, *Phys. Rev.*, *B*, **27**, 985 (1983).
29. A.C. Boccara, C. Pickering, and J. Rivory, eds., *Proceedings of the First International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam, (1993); published as *Thin Solid Films*, **233–234** (1993).
30. R.W. Collins, D.E. Aspnes, and E.A. Irene, eds., *Proceedings of the Second International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam, (1998); published as *Thin Solid Films*, **313–314** (1998).
31. A. Moritani and J. Nakai, *Appl. Opt.*, **21**, 3231 (1982).
32. B. Drevillon, *Prog. Cryst. Growth Charact. Mater.*, **27**, 1 (1993).
33. J. Costa, J. Campmany, A. Canillas, J.L. Andujar, and E. Bertan, *Rev. Sci. Instrum.*, **68**, 3135 (1997).
34. I. An, Y. M. Li, H .V. Nguyen, and R. W. Collins, *Rev. Sci. Instrum.*, **63**, 3842 (1992).
35. J.C. Farmer and R.H. Muller, *J. Electrochem. Soc.*, **132**, 313 (1985).
36. S.T. Mayer and R.H. Muller, *J. Electrochem. Soc.*, **135**, 2133 (1988).
37. J.M. Bennett and H.E. Bennett in: *Handbook of Optics*, edited by W.G. Driscoll, McGraw-Hill, New York (1978), Chapt. 10.
38. P. Chindaudom and K. Vedam in: *Optical Characterization of Real Films and Surfaces*, edited by K. Vedam, Academic, New York (1994), p. 191.
39. H.G. Tompkins, in: this Handbook, Chapt. 4.
40. J.A. Zapien, R.W. Collins, and R. Messier, *Rev. Sci. Instrum.*, **71**, 3451 (2000).
41. N.V. Nguyen, B.S. Pudliner, I. An, and R.W. Collins, *J. Opt. Soc. Am. A*, **8**, 919 (1991).
42. S.H. Russev, *Appl. Opt.*, **28**, 1504 (1989).
43. M. Ebert and D.E. Aspnes, *Rev. Sci. Instrum.*, **72**, 3477 (2001).

44. I. An and R.W. Collins, *Rev. Sci. Instrum.*, **62**, 1904 (1991).
45. S.D. Yoo, N.V. Edwards, and D.E. Aspnes, *Thin Solid Films*, **313–314**, 143 (1998).
46. D.E. Aspnes, *J. Opt. Soc. Am.*, **64**, 812 (1974).
47. R.J. King and M.J. Downs, *Surf. Sci.*, **16**, 288 (1969).
48. J.R. Zeidler, R.B. Kohles, and N.M. Bashara, *Appl. Opt.*, **13**, 1115 (1974).
49. R.W. Collins, *Proc. Soc. Photo-Opt. Instrum. Eng.*, **617**, 62 (1986).
50. A.A. Studna, D.E. Aspnes, L.T. Florez, B.J. Wilkens, J.P. Harbison, and R.E. Ryan, *J. Vac. Sci. Technol. B*, **7**, 3291 (1989).
51. G. Laurence, F. Hottier, and J. Hallais, *Rev. Phys. Appl. (Paris)*, **16**, 579 (1981).
52. W. Budde, *Appl. Opt.*, **1**, 201 (1962).
53. R.W. Collins, I. An, J. Lee, and J.A. Zapien, in: this Handbook, Chapter 7.
54. D.E. Aspnes and A.A. Studna, *Rev. Sci. Instrum.*, **49**, 291 (1978).
55. R.H. Muller, in *Advances in Electrochemistry and Electrochemical Engineering*, Vol. 9, edited by R.H. Muller, Wiley, New York (1973), p. 167.
56. F.L. McCrackin, *J. Opt. Soc. Am.*, **60**, 57 (1970).
57. D. Chandler-Horowitz and G.A. Candela, *Appl. Opt.*, **21**, 2972 (1982).
58. M. Yamamoto and O.S. Heavens, *Surf. Sci.*, **96**, 202 (1980).
59. J.A. Woollam and P.G. Snyder, *Mater. Sci. Eng. B*, **5**, 279 (1990).
60. K. Vedam, P.J. McMarr, and J. Narayan, *Appl. Phys. Lett.*, **47**, 339 (1985).
61. Y.J. van der Meulen and N.C. Hien, *J. Opt. Soc. Am.*, **64**, 804 (1974).
62. B. Johs, *Thin Solid Films*, **234**, 395 (1993).
63. J.M.M. de Nijs, A.H.M. Holtslag, A. Hoeksta and A. van Silfhout, *J. Opt. Soc. Am. A*, **5**, 1466 (1988).
64. I. An, J. Lee, B. Hong, and R.W. Collins, *Thin Solid Films*, **313–314**, 79 (1998).
65. G.E. Jellison, Jr., *Appl. Opt.*, **30**, 3354 (1991).
66. C. Chen, I. An, and R.W. Collins, unpublished data.
67. W.R. Hunter, *J. Opt. Soc. Am.*, **66**, 94 (1976).
68. D.E. Aspnes, *Opt. Commun.*, **8**, 222 (1973).
69. M. Schubert, *Phys. Rev. B*, **53**, 4265 (1996).
70. P.S. Hauge, *Surf. Sci.*, **56**, 148 (1976).
71. J. Lee, P.I. Rovira, I. An, and R.W. Collins, *Rev. Sci. Instrum.*, **69**, 1800 (1998).
72. R.W. Collins and J. Koh, *J. Opt. Soc. Am. A*, **16**, 1997 (1999); J. Lee, J. Koh, and R.W. Collins, *Rev. Sci. Instrum.*, **72**, 1742 (2001).
73. E. Franke, C.L. Trimble, M.J. DeVries, J.A. Woollam, M. Schubert, and F. Frost, *J. Appl. Phys.*, **88**, 5166 (2000).
74. O.P.A. Lindquist, K. Jarrendahl, S. Peters, J.T. Zettler, C. Cobet, N. Esser, D.E. Aspnes, A. Henry, and N.V. Edwards, *Appl. Phys. Lett.*, **78**, 2715 (2001).
75. A. Roeseler, *Infrared Spectroscopic Ellipsometry*, Akademie, Berlin (1990).
76. T.E. Tiwald, D.W. Thompson, J.A. Woollam, and S.V. Pepper, *Thin Solid Films*, **313–314**, 718 (1998).
77. B. Drevillon, *Thin Solid Films*, **313–314**, 625 (1998).
78. M. Erman and J.B. Theeten, *J. Appl. Phys.*, **60**, 859 (1986).

Next Page

79. R.F. Cohn, J.W. Wagner, and J. Kruger, *Appl. Opt.*, **27**, 4664 (1988).
80. A. Michaelis and M. Schweinsberg, *Thin Solid Films*, **313–314**, 756 (1998).
81. C. Pickering, J. Russell, V. Nayar, J. Imschweiler, H. Wille, S. Harrington, C. Wiggins, J.-L. Stehle, J.-P. Piel, J. Bruchez, *Thin Solid Films*, **313–314**, 446 (1998).
82. R.W. Collins, I. An, H. Fujiwara, J. Lee, Y. Lu, J. Koh, and P.I. Rovira, *Thin Solid Films*, **313–314**, 18 (1998).

6 Polarization Modulation Ellipsometry

Gerald E. Jellison, Jr. and Frank A. Modine

Solid State Division, Oak Ridge National Laboratory

6.1 Introduction

All ellipsometers consist of a light source, a polarization state generator (PSG), a sample, and a polarization state analyzer (PSA) (see Fig. 6.1). The polarization state of the light beam changes three times as the beam progresses from the source to the detector: 1) after the PSG, 2) after the sample, and 3) after the PSA. Spectroscopic instruments use either a monochromator before the PSG, thus illuminating the sample with quasimonochromatic light, or a monochromator/spectrograph between the PSA and the detector, where the sample is illuminated using white light. Spectroscopic ellipsometers ordinarily achieve high sensitivity by modulating the polarization state of a light beam. The polarization state modulation can be introduced either before the sample (in the PSG), after the sample (in the PSA), or both. Generally speaking, spectroscopic ellipsometers come in two different varieties depending on how the polarization state is modulated: rotating element ellipsometers physically rotate one or more of the optical elements in either the PSG and/or in the PSA, while phase modulation ellipsometers do not physically rotate any optical element. This difference is somewhat arbitrary, since rotating compensator ellipsometers may be considered as phase modulated ellipsometers. However, this chapter will only be concerned with phase modulated ellipsometers (PMEs, sometimes called polarization modulation ellipsometers). These ellipsometers incorporate one or more oscillating retarders, where the oscillating phase retardation is obtained by applying a mechanical stress, an electric field, or a magnetic field to a transparent optical material.

Rotating element ellipsometers are commercially available from several manufacturers. The earliest versions rotated the polarizer following the sample (the analyzer), but others rotate the first polarizer. These instruments have the inherent limitations that only linearly polarizing optical elements are used, and the light beam incident upon the sample contains no circularly polarized component nor is the ellipticity of the

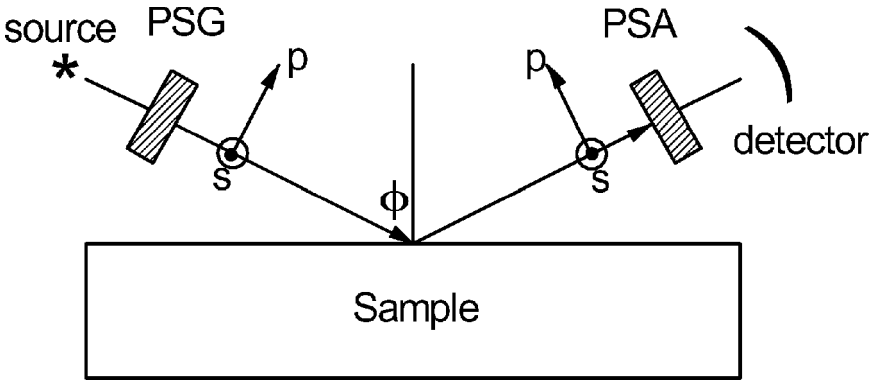


Figure 6.1 Schematic diagram of an ellipsometry experiment. The light beam comes from the light source, passes through the polarization state generator (PSG), which polarizes the light beam incident upon the sample. The sample changes the polarization state of the light beam upon reflection. The polarization state analyzer (PSA) repolarizes the reflected light beam, which is then detected. For isotropic samples, the eigenmodes of reflection are defined by the plane of incidence, with the p-polarization in the plane of incidence and the s-polarization perpendicular to the plane of incidence. The angle ϕ is the angle of incidence.

emerging light beam altered. As a result, polarizer-only instruments can measure only the first three rows and three columns of the sample Mueller matrix. More recently, instruments have become available where a compensating element (either in the PSG or the PSA) is the rotated element. The inclusion of a compensating element *does* allow the fourth row and/or column of the sample Mueller matrix to be measured. The data acquisition speed of any rotating element ellipsometer is fundamentally limited to the rotational velocity of the rotating element, usually less than 100 Hz. Rotating element ellipsometers are discussed in detail in Chapter 5.

Ellipsometers that employ a photoelastic modulator (PEM) as an oscillating retarder are also available commercially, although they are not used as widely as rotating element ellipsometers. As will be discussed in Section 6.2, the retardation of a PEM varies sinusodially and is produced by applying an oscillating mechanical stress to a transparent optical element. There are other oscillating retarder devices, such as electro-optic and Faraday modulators, but PEMs have several advantages: 1) greater numerical aperture, 2) lower operating power requirements, and 3) a wider choice of optical materials. Consequently, only PEM-based polarization modulation ellipsometers are much used, and this chapter will

focus on these ellipsometers. We note, however, that the mathematical formalism is the same for electro-optic modulators and PEMs, so many of the equations developed in this chapter would be applicable to ellipsometers constructed with electro-optic and Faraday modulators. The fundamental frequency of a PEM is generally 10 to 100 kHz, and no physical rotation of any optical element is needed, so PEM-based ellipsometers can acquire data 100 to 1000 times faster than rotating element ellipsometers. There are two main disadvantages of ellipsometers using PEMs: 1) The phase retardation is wavelength dependent, which either reduces their sensitivity and accuracy at some wavelengths and complicates the data analysis or it necessitates close control of the modulation amplitude as a function of wavelength. 2) The time dependence of the light intensity that is produced by an oscillating retardation is quite complicated, which introduces considerable complexity into the data analysis.

In this chapter, we will discuss the experimental aspects of PMEs, where it will be assumed that ellipsometers are used in the specular reflection mode. Although most ellipsometers are operated in the reflection mode, it is also possible for them to operate in the transmission mode, where the instrument is sometimes referred to as a dichograph or a polarimeter. It is also possible to operate the same instrument in the bidirectional ellipsometry or non-specular reflection configuration, where the PSA is not aligned in the specular direction. This configuration has been used for Mie scattering measurements, for the examination of different diffraction orders from gratings, or for other scattering measurements.

The discussion in this chapter will be limited to PEM-based spectroscopic ellipsometers that operate in the ultraviolet-visible-near infrared region of the spectrum. As a result, infrared ellipsometers will not be discussed, but will be discussed in Chapter 11. There are also several instruments that use one or more PEMs, but are generally used at a single wavelength; again, these instruments will not be discussed in this chapter.

Much of the analysis presented in this chapter will assume a working knowledge of Stokes vectors and Mueller matrices. If the reader is unfamiliar with these methods of representing polarized light, it is suggested that Chapter 1 of this volume be read first. Section 6.2 will discuss photoelastic modulators, including a discussion of the mathematical representation of PEMs and a description of some of the common types of PEMs. Several different PME configurations will be discussed in Section 6.3. Section 6.4 will discuss polarization modulation ellipsometers from the perspective of the time-dependent intensity of the light beam incident upon the detector. Both single PEM instruments and dual PEM instruments will be discussed. As with all ellipsometers, calibration is an essential part of ellipsometer operation, which will be discussed in Section 6.5. The time-dependent intensity of the light beam incident upon the detector

in a PME instrument is very complicated, and special techniques must be employed to deconvolute the intensity waveform; these techniques will be discussed in Section 6.6. Errors, both systematic and random, play an important role in the interpretation of any spectroscopic ellipsometry experiment, and these will be discussed in Section 6.7 for the case of PMEs. Section 6.8 will list a bibliography for further reading, and the numbered references.

6.2 The Photoelastic Modulator (PEM)

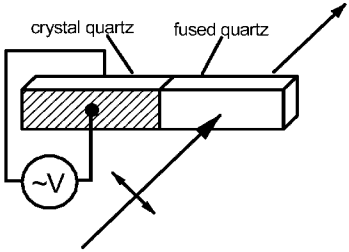
6.2.1 General Description and Historical Perspective

Although there are several different designs of PEMs, they all share common characteristics. A nominally isotropic optical element is set into physical oscillation by some external driving force, usually operating on the driving element of the PEM. This oscillation is resonant, and the frequency is determined by the physical dimensions of the optical element and the speed of sound in the material. Since these quantities also depend upon temperature, the resonant frequency of the PEM will also depend upon its temperature. The resonance quality factor Q can be extremely high for some PEM designs, often exceeding 10,000. This means, of course, that the frequency of the device is very stable, and that very little energy is needed to keep the device in oscillation. Usually a quarter to a full acoustic wave is excited along the length of a bar that is several cm in length to provide a useful optical aperture. Electro-optic devices are not resonant in the same way, meaning that the frequency is not as stable, and significant energy is required to keep the device in oscillation, but it can be driven at any selected frequency.

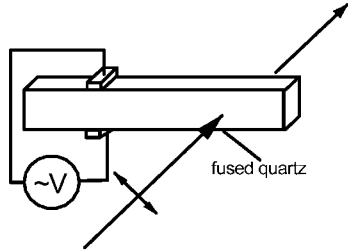
Most PEMs are used in conjunction with a linear polarizer oriented at ±45° with respect to the major oscillating direction of the PEM. In this situation, the resulting polarization state of the light beam emerging from the polarizer-PEM device is dynamically elliptically polarized; that is, the polarization state oscillates between linear, elliptical, and circular polarization states. This is shown schematically in Fig. 6.2a. At the points in the cycle where the optical element of the PEM is unstrained, the polarization state is linearly polarized as determined by the linear polarizer. At other points in the cycle, the optical element is strained, resulting in elliptical, linear, or circular polarized light, depending upon the instantaneous strain of the optical element.

The development of the PEM can be traced to Decker and Muller,[1] who in 1957 described an optical polarization modulator in which a glass block was excited in a resonant oscillation by a piezoelectric transducer

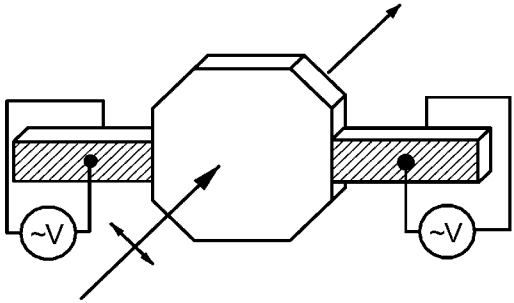
a) Ellipical Polarization States



b) Double Bar PEM



c) Single Bar PEM



d) Octagonal PEM

Figure 6.2 a) The various polarization states are produced by a PEM, shown schematically, where the tip of the electric field vector precesses about the circle, ellipse, or line. A complete cycle will involve several of these polarization states. b) The double bar PEM, first constructed by Kemp and Modine (ref. 4, c). The single bar PEM, first constructed by Badoz and coworkers (ref. 7, d). The octagonal PEM, now sold by Hinds Instruments.

placed on one side of the block, and a transducer on the other side of the block sensed the oscillation and provided an electrical feedback to an amplifier driving the oscillation. Although Decker and Mueller were concerned with optical communications rather than with optical spectroscopy, the design of their instrument encompasses the basic elements of the PEM

that was developed later for spectroscopic purposes by Billardon and Badoz.[2]

A similar but more elegant PEM design was developed simultaneously by Jasperson and Schnatterly[3] and by Kemp.[4] (See Fig. 6.2b.) This design, which was published in 1969, soon became available commercially as a 50 kHz modulator that remains popular today for use in ellipsometry and other forms of optical spectroscopy that require high-frequency polarization modulation. The essence of this design is that a fused quartz bar is mechanically bonded to a $-18°$ x-cut crystal-quartz transducer with a matched resonant frequency. The principal advantages of this design are:

1) Only a single piezoelectric transducer is required.
2) The strength and high Q factor of quartz allow large phase retardations to be obtained at low input power.
3) Little static strain is introduced into the optical element by its bond to the transducer because the elements vibrate sympathetically in a longitudinal mode and can be joined with a flexible cement at an antinode where little stress is transmitted.
4) The electrical circuit for this PEM is just that of a crystal oscillator with controllable amplitude.

Many other variations in PEM design abound, but only a few will be mentioned.

1) In an early and somewhat different design concept by Mollenauer et al.[5], a glass block was held between two steel bars connected by stacks of ceramic piezoelectric transducers. The transducers harmonically squeeze the glass block between the bars to produce the oscillating retardation at a mechanical resonance.
2) In a simplification of the Billardon and Badoz design[2] by Baldachini and Mollenauer,[6] a single ceramic transducer was cemented between a heavy weight and a fused quartz bar that is driven at its fundamental resonance by connecting the transducer in a self-oscillating circuit.
3) In another design[7], the separate drive and sense transducers of the Billardon and Badoz PEM are used, but the transducers are cemented to the sides of a glass bar and a shear stress, instead of a longitudinal stress, provides the mechanical coupling between the transducers and the glass bar. (See Fig. 6.2c.)

4) Most PEM designs use bars with a high aspect ratio that are excited in a longitudinal mode, but other shapes and modes can be used. One successful design uses an octagonal shaped plate in which a two-dimensional standing wave is excited by a quartz transducer that is frequency matched to the resonance of the plate.[8] (See Fig. 6.2d.) Either one or two transducers can be used. The octagonal design is particularly advantageous for the IR because twice as much stress birefringence can be obtained in a given material thickness, and more birefringence is required to obtain quarter-wave or half-wave retardation at long IR wavelengths.

In the development of PEMs, their operating wavelengths have been extended into the UV[9] and the IR,[10] as materials such as MgF_2 and KRS5 were incorporated into PEM designs. A variety of optical materials are now used in commercially available PEMs that operate in various wavelength regions from the vacuum ultraviolet to the far infrared.[11] The PEMs operate at a variety of modulation frequencies, and they are available as calibrated instruments that give a specified retardation at a selectable wavelength.

The performance of PEMs for various applications has for the most part been considered as part of the performance of the optical instruments in which they are incorporated. However, there are performance considerations that are specific to PEM design and fabrication. Although PEMs are electrically driven to oscillate at a highly-tuned mechanical resonance (resulting in the very high Q of the device), the optical retardation δ is generally a complicated function of time.

One complication arises from the residual static strain in the optical element. Some static strain is introduced in the fabrication of an optical element, particularly by cutting and grinding, but, thermal annealing or chemical etching can reduce this strain to minimal values. The rigid cements that bond the optical element to a piezoelectric transducer or to its mechanical supports usually introduce the largest static strains. Good PEM design and fabrication procedures can reduce the strain, but it cannot be eliminated. The static strain introduces a static retardation to the PEM, usually denoted δ_o, which is less than 0.2° (~0.003 rad) at 600 nm for a well-designed and constructed PEM.

If the static retardation is collinear with the modulated strain, then the static strain-induced retardation can be easily incorporated into the mathematical formalism describing the PEM (this will be described in the next subsection). However, if the tensor of the static strain is not collinear with the tensor of the dynamic strain, then the fast axis of the

dynamic retarder depends on time. This case has been analyzed and methods for eliminating the errors that can occur in interpreting spectroscopic measurement have been identified,[12, 13] but will not be discussed further in this chapter. Most modulators do not possess this defect to a significant degree.

If the acoustic excitation of the optical element is not excited in a single standing wave at the resonant frequency, then the time-dependent retardation of the PEM will contain higher harmonics. The generation of these higher harmonics depends upon the harmonic purity of the electrical excitation that is applied to a PEM and also upon the mechanical mounting of the optical element and the transducer. The situation has also been analyzed, and correction procedures developed.[14]

6.2.2 Mathematical Description of a PEM

In most cases, PEMs consist of two elements (see Figures 6.2b–6.2d). The driving element responds to the external signal and dynamically strains the optical element, through which the light beam passes. During the physical oscillation of this optical element, each part of the device will be in compression during half of the cycle, and in expansion during the other half of the cycle. If the optical element is in expansion along its major axis (x-axis), then it will be in compression in the y-axis; similarly, compression in the x-axis will result in expansion in the y-axis. This compression or expansion will increase or decrease the refractive index for light polarized in that direction, making the optical element optically uniaxial. The optical retardation of the device is just that of any strained material (see Born and Wolf) and can be expressed as

$$\delta = \frac{d}{\lambda}\pi n^3(\lambda)(q_{11} - q_{12})(P_{xx} - P_{yy}) = \frac{d}{\lambda}\pi n^3(\lambda)Q(\lambda)P, \qquad (6.1)$$

where d is the thickness of the optical element, λ is the wavelength of light, $n(\lambda)$ is the unstrained refractive index of the element at wavelength λ, $P_{xx}(P_{yy})$ are the strains in the $xx(yy)$ directions, and $q_{11}(q_{12})$ are the strain-optic coefficients. The z-direction is defined as being along the axis of the incident light beam, perpendicular to the front and back surfaces of the PEM. At any particular point (x, y) in the optical element, the strain will be time-dependent. If it can be assumed that the strain is harmonic, then the strain is given by[15]

$$P(x, y, t) = P_0(x, y) + \sum_k P_k \sin(k\omega t + \phi_k) \qquad (6.2)$$

where $P_0(x,y)$ is the static strain at (x,y) through the optical element, ω is the oscillating frequency of the optical element, and ϕ_k is the phase for the k^{th} harmonic.

In Eq. (6.2), and throughout this chapter, it will be assumed that the direction of the static strain is along either the x- or the y-axis of the optical element. With this simplifying assumption, many of the tensor characteristics of the strain can be simplified. If this assumption cannot be made, then the total strain parameter P cannot be treated as a scalar[12,13], as is done in Eqs. (6.1) and (6.2).

If it is assumed that the first harmonic term is the dominant term (this is usually a good approximation, but see Ref. 14 for a treatment of the case where the higher harmonics are important), and the x,y dependence is suppressed, then the dynamic retardation is given by[15]

$$\begin{aligned}\delta(\lambda, t) &= \frac{d}{\lambda}\pi n^3(\lambda)Q(\lambda)[P_0 + P_1\sin(\omega t + \phi_1)] \\ &= A(\lambda)\sin(\omega t + \phi_1) + \delta_0(\lambda).\end{aligned} \tag{6.3}$$

In most PEM experiments, the amplitude of modulation P_1 is controlled by an external voltage V_m. Often, it is convenient to adjust this control voltage such that $A(\lambda) = A_o = 2.4048$. (The reason for this will be discussed below). If it is assumed that the relationship is linear, then

$$P_1 = K_m V_m, \tag{6.4}$$

where K_m is a proportionality constant. Therefore,

$$V_m = \frac{P_1}{K_m} = \frac{A_o\lambda}{K_m d\pi n^3(\lambda)Q(\lambda)}. \tag{6.5}$$

As can be seen from Eqs. (6.3) and (6.5),[15]

$$V_m(\lambda)\,\delta(\lambda) = \text{constant}. \tag{6.6}$$

If the wavelength λ is considerably longer than the wavelength of the band edge of the optical element (for fused silica, this is ~9 eV), then it is a good approximation to use the Cauchy expression for the product:

$$n^3(\lambda)Q(\lambda) = \sum_j \frac{\alpha_j}{\lambda^2}. \tag{6.7}$$

Therefore,[15]

$$\delta_o(\lambda)\lambda = \pi d P_0 \sum_j \frac{\alpha_j}{\lambda^2}, \tag{6.8a}$$

$$\frac{\lambda}{V_m} = \frac{\pi d K_m}{A_o} \sum_j \frac{\alpha_j}{\lambda^2}. \tag{6.8b}$$

Although an infinite sum is shown in Eqs. (6.7) and (6.8), it is usually sufficient to keep only the first two terms ($j = 0$ and $j = 1$), where the second term is treated as a perturbation. If the experiment employing a PEM is spectroscopic and the value of $A(\lambda) = A_o$ is kept constant (such as $A_o = 2.4048$), then the required control voltage V_m is nearly proportional to λ. On the other hand, the static retardation $\delta_o(\lambda)$ of the PEM is inversely proportional to λ.

6.2.3 Stokes Vector Descriptions of the PSG and PSA

The most convenient and accurate way to describe the polarization effects of optical elements in ellipsometers is by the Stokes vector representation (see Chapter 1). The Stokes vector is a four-element real vector that can describe any polarization state, including partially depolarized light. The matrices that are used to describe the transformations of polarization state due to particular optical elements are Mueller matrices, which are 4×4 real matrices. For the purposes of this discussion, we will assume that all Stokes vectors and Mueller matrices are normalized ($S_1 = m_{1,1} = 1$).

In PME instruments, the most common arrangement of optical elements consists of polarizers and polarizer-PEM pairs, which can be azimuthally rotated about the input or output beam direction. In the visible region of the spectrum, polarizers are nearly perfect, giving extinction ratios of greater than $1{:}10^5$. In this chapter, it will be assumed that polarizers are perfect.

There are four different polarizer-PEM combinations used in spectroscopic PMEs:

1) *Polarizer acting as the PSG.* If a single polarizer acts as the PSG, the Stokes vector for the PSG is particularly simple, and is given by

$$\mathbf{S}_{\mathrm{PSG}} = \mathbf{R}(-\theta_p)\mathbf{M}_{\mathrm{p}}\mathbf{R}(\theta_{\mathrm{p}})\begin{pmatrix}1\\0\\0\\0\end{pmatrix} = \begin{pmatrix}1\\C_p\\S_p\\0\end{pmatrix}, \tag{6.9}$$

where the short-hand notation $S_p = \sin(2\theta_p)$; $C_p = \cos(2\theta_p)$ has been used. Note that the last element is 0, resulting from the fact that a linear polarizer cannot generate a polarization state containing any circularly polarized component.

2) *Polarizer acting as the PSA.* Similarly, if a single polarizer acts as the PSA, the transposed Stokes vector representing the PSA is very simple:

$$\mathbf{S}_{\mathrm{PSA}}^{\mathrm{T}} = (1 \;\; 0 \;\; 0 \;\; 0)\mathbf{R}(-\theta_p)\mathbf{M_p}\mathbf{R}(\theta_p) = (1 \;\; C_p \;\; S_p \;\; 0). \tag{6.10}$$

Again, the last element is 0.

3) *Polarizer-PEM pair acting as the PSG.* The most common configuration of the PSG in PMEs is the polarizer-PEM pair. If it is assumed that the input light is unpolarized, and that the azimuthal angles of the polarizer and PEM are θ_p and θ_m, then the Stokes vector for the emerging light beam is given by

$$\mathbf{S}_{\mathrm{PSG}} = \mathbf{R}(-\theta_m)\mathbf{M}_{\mathrm{m}}\mathbf{R}(\theta_m)\mathbf{R}(-\theta_p)\mathbf{M_p}\mathbf{R}(\theta_p)\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \tag{6.11}$$

where the **R** matrices are rotational matrices and $\mathbf{M_p}$ is the Mueller matrix for a polarizer. The Muller matrix for the modulator is

$$\mathbf{M}_{\mathrm{m}} = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \cos(\delta) & \sin(\delta)\\ 0 & 0 & -\sin(\delta) & \cos(\delta)\end{pmatrix} = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & Y_\delta & X_\delta\\ 0 & 0 & -X_\delta & Y_\delta\end{pmatrix}. \tag{6.12}$$

The quantities Y_δ and X_δ contain all the time dependence in the expression for the PSG. Using Eq. (6.3) and assuming that the relative phase ϕ_1 is zero, then

$$\begin{aligned} X_\delta &= \sin(A\sin(\omega t) + \delta_o)\\ &\sim \sin(A\sin(\omega t)) + \delta_o \cos(A\sin(\omega t)) = X + \delta_o Y \end{aligned} \tag{6.13a}$$

$$\begin{aligned} Y_\delta &= \cos(A\sin(\omega t) + \delta_o)\\ &\sim \cos(A\sin(\omega t)) - \delta_o \sin(A\sin(\omega t)) = Y - \delta_o X, \end{aligned} \tag{6.13b}$$

where the approximation has been made that δ_o is small. The basis functions for the polarizer-PEM pair are the

quantities X and Y, which are related to the trigonometric basis functions using an infinite series of integer Bessel functions.

$$X(t) = \sin(A\sin(\omega t)) = 2\sum_{j=0} J_{2j-1}(A)\sin((2j-1)\omega t), \qquad (6.14a)$$

$$Y(t) = \cos(A\sin(\omega t)) = J_0(A) + 2\sum_{j=1} J_{2j}(A)\sin(2j\omega t). \qquad (6.14b)$$

The integer Bessel functions are functions of the Bessel angle A, measured in radians, which is proportional to the magnitude of modulation and is usually proportional to the drive voltage (see Section 6.2.2). If the drive voltage is set to a voltage such that $A_o = 2.4048$, then $J_o(A_o) \sim 0$ and a significant simplification results since the dc term in Eq. (6.14b) vanishes. In addition, the first two integer Bessel functions are near their maxima for this value of A and higher order integer Bessel functions asymptotically approach 0 [$J_1(A_o) = 0.5192$, $J_2(A_o) = 0.4318$, $J_3(A_o) = 0.1990$, $J_4(A_o) = 0.0647$, $J_5(A_o) = 0.0164$].

Multiplying out Eq. (6.1) yields the Stokes vector for the polarizer-PEM pair:

$$\mathbf{S}_{\text{PSG}} = \begin{pmatrix} 1 \\ C_m C_b + S_m S_b Y_\delta \\ S_m C_b - C_m S_b Y_\delta \\ S_b X_\delta \end{pmatrix}, \qquad (6.15)$$

where the shorthand notation $S_m = \sin(2\theta_m)$; $C_m = \cos(2\theta_m)$, etc. has been used and $\theta_b = \theta_m - \theta_p$. Most polarizer-PEM pairs are configured such that the polarizer is set close to $\pm_b 45°$ with respect to the modulator. If $\theta_b = \pm_b 45° + \varepsilon_b$ then $C_b = -\pm_b 2\varepsilon_b$ and $S_b = \pm_b 1$. The Stokes vector for the polarizer-PEM pair is then given by

$$\mathbf{S}_{\text{PSG}} = \begin{pmatrix} 1 \\ \pm_b(-2C_m\varepsilon_b + S_m Y_\delta) \\ \mp_b(2S_m\varepsilon_b + C_m Y_\delta) \\ \pm_b X_\delta \end{pmatrix}. \qquad (6.16)$$

Note: The $\pm_b$ notation indicates that the prefactor equals $+1$ when $\theta_b = +45°$ and equals -1 when $\theta_b = -45°$.

The light emerging from a polarizer-PEM pair will always have a circularly polarized component (from the fourth component of the Stokes vector). There also will be a linearly polarized component, but only a linear combination of the second and third components of the Stokes vector will be sampled. Therefore, the polarizer-PEM pair cannot create all possible polarization states without being physically rotated.

4) *PEM-polarizer pair as the PSA.* If the PSA consists of a PEM-polarizer pair (where the light beam first passes through the PEM), then the transposed Stokes vector is given by

$$\mathbf{S}_{\mathrm{PSA}}^{\mathrm{T}} = (1 \;\; 0 \;\; 0 \;\; 0)\mathbf{R}(-\theta_p)\mathbf{M_p}\mathbf{R}(\theta_p)\mathbf{R}(-\theta_m)\mathbf{M_m}\mathbf{R}(\theta_m) \qquad (6.17)$$

The matrices can be multiplied out, giving

$$\mathbf{S}_{\mathrm{PSA}} = (1 \quad C_m C_b + S_m S_b Y_b \quad S_m C_b - C_m S_b Y_\delta \quad -S_b X_\delta). \qquad (6.18)$$

This matrix is very similar to the transpose of the Stokes vector of the polarizer-PEM pair acting as the PSG, with only a sign change in the fourth component.

As with all ellipsometers, it is critical to calibrate both the PSG and the PSA as accurately as possible. If the PSG or the PSA is a single polarizer, then the only calibration needed is the azimuthal angle θ_m with respect to the plane of incidence. If a PEM is included in either the PSG or the PSA, then the calibration procedure is more complicated and four calibrations need to be made. As with the simple polarizer PSG or PSA, the azimuthal angle θ_m of the PEM with respect to the plane of incidence needs to be calibrated. In addition to the θ_m calibration, the azimuthal angle of the polarizer with respect to the PEM θ_b must also be calibrated, where this is normally set to ±45°. Two calibrations must also be made as a function of wavelength: the modulator control voltage $V_m(\lambda)$ required to give $A = 2.4048$, and the modulator static strain $\delta(\lambda)$. Both $V_m(\lambda)$ and $\delta(\lambda)$ are strong functions of wavelength (see Eqs. [6.8]), but can be easily parameterized using a Cauchy expansion. As a result, more accurate calibrations are obtained if the calibration data are then fit to the expressions in Eqs. (6.8), retaining only the fitted Cauchy coefficients. Usually only two or three terms need to be retained. Calibrations will be the subject of Section 6.6.

6.3 Experimental Configurations of Polarization Modulation Ellipsometers

Spectroscopic ellipsometers that use photoelastic modulators can exist in several different experimental configurations. In this section, we will discuss four different configurations that have been presented in the literature, shown schematically in Figs. 6.3 through 6.6. This section will concentrate on presenting the experimental apparatus that is used to make the measurements. The mathematical formalism is similar for all cases, and will be presented in later sections.

6.3.1 Polarization Modulation Ellipsometry (PME) with Analog Data Acquisition (Jasperson and Schnatterly[3])

The first spectroscopic ellipsometer was a polarization modulation ellipsometer, constructed by Jasperson and Schnatterly[3] (J&S) and is shown schematically in Fig. 6.3. The light source for this ellipsometer was a quasimonochromatic light beam coming from a xenon arc lamp filtered through a monochromator. This light beam was then dynamically elliptically polarized with the polarizer-PEM pair before it was incident upon the sample. The sample changes the polarization state of the light beam, which is then linearly polarized by the analyzer. The intensity of the light beam is then detected by the photomultiplier tube placed after the analyzer. For some of Jasperson and Schnatterly's measurements, the analyzer was removed.

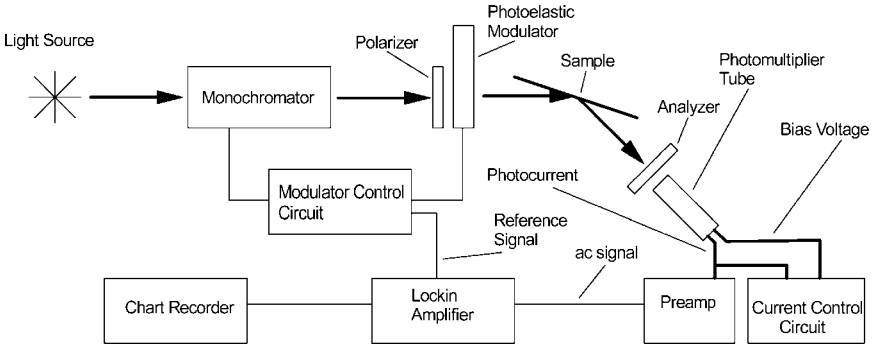


Figure 6.3 The experimental system of Jasperson and Schnatterly,[3] where the data acquisition is analog.

The photomultiplier tube (PMT) is controlled by a special electronic circuit, whereby the dc photocurrent is fed back to the bias control circuit so that the total dc photocurrent is kept constant. That is, if the ellipsometer is set at a particular wavelength where the light intensity from the monochromator is strong, then the control circuit will decrease the dynode bias voltage on the PMT such that the dc current is equal to some preset value. Similarly, if the ellipsometer is set in the ultraviolet (where the light level from the xenon lamp is small) or in the infrared (where the quantum efficiency of the PMT is small), the dynode bias voltage is increased such that the dc current is equal to the same value. This type of arrangement has a particular advantage in that the gain ranges on the detection apparatus (such as lock-in amplifiers or waveform digitizers) need not be changed. In addition, if the time constant of the circuit is set to ~10 milliseconds, then the PMT bias circuit can respond to fluctuations in the lamp intensity (such as due to 60 Hz line noise), and this noise will not appear on the output signal.

As will be discussed later, the intensity waveform from a polarization modulation ellipsometer consists of a dc component, a 1f component, a 2f component, as well as higher harmonics, where f is the frequency of the PEM (~50 kHz). The dc component is constrained to be constant due to the PMT bias control circuitry. Either or both of the 1f and 2f signals can be detected with a lock-in amplifier, as shown in Fig. 6.3. The higher harmonics do not give any further information, since the prefactors for all odd harmonics are the same, as are the prefactors for all even harmonics.

Jasperson and Schnatterly's ellipsometer used a single lock-in amplifier where the output was directed to a chart recorder. Later implementations[16, 17] of this instrument directed the output to a computer. Depending upon the azimuthal orientations of the polarizer-PEM pair and the analyzing polarizer, the parameters N, S, or C were measured (see Chapters 1 and 3 for definitions of the measured parameters N, S, and C), where N or C are proportional to the 2f amplitude, and S is proportional to the 1f amplitude.

Note that the J&S ellipsometer can measure only one parameter (N, S, or C) at a time. If a second lock-in amplifier were used, then it would be possible to measure two parameters [(N, S) or (C, S)] at a time. A complete determination of N, S, and C requires multiple spectral scans.

6.3.2 Phase Modulated Ellipsometry (PME) with Digital Data Acquisition

Figure 6.4 shows a schematic diagram of the polarization modulation ellipsometer constructed by Drévillon et al.[18] This instrument differs from the J&S instrument in two primary ways: 1) the light source is not monochromatized before the sample, but after the analyzing

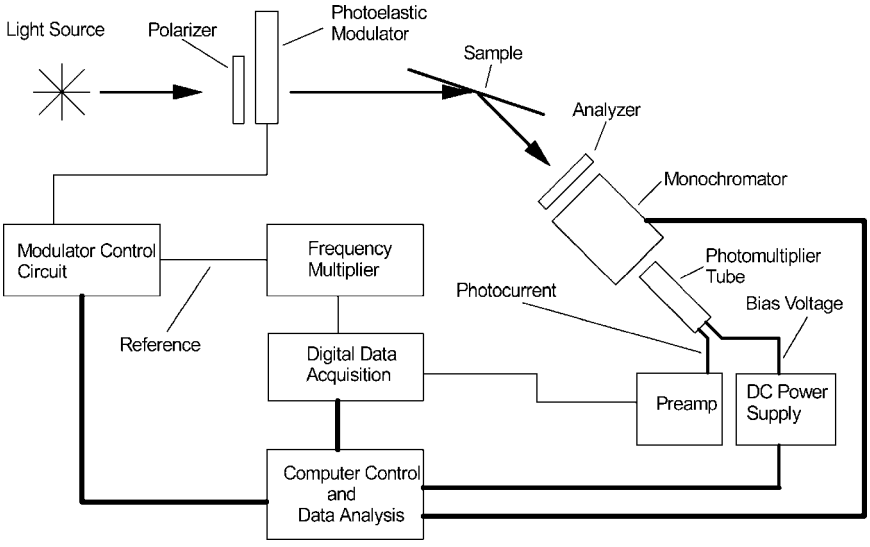


Figure 6.4 The experimental polarization modulation ellipsometer scheme of Drévillon et al.[18] using digital data acquisition.

polarizer; 2) the signal from the PMT is not fed into a lock-in amplifier, but rather digitized and fed into the computer.

The placement of the monochromator after the sample has a particular advantage for situations where the sample may get hot. When the incident light beam is first monochromatized, the detector sees the reflected light beam, as well as any black body radiation given off by the sample (this is in addition to any stray room light that may get to the detector). On the other hand, the placement of the monochromator after the sample will reduce considerably stray room light and black body radiation from the sample.

The waveform from the Drévillon et al. instrument is first digitized using a waveform digitizer. Since the PEM frequency is 50 kHz, and one is interested in some of the higher harmonics, the waveform must be digitized at least at ~1 MHz. However, once the waveform has been digitized, it is possible to determine the amplitudes of all harmonics from a single measurement. Thus, two independent parameters [(N, S) or (C, S)] can be simultaneously determined. In addition, the amplitudes of the third and fourth harmonics can be used to calibrate the instrument (see Ref. 14). A variation of this instrument is presently available commercially.[19]

One improvement to this instrument is to incorporate optical fibers in the input and output arms of the ellipsometer. This allows the light source and the monochromator to be placed arbitrarily, making it much easier to

use the instrument for *in situ* ellipsometry measurements of film growth. The ellipsometers of Duncan and Henck[20] and Drévillon et al.[21] incorporate this capability.

Another improvement, also incorporated into the ellipsometers of Duncan and Henck[20] and Drévillon et al.[21], is to replace the single PMT either with a photodiode array[20], or with several PMTs.[21] With this improvement, the intensity waveforms from several different wavelengths can be acquired at the same time. The use of a photodiode array is common in rotating element ellipsometers (see Chapter 7 by Collins), but the incorporation into instruments using PEMs is complicated by the speed required for data acquisition. Since rotating element ellipsometers require much slower digitization speeds, these instruments can often use a single A/D channel to collect all the data. PEM systems require digitization speeds near 1 MHz, so often a separate A/D channel is required for each wavelength sampled. In addition, the waveform is more complicated than a simple Fourier expansion, requiring more extensive analysis of the waveform. However, present systems are capable of collecting 16 to 48 channels (wavelengths) of data simultaneously.[19]

6.3.3 Two-Channel Spectroscopic Polarization Modulation Ellipsometer (2-C SPME)

The ellipsometers discussed above are capable of only measuring two parameters at a time. However, there are situations where it is important to measure all three parameters (N, S, and C) simultaneously. This has been addressed by the 2-C SPME of Jellison and Modine,[22] shown schematically in Fig. 6.5.

The main difference between this instrument and the previous instruments is that the analyzing polarizer is replaced with a Wollaston prism (any similar beam-splitting prism could also be used). This prism will split the incoming light beam, with one polarization being deflected in one direction and the other polarization deflected in another direction. Two independent PMT detectors are used, one for each light beam from the Wollaston prism. As above, the two PMTs are controlled to give constant output current, and the output waveform from each PMT is analyzed using two lock-in amplifiers, one set at 1f and the other at 2f. If the Wollaston prism is set at ±45° with respect to the plane of incidence and if the azimuthal angle of the polarizer-PEM pair is set to ±22.5° (or, equivalently, to ±67.5°), then the two 1f channels can be used to measure S, while the two 2f channels will be proportional to $(N - C)$ and $(N + C)$. As a result, it is possible to simultaneously measure N, S, and C with this instrument.

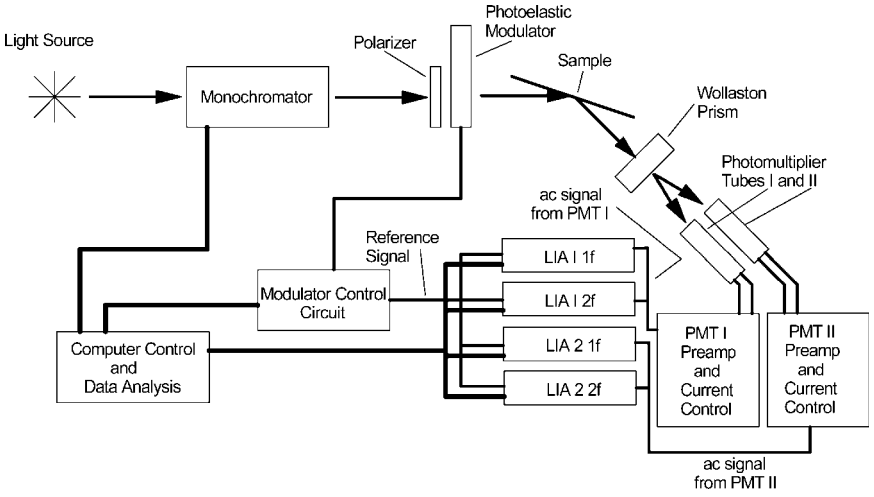


Figure 6.5 The 2-Channel Spectroscopic Polarization Modulation Ellipsometer (2-C SPME) of Jellison and Modine.[18]

The main advantage of this instrument is that N, S, and C are measured simultaneously. For many samples, this is redundant, but for situations where the sample surface is depolarizing, the measurement of all three becomes necessary. As will be discussed later, the sum $(N^2 + S^2 + C^2)$ is equal to 1 if there are no depolarization elements in the light train. However, if there are depolarization elements (such as the sample itself!) then the sum $(N^2 + S^2 + C^2)$ will be less than 1. For some cases, this sum can be related to the degree of light polarization in the system.

The use of the Wollaston prism does incur some disadvantages. First of all, the angle of the two light beams from the Wollaston will be a function of the wavelength of light. Therefore, it is not possible to use a Wollaston in conjunction with fiber optics, nor is it possible to easily place the monochromator after the sample. Secondly, four lock-in amplifiers are required, which increases the cost of the instrument.

6.3.4 Two-Modulator Generalized Ellipsometer (2-MGE)

Each of the ellipsometers discussed above utilizes a single polarizer-PEM pair. However, it is possible to incorporate two polarizer-PEM pairs, one in the PSG arm of the instrument, the other in the PSA arm.[23,24] The arrangement of the polarizer-PEM pair is as shown in Fig. 6.6: the light

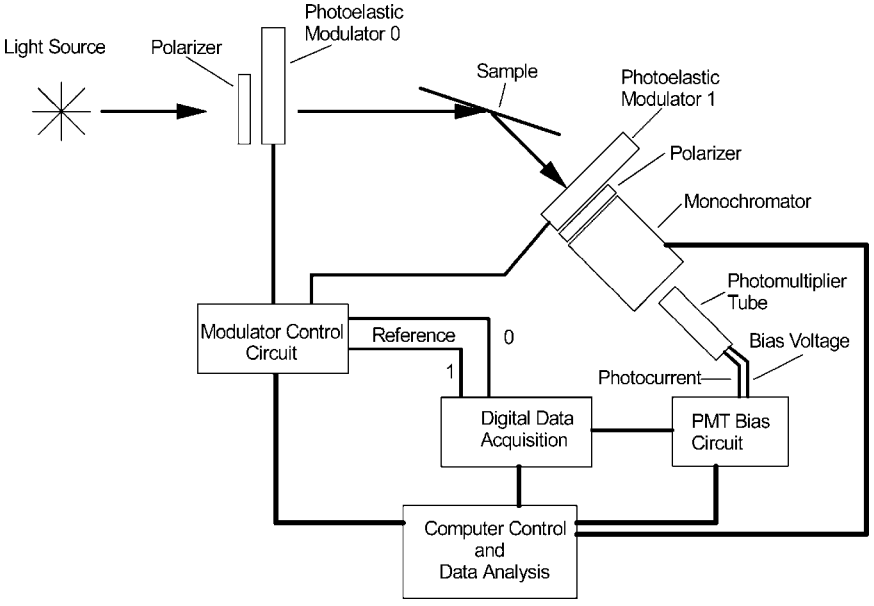


Figure 6.6 Schematic diagram of the 2-Modulator Generalized Ellipsometer (2-MGE) of Jellison and Modine.[23]

passes through the polarizer and then the PEM in the PSG arm, and the reflected light passes through the PEM and then the polarizer. The resonant frequencies of the two PEMs must be significantly different (such as 50 kHz and 60 kHz for the instrument described in Ref. 24).

The main advantage of this type of instrument is that many more parameters can be measured. While the PMEs described to this point will measure one, two, or three of the parameters *N*, *S*, and *C*, the 2-MGE can measure eight parameters simultaneously. These parameters are related to the Mueller matrix of the sample, and the azimuthal angles of the PSG and the PSA determine which elements of the sample Mueller matrix are measured.

For example if either the PSG or the PSA is set to ±45° and the other is set to 0° or 90°, then the 2-MGE can measure the parameters *N*, *S*, and *C* and five additional parameters. For anisotropic, non-depolarizing samples, the instrument in this configuration will completely characterize the sample Mueller matrix; that is, there is no more information that can be obtained from a reflection ellipsometry experiment at that angle of incidence, even though seven elements of the reduced sample Mueller matrix remain unmeasured. This ability to handle anisotropic samples is one of the main strengths of the 2-MGE.

6.4 Light Intensity Through a Polarization Modulation Ellipsometer

The light intensity through any ellipsometer can be obtained by the matrix multiplication

$$\text{Intensity} = \mathbf{S_{PSA}}^{\mathbf{T}}\,\mathbf{M}\,\mathbf{S_{PSG}}. \tag{6.19}$$

The matrix $\mathbf{M}$ is the Mueller matrix for all optical elements between the PSG and the PSA; these elements include the sample, and any other optical elements between the PSG and the sample and between the sample and the PSA. The vector $\mathbf{S_{PSG}}$ is the Stokes vector for the light beam leaving the PSG, and $\mathbf{S_{PSA}}^{\mathbf{T}}$ is the transposed Stokes vector representing the polarization effects of the PSA, both of which have been discussed in Section 6.2.

The intensity waveform from any PME instrument is complicated by the fact that the basis functions for the intensity are not standard sines and cosines of rotating element ellipsometers, but rather sines of sines and cosines of sines [see the expressions for $X(t)$ and $Y(t)$ in Eqs. (6.13) and (6.14)]. Therefore, the intensity waveform is most conveniently written in terms of these basis functions. In this section, we will present the intensity waveform expression for various PMEs in terms of a dc term and coefficients of the basis functions $X(t)$ and $Y(t)$. The coefficients are essentially the measured quantities of the PMEs, and techniques for extracting these from the waveform will be presented in Section 6.5.

6.4.1 Mueller Matrices for Various Samples

The normalized Mueller matrix for all of the optical elements between the PSG and the PSA will be of the form

$$\mathbf{M} = \begin{pmatrix} 1 & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{pmatrix}, \tag{6.20}$$

which contains 15 unknown matrix elements. Most ellipsometers do not measure the absolute reflectance or transmission of the sample. This is equivalent to normalizing the sample Mueller matrix to the m_{11} term, which is then set to 1 in Eq. (6.20). Fortunately, the Mueller matrix for most ellipsometry samples is considerably simpler, where many of the

elements are known to be 0 or to be equal to other elements. If it can be assumed that the sample and intervening optics are non-depolarizing, then the 15 elements of the normalized sample Mueller matrix are not independent, and are related by nine bilinear constraints[25] (see Chapter 3, Section 3.2.3)

1) *Free Space* If the ellipsometer is placed in the straight-through configuration, the sample is effectively free space, which will not change the polarization state at all. Therefore, the Mueller matrix for free space is just the 4×4 identity matrix:

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{6.21}$$

Free space measurements are generally used for calibrations.

2) *Isotropic sample* Most ellipsometry measurements are performed on isotropic samples without intervening optics or depolarization effects. In this case, the sample Mueller matrix is also particularly simple and is given by

$$\mathbf{M} = \begin{pmatrix} 1 & -N & 0 & 0 \\ -N & 1 & 0 & 0 \\ 0 & 0 & C & S \\ 0 & 0 & -S & C \end{pmatrix}. \tag{6.22}$$

The parameters N, S, and C are the associated ellipsometry parameters, and are given by

$$N = \cos(2\psi), \tag{6.23a}$$

$$S = \sin(2\psi)\sin(\Delta), \tag{6.23b}$$

$$C = \sin(2\psi)\cos(\Delta). \tag{6.23c}$$

These are related to the traditional ellipsometry parameters ψ and Δ (see Chapter 1) by

$$\rho = \frac{r_{pp}}{r_{ss}} = \frac{C + iS}{1 + N} = \tan(\psi)e^{i\Delta}. \tag{6.23d}$$

Obviously, the parameters N, S, and C are not independent and are constrained by the equation

$$N^2 + S^2 + C^2 = 1 \tag{6.23e}$$

3) *Isotropic sample with windows* Many ellipsometry measurements necessarily include additional optical elements. If the experiment takes place *in situ*, then windows must separate the PSG and the PSA of the ellipsometer from the sample being investigated. If a small spot size is needed, then a focusing lens and collection lens are required. In both of these cases, there are additional optical elements (generically called windows here) between the PSG and the PSA, which can alter the polarization state of the light coming from the PSA or reflected from the sample due to the inherent strain of the optic. Normally, the windows are approximated as simple retarders where the retardation is small, and the total sample Mueller matrix (including the windows) is given by[26]

$$\begin{aligned}\mathbf{M}_{\mathrm{G}} &= \mathbf{M}_{\mathrm{W0}}\mathbf{M}\mathbf{M}_{\mathrm{W1}} \\ &= \begin{pmatrix} 1 & -N & 0 & S_0N \\ -N & 1 & S_1S & -S_0 - S_1C \\ 0 & S_0S & C - WC & S + WC \\ -S_1N & S_1 + S_0C & -(S + WC) & C - WC \end{pmatrix}.\end{aligned} \tag{6.24}$$

If the retardation of the windows is δ_0, δ_1 and the azimuthal angle of the fast axis is θ_0, θ_1, then $S_{0,1} = \delta_{0,1}\sin(2\theta_{0,1})$ and $W = \delta_0\cos(2\theta_0) + \delta_1\cos(2\theta_1)$. Using the first-order approximation, the effect of the windows on the experiment can be expressed with only three parameters, since the two cosine terms are additive.

4) *Anisotropic samples* Another example of a sample Mueller matrix occurs if the sample is anisotropic (see Chapters 3 and 9). The sample Mueller matrix for an anisotropic sample is given by[24]

$$\mathbf{M} = \begin{pmatrix} 1 & -N - \alpha_{ps} & C_{sp} + \zeta_1 & S_{sp} + \zeta_2 \\ -N - \alpha_{sp} & 1 - \alpha_{sp} - \alpha_{ps} & -C_{sp} + \zeta_1 & -S_{sp} + \zeta_2 \\ C_{ps} + \xi_1 & -C_{ps} + \xi_1 & C + \beta_1 & S + \beta_2 \\ -S_{ps} + \xi_2 & S_{ps} + \xi_2 & -S + \beta_2 & C - \beta_1 \end{pmatrix}, \tag{6.25}$$

where α_{ps}, α_{sp}, β_1, β_2, ζ_1, ζ_2, ξ_1, and ξ_2 are defined in terms of the seven ellipsometry parameters in the NSC

representation: N, S, C, S_{sp}, C_{sp}, S_{ps}, and C_{ps} (see Chapter 3, Section 3.2.3). If the sample is nondepolarizing, then the sum of the squares of the seven ellipsometry parameters in the NSC representation is equal to 1.

6.4.2 Intensity for a Standard PME

The most common PME configuration used consists of a polarizer-PEM pair acting as the PSG, and a polarizer acting as the PSA (the P PEM S A configuration). This is the configuration used by Jasperson and Schnatterly[3], Drevillon et al.[18], Duncan and Henck[20], and Jellison and Modine[22], and is the configuration of the commercial PME.[19] Using Eqs. (6.19), (6.10), and (6.14), it is a simple matter to calculate the intensity of the light beam incident upon the detector in this configuration:

$$I(t) = (1 \quad C_{p1} \quad S_{p1} \quad 0)\mathbf{M}\begin{pmatrix} 1 \\ C_m C_b + S_m S_b Y_\delta \\ S_m C_b - C_m S_b Y_\delta \\ S_b X_\delta \end{pmatrix}. \tag{6.26}$$

The time dependence of the intensity is implicitly included in the two basis functions X_δ and Y_δ, which include PEM static retardation. Using Eqs. 6.13, the contribution of PEM static retardation can be separated out:

$$\begin{aligned} I(t) = I_{dc} + I_X X_\delta + I_Y Y_\delta = I_{dc} + (I_X - \delta_o I_Y)\, X(t) \\ + (I_Y + \delta_o I_X)\, Y(t), \end{aligned} \tag{6.27}$$

where it is assumed that the static retardation δ_o of the modulator is small and therefore first-order corrections are sufficient. The intensity waveform is then entirely determined by the two time-dependent basis functions X and Y, as well as the three parameters I_{dc}, I_X, and I_Y, which are given by

$$\begin{aligned} I_{dc} = 1 + C_{p1} m_{21} + S_{p1} m_{31} + C_b[C_m(m_{12} + C_{p1} m_{22} \\ + S_{p1} m_{32}) + S_m(m_{13} + C_{p1} m_{23} + S_{p1} m_{33})], \end{aligned} \tag{6.28a}$$

$$I_X = S_b(m_{14} + C_{p1} m_{24} + S_{p1} m_{34}), \tag{6.28b}$$

$$\begin{aligned} I_Y = S_b[-C_m(m_{13} + C_{p1} m_{23} + S_{p1} m_{33}) \\ + S_m(m_{12} + C_{p1} m_{22} + S_{p1} m_{32})]. \end{aligned} \tag{6.28c}$$

Because the PSA contains only a polarizer, the terms I_{dc}, I_X, and I_Y do not depend upon elements of the fourth row.

Another possible configuration for a single modulator PME would be the P S PEM A arrangement where the PEM-polarizer pair is placed after the sample. In this case, a single polarizer is used for the PSG and a PEM-polarizer pair is used for the PSA. Again, it is a simple matter to calculate the time-dependent intensity of the light beam incident upon the detector:

$$I(t) = (1 \quad C_m C_b + S_m S_b Y_\delta \quad S_m C_b - C_m S_b Y_\delta \quad -S_b X_\delta)\mathbf{M}\begin{pmatrix} 1 \\ C_{p0} \\ S_{p0} \\ 0 \end{pmatrix}. \quad (6.29)$$

This has the same general form as Eq. (6.27), where the coefficients are given by

$$I_{dc} = 1 + C_{p0}m_{12} + S_{p0}m_{13} + C_b[C_m(m_{21} + C_{p0}m_{22} + S_{p0}m_{23}) + S_m(m_{31} + C_{p0}m_{32} + S_{p0}m_{33})], \quad (6.30a)$$

$$I_X = -S_b(m_{41} + C_{p0}m_{42} + S_{p0}m_{43}), \quad (6.30b)$$

$$I_Y = S_b[-C_m(m_{31} + C_{p0}m_{32} + S_{p0}m_{33}) + S_m(m_{21} + C_{p0}m_{22} + S_{p0}m_{23})]. \quad (6.30c)$$

Because the PSG contains only a polarizer, the terms I_{dc}, I_X, and I_Y do not depend upon elements of the fourth column. Neither the P PEM S A configuration or the P S PEM A configuration is sensitive to the m_{44} element of **M**.

Nearly all PMEs are operated with the polarizer-PEM azimuthal angle set to $\theta_b = \pm_b 45°$. If first-order errors in θ_b are included, then $S_b = \pm_b 1$ and $C_b = -\pm_b 2\ \varepsilon_b$ and this error only affects the I_{dc} term to first order. (Again, note that the $\pm_b$ notation indicates that the prefactor equals $+1$ when $\theta_b = +45°$ and equals -1 when $\theta_b = -45°$.) In addition, the sample Mueller matrix is often much simpler than the most general form used above. For example, if it can be assumed that the sample is isotropic, then the sample Mueller matrix simplifies to Eq. (6.22), and the terms I_{dc}, I_X, and I_Y are also simplified. For the P PEM S A configuration, these terms are given by

$$I_{dc} = 1 - C_{p1}N - \pm_b 2\varepsilon_b[C_m(-N - C_{p1}) + S_m S_{p1} C], \quad (6.31a)$$

$$I_X = \pm_b S_{p1} S, \quad (6.31b)$$

$$I_Y = \pm_b[-C_m S_{p1} C + S_m(-N + C_{p1})]. \quad (6.31c)$$

For the P S PEM P configuration, these terms are given by

$$I_{dc} = 1 - C_{p0}N - \pm_b 2\,\varepsilon_b[C_m(-N + C_{p0}) + S_m S_{p0} C], \quad (6.32a)$$

$$I_X = \pm_b S_{p0} S, \quad (6.32b)$$

$$I_Y = \pm_b[-C_m S_{p0} C + S_m(-N + C_{p0})]. \quad (6.32c)$$

As can be seen from Eqs. (6.31) and (6.32) above, the only difference between the coefficients for the P PEM S A and P S PEM A configurations is the permutation of the $0 \Leftrightarrow 1$ indices. Therefore, there is no difference between the P PEM S A configuration and the P S PEM A configuration for isotropic samples; the same information can be obtained whether the polarizer-PEM pair is on the input side of the sample or on the output side.

Many calibrations of PMEs are done in the straight-through alignment, where the sample Mueller matrix becomes just the Mueller matrix of free space (Eq. 6.21). In this case, Eqs. (6.31) and (6.32) can be simplified by setting $N = S = 0$; $C = 1$:

$$I_{dc} = 1 - \pm_b 2\varepsilon_b[C_m C_{p0} + S_m S_{p0}] = 1 - \pm_b 2\varepsilon_b C_{m-p}, \quad (6.33a)$$

$$I_X = 0, \quad (6.33b)$$

$$I_Y = \pm_b[-C_m S_{p0} + S_m C_{p0}] = \pm_b S_{m-p}. \quad (6.33c)$$

where the quantities $C_{m-p} = \cos[2(\theta_m - \theta_p)]$ and $S_{m-p} = \sin[2(\theta_m - \theta_p)]$. Obviously, there can only be a relative azimuthal angle, since the sample (free space) cannot define a plane of incidence. The total intensity can be written as

$$I(t) = 1 - \pm_b 2\varepsilon_b C_{m-p} - \pm_b \delta_o S_{m-p}\, X(t) + \pm_b S_{m-p}\, Y(t). \quad (6.34)$$

If θ_{m-p} is set to ±45°, then the coefficient of the X basis function is proportional to the static retardation δ_o and the static retardation of the PEM can be determined.

6.4.3 Intensity for the 2-Modulator Generalized Ellipsometer (2-MGE)

As discussed in Section 6.3, the 2-modulator generalized ellipsometer (2-MGE) uses a polarizer-PEM pair in both the PSG and the

PSA, where the resonant frequencies of the two PEMs are significantly different (such as 50 and 60 kHz). Although this complicates both the experimental setup and the data reduction, considerably more information becomes available than is available with single PEM systems. Single PEM systems produce an intensity that can be expressed with a dc term and two independent coefficients of the basis functions X and Y. As will be shown below, the intensity waveform of the 2-MGE is considerably more complicated, but it can be expressed with a dc term and eight coefficients of basis functions. These additional parameters can be related directly to elements of the sample Mueller matrix, where the specification of which Mueller matrix elements are incorporated into the expression of the intensity waveform depends upon the azimuthal angular orientation of the PSG and the PSA. If the correct azimuthal angles are chosen, *it is possible to completely characterize the sample Mueller matrix of a non-depolarizing sample using the 2-MGE.*

The intensity of the time-dependent light beam for the 2-MGE can be calculated in the same manner as above for single PEM PMEs. The Stokes vectors for the polarizer-PEM PSG and for the PEM-polarizer PSA are given in Eqs. (6.15) and (6.18). From these, the intensity can be calculated:

$$I(t) = (1 \quad C_{m1}C_{b1} + S_{m1}S_{b1}Y_{1\delta} \quad S_{m1}C_{b1} - C_{m1}S_{b1}Y_{1\delta} \quad -S_{b1}X_{1\delta})\mathbf{M}\begin{pmatrix} 1 \\ C_{m0}C_{b0} + S_{m0}S_{b0}Y_{0\delta} \\ S_{m0}C_{b0} - C_{m0}S_{b0}Y_{0\delta} \\ S_{b0}X_{0\delta} \end{pmatrix} \tag{6.35}$$

where the PSG polarizer and PEM are identified with the subscript 0, and the PSA polarizer and PEM are identified with the subscript 1. The total intensity can be written similarly to Eq. (6.27), but now with eight basis functions and eight associated coefficients:

$$I(t) = I_{dc} + I_{X0}\,X_{0\delta} + I_{Y0}\,Y_{0\delta} + I_{X1}\,X_{1\delta} + I_{Y1}\,Y_{1\delta} + I_{X0X1}\,X_{0\delta}\,X_{1\delta} + I_{X0Y1}\,X_{0\delta}\,Y_{1\delta} + I_{Y0X1}\,Y_{0\delta}\,X_{1\delta} + I_{Y0Y1}\,Y_{0\delta}\,Y_{1\delta}. \tag{6.36a}$$

This intensity is written with the basis functions $X_{0\delta}$, etc., which implicitly contain the modulator static retardations δ_0 and δ_1. If the static strain-induced retardations are treated to first order, then the intensity can be rewritten in terms of unstrained basis functions:

$$I(t) = (1 \quad X_0 \quad Y_0 \quad X_1 \quad Y_1 \quad X_0X_1 \quad X_0Y_1 \quad Y_0X_1 \quad Y_0Y_1)$$
$$\times \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \delta_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\delta_0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \delta_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\delta_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \delta_1 & \delta_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\delta_1 & 1 & 0 & \delta_0 \\ 0 & 0 & 0 & 0 & 0 & -\delta_0 & 0 & 1 & \delta_1 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\delta_0 & -\delta_1 & 1 \end{pmatrix} \begin{pmatrix} I_{dc} \\ I_{X0} \\ I_{Y0} \\ I_{X1} \\ I_{Y1} \\ I_{X0X1} \\ I_{X0Y1} \\ I_{Y0X1} \\ I_{Y0Y1} \end{pmatrix} \tag{6.36b}$$

To simplify the expressions, assume that $\theta_{b0} = \pm_{b0}45°$, $\theta_{b1} = \pm_{b1}45°$. Therefore, $C_{b0} = C_{b1} = 0$ and $S_{b0} = \pm_{b0}1$, $S_{b1} = \pm_{b1}1$. The intensity expression of Eq. (6.35) becomes

$$I(t) = (1 \quad S_{m1}S_{b1}Y_{1\delta} \quad -C_{m1}S_{b1}Y_{1\delta} \quad -S_{b1}X_{1\delta})\mathbf{M}\begin{pmatrix} 1 \\ S_{m0}S_{b0}Y_{0\delta} \\ -C_{m0}S_{b0}Y_{0\delta} \\ S_{b0}X_{0\delta} \end{pmatrix}. \tag{6.37}$$

The values of the nine coefficients in Eqs. (6.36a) and (6.36b) are functions of the individual elements of the sample Mueller matrix and are given by

$$I_{dc} = 1 \tag{6.38a}$$

$$I_{X0} = \pm_{b0}m_{14} \tag{6.38b}$$

$$I_{Y0} = -\pm_{b0}(C_{m0}m_{13} - S_{m0}m_{12}) \tag{6.38c}$$

$$I_{X1} = -\pm_{b1}m_{41} \tag{6.38d}$$

$$I_{Y1} = -\pm_{b1}(C_{m1}m_{31} - S_{m1}m_{21}) \tag{6.38e}$$

$$I_{X0X1} = -\pm_{b0}\pm_{b1}m_{44} \tag{6.38f}$$

$$I_{X0Y1} = \pm_{b0}\pm_{b1}(S_{m1}m_{24} - C_{m1}m_{34}) \tag{6.38g}$$

$$I_{Y0X1} = \pm_{b0}\pm_{b1}(-S_{m0}m_{42} + C_{m0}m_{43}) \tag{6.38h}$$

$$I_{Y0Y1} = \pm_{b0}\pm_{b1}(C_{m0}C_{m1}m_{33} - S_{m0}C_{m1}m_{32} - C_{m0}S_{m1}m_{23} + S_{m0}S_{m1}m_{22}) \tag{6.38i}$$

As can be seen by Eqs. (6.38), all elements of the sample Mueller matrix are be included in the value of the intensity and can therefore be measured by the 2-MGE. The corners of the sample Mueller matrix m_{14}, m_{41}, and m_{44} are always included in the coefficients I_{X0}, I_{X1}, and I_{X0X1} and so are *always* measured (contrast this to the one-modulator PMEs where m_{44} can *never* be measured). If the azimuthal orientations of the PSG and PSA are restricted to modulo 45°, then each of the intensities will be proportional to just one of the sample Mueller matrix elements. This can be shown schematically by the following scheme, where the appropriate coefficient is placed in the sample Mueller matrix, indicating that it can be measured with the given azimuthal orientation of the PSG and PSA. If an element of the sample Mueller matrix cannot be measured, this is indicated by a large dot (•).

$$\mathbf{M}(\theta_{m0} = 0°,90°;\theta_{m1} = 0°,90°) = \begin{pmatrix} 1 & \bullet & I_{Y0} & I_{X0} \\ \bullet & \bullet & \bullet & \bullet \\ I_{Y1} & \bullet & I_{Y0Y1} & I_{X0Y1} \\ I_{X1} & \bullet & I_{Y0X1} & I_{X0X1} \end{pmatrix} \quad (6.39a)$$

$$\mathbf{M}(\theta_{m0} = 0°,90°;\theta_{m1} = \pm 45°) = \begin{pmatrix} 1 & \bullet & I_{Y0} & I_{X0} \\ I_{Y1} & \bullet & I_{Y0Y1} & I_{X0Y1} \\ \bullet & \bullet & \bullet & \bullet \\ I_{X1} & \bullet & I_{Y0X1} & I_{X0X1} \end{pmatrix} \quad (6.39b)$$

$$\mathbf{M}(\theta_{m0} = \pm 45°;\theta_{m1} = 0°,90°) = \begin{pmatrix} 1 & I_{Y0} & \bullet & I_{X0} \\ \bullet & \bullet & \bullet & \bullet \\ I_{Y1} & I_{Y0Y1} & \bullet & I_{X0Y1} \\ I_{X1} & I_{Y0X1} & \bullet & I_{X0X1} \end{pmatrix} \quad (6.39c)$$

$$\mathbf{M}(\theta_{m0} = \pm 45°;\theta_{m1} = \pm 45°) = \begin{pmatrix} 1 & I_{Y0} & \bullet & I_{X0} \\ I_{Y1} & I_{Y0Y1} & \bullet & I_{X0Y1} \\ \bullet & \bullet & \bullet & \bullet \\ I_{X1} & I_{Y0X1} & \bullet & I_{X0X1} \end{pmatrix} \quad (6.39d)$$

If the PSG (PSA) is set to ±45°, then the third column (row) of the sample Mueller matrix cannot be measured. Similarly, if the PSG (PSA) is set to 0°, 90°, then the second column (row) of the sample Mueller matrix cannot be measured.

The two most useful configurations are those shown in Eqs. (6.39b) and (6.39c), where the upper right or the lower left 2 × 2 block of the sample Mueller matrix is measured. Either of these configurations can be used to totally characterize most anisotropic samples. Measuring both configurations overdetermines most systems, but the additional information

can be used to reduce systematic errors by zone averaging. If all four configurations are measured, then the entire sample Mueller matrix can be determined, where several of the elements will be measured in more than one configuration. The most complete measurement would involve 16 measurements, where $\theta_{m0} = -45°, 0°, +45°, 90°$ and $\theta_{m1} = -45°, 0°, +45°, 90°$. The major advantage of the complete measurement is that many of the systematic errors associated with azimuthal angle alignment and modulator calibration will be eliminated to first order. For nearly all situations, the two-zone measurement is sufficient even for the most demanding measurement applications.

6.5 Waveform Analysis

The intensity waveforms for various configurations of PMEs, given in Section 6.4, contain the measured quantities for the instrument, but these measured quantities are in terms of coefficients of basis functions. For single modulator PMEs, these are the I_X and I_Y terms of Eq. (6.27), and for the 2-MGE, these are the eight coefficients of Eq. (6.36). Therefore, techniques must be developed to determine, as accurately as possible, these coefficients. This data reduction step can be done computationally, by first digitizing the waveform and then performing a Fourier-like analysis, or electronically using phase-sensitive detection. As a consequence, it is also very important to be able to perform this analysis as fast as possible.

It has already been stated that the intensity waveform from PME experiments is quite complicated. This is emphasized in Figure 6.7, which shows an example of the light intensity waveform from a single modulator PME and from a 2-MGE. The relevant parameters are shown in the figure caption, and the resulting Fourier transform is also shown in the right half of Figure 6.7. The waveform for the single modulator PME contains a dc component and Fourier components at both even and odd harmonics. The first and second harmonics are clearly the most important, but the third harmonic is also significant. The higher harmonics dampen very quickly, and the 5^{th} and higher harmonics are usually not measurable. The waveform for a 2-MGE, where there are two PEMs contributing to the intensity waveform, is even more complicated than for the single PEM ellipsometer case. Not only do the two PEMs have different operating frequencies, but the phases of the two PEMs are generally not synchronized. This has the consequence that the waveform will never actually repeat itself (although, practically speaking, it will come close). A Fourier transform power spectrum can still be calculated (see Fig. 6.7, lower right), and it shows the even and odd harmonics of each of the PEMs, as well as many of the sum and difference frequencies.

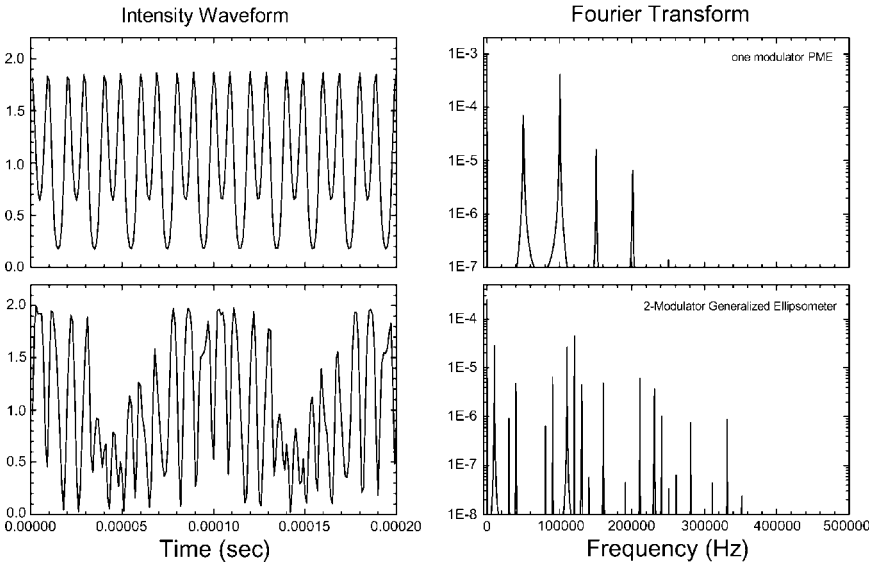


Figure 6.7 Intensity waveforms (left) and resultant Fourier transforms (right) for a one-modulator PME (top) and for the 2-modulator generalized ellipsometer (2-MGE). These waveforms were calculated for a silicon sample at 64.47° angle of incidence (N = 0.800, S = 0.344, C = −0.492) at 350 nm. The azimuthal angle of the PSG was 0° and 45° for the PSA. The PSG PEM was set to A = 2.4048, operating at 50.2 kHz with a static retardation of 0.005 rad. The PSA PEM (for the 2-MGE) was set to A = 2.4048, operating at 60.3 kHz with a static retardation of 0.025 rad. The waveforms were assumed to be digitized with a 1 MHz sampling rate and 4096 samples were collected.

Note that the sample chosen for this illustration (silicon) is isotropic; if an anisotropic sample were chosen, then even more frequencies would be observed in the Fourier transform power spectrum.

One unique aspect of PME experiments arises from the fact that the PEMs are highly resonant devices. This has been discussed in Section 6.2.2 above, and these characteristics are important in the analysis of the waveform in a PME measurement. Since the frequency is extremely stable at a given temperature, one can perform the waveform analysis *assuming* that the frequency is accurately determined. However, since the PEM is so highly resonant, the PEM is freely running, and its phase cannot be externally set. However, the phases of the PEMs can be measured from their monitor signals, which then can be used in the waveform analysis. Note that this disadvantage does not apply to rotating element ellipsometers, since the rotation of each of the elements is usually controlled externally.

For single modulator PME experiments, this usually does not present a major problem, since the reference signal from the PME can be used to synchronize lock-in amplifiers or used to trigger waveform digitizers. If there are more than one PEM in the experiment, both of which want to "control the experiment," then simple synchronization from a single resonant source is no longer possible, and some other method must be found to synchronize waveform capture with respect to the relative phases of the two PEMs.

There are two common techniques used to determine the coefficients of the basis functions from the intensity waveform of a PME experiment. The oldest technique, in common practice with single-modulator PMEs, uses phase-sensitive detection. Here the monitor signal (which is proportional to the actual dynamic retardation of the PME, possibly phase shifted) from the PEM is used as the reference signal of the phase-sensitive detection device, such as a lock-in amplifier. Another technique, pioneered by Drévillon and co-workers,[18] first digitizes the light intensity waveform and then uses Fourier analysis techniques to determine the coefficients I_X and I_Y. The second technique is not as easy to implement, but does eliminate the need for multiple phase-sensitive detectors on the same channel and does produce more information, since the higher harmonics can then be used to characterize the modulation amplitude.

6.5.1 Basis Function

From Eqs. (6.13), the basis functions X and Y are given by

$$X(t) = \sin(A\sin(\omega t)) \tag{6.40a}$$

$$Y(t) = \cos(A\sin(\omega t)). \tag{6.40b}$$

There is an implicit assumption in the formulation of this basis set that the relative phase is set to 0. In this context, the relative phase is the actual phase of the PEM with respect to the monitor signal used to synchronize the waveform analysis, which can be non-zero from phase shifts in cables and other electronics. If the phase cannot be assumed to be zero, then Eqs. (6.40a and 6.40b) no longer form a complete basis set, and a more complete statement of the basis functions would be

$$X = \sin(A\sin(\omega t + \phi)) \tag{6.40c}$$

$$Y = \cos(A\sin(\omega t + \phi)), \tag{6.40d}$$

where ϕ is the phase of the PEM with respect to the monitor signal. This phase must be carried into the integer Bessel function expansions for X and Y, also shown in Eqs. (6.6).

$$X = \sin(A\sin(\omega t + \phi)) = 2\sum_{j=0} J_{2j-1}(A)\sin((2j-1)(\omega t + \phi)), \quad (6.41a)$$

$$Y = \cos(A\sin(\omega t + \phi)) = J_0(A) + 2\sum_{j=1} J_{2j}(A)\sin(2j(\omega t + \phi)). \quad (6.41b)$$

For unknown phase, an alternative complete basis function set for a one-modulator system would be:

$$X_s = \sin(A\sin(\omega t)) \quad (6.42a)$$

$$Y_s = \cos(A\sin(\omega t)) \quad (6.42b)$$

$$X_c = \sin(A\cos(\omega t)) \quad (6.42c)$$

$$Y_c = \cos(A\cos(\omega t)). \quad (6.42d)$$

The expressions above assume that the phase is the same for the X and the Y basis functions. However, in the integer Bessel function expansion (Eqs. (6.41) above), it can be seen that X is made up of odd harmonics, while the Y basis function is made up of even harmonics. Depending upon the electronics, the relative phases for the two basis functions may be different, since phase shifts through amplifiers often depend upon frequency. Furthermore, the electronics usually will not have perfect frequency response, and the higher harmonics will be progressively attenuated compared to the 1f signal.

If it is assumed that the electronics do not introduce attenuation into the higher harmonics, the intensity of the waveform can be re-written using sines and cosines as the basis functions. From Eq. (6.27) it was shown that:

$$I(t) = I_{dc} + I_X X_\delta + I_Y Y_\delta = I_{dc} + (I_X - \delta_o I_Y)\, X + (I_{Y\,+}\, \delta_o I_X)\, Y. \quad (6.43)$$

If Eqs. (6.41) are used in (6.23), then

$$\begin{aligned} I(t) = I_{dc} &+ (I_Y + \delta_o I_X)J_0(A) + 2(I_X - \delta_o I_Y)\,[J_1(A)\sin(\omega t) \\ &+ J_3(A)\sin(3\omega t) + \cdots] + 2(I_Y + \delta_o I_X)[J_2(A)\sin(2\omega t) \\ &+ J_4(A)\sin(4\omega t) + \cdots]. \end{aligned} \quad (6.44)$$

That is, the intensity waveform has a dc component, which is $I_{dc} + (I_Y + \delta_o I_X)J_0(A)$, a first harmonic component $2J_1(A)$ $(I_X - \delta_o I_Y)$, a second harmonic component $2J_2(A)$ $(I_Y + \delta_o I_X)$, a third harmonic component $2J_3(A)$ $(I_X - \delta_o I_Y)$, etc. Although the intensity waveform in the Fourier basis set contains an infinite number of terms, there is no additional information concerning the sample, since all the odd harmonics are proportional to $(I_X - \delta_o I_Y)$ and all the even harmonics are proportional to $(I_Y + \delta_o I_X)$. However, there is information concerning the amplitude of modulation. If the coefficient of the 3rd harmonic is divided by the coefficient of the 1st harmonic, this result is $J_3(A)/J_1(A)$, from which the Bessel angle A can be determined.

6.5.2 Phase-Sensitive Detection

A common way to analyze the intensity waveform for a PME experiment (as well as many other kinds of experiments utilizing PEMs) is to convert the intensity waveform into a voltage waveform, which then can be characterized by a lock-in amplifier, where the PEM monitor signal is used as the reference input. The I_X component is measured using the 1f signal (where the frequency of the PEM is $f = 2\pi\omega$), and the I_Y component is measured using the 2f signal. As electronic phase shifts can make the phase of the intensity waveform different from the phase of the monitor signal, the phase adjustment on the lockin must be used to match the phases. The signals must then be normalized by the basis function prefactors: $2J_1(A)$ for 1f and $2J_2(A)$ for 2f and any corrections made to frequency response attenuation. The relative phase difference between the PEM and the monitor signal ϕ is corrected using the phase adjustment of the lock-in amplifier.

In most implementations of PMEs, it is assumed that the 1f and 2f signals are normalized to the dc component. If photomultiplier tubes (PMTs) are used for the waveform detector, then this can be done directly in the control circuitry of the PMT by dynamically controlling the bias voltage across the dynode chain to give constant dc current out (this was discussed in Section 6.3). It may also be useful to measure the dc current, but this can be accomplished using a simple ammeter.

If the waveform is normalized to the dc intensity, then the signals from the lock-in amplifier are given by

$$R_{1\omega} = 2\, J_1(A)\, I_X \,/\, (I_{dc} + J_0(A)\, I_Y) \tag{6.45a}$$

$$R_{2\omega} = 2\, J_2(A)\, I_Y \,/\, (I_{dc} + J_0(A)\, I_Y). \tag{6.45b}$$

The coefficients of the basis functions I_X, I_Y and I_{dc} are given in terms of the elements of the sample Mueller matrix (see Eqs. (6.31), (6.32), and (6.33) for one-modulator PMEs). The Bessel angle A is proportional to the modulator amplitude and is often set to 2.4048 radians, where $J_0(A) = 0$; this obviously simplifies the data analysis in that it eliminates the denominator dependence on the coefficient I_Y.

6.5.3 Digital Waveform Analysis

It is also possible to obtain the basis function coefficients from the intensity waveform using modern digital techniques. This is somewhat more difficult, in that the waveform must first be digitized and then the basis function coefficients obtained using Fourier-like analysis. Whereas the phase-sensitive detection scheme described above can obtain one parameter for each phase-sensitive detector, the digital analysis approach can determine many more parameters, which then can be used to increase the information content and/or the accuracy of the measurements.

It first might appear that waveform analysis from PEM experiments using a digitized waveform would be just a simple application of a fast Fourier transform (FFT). However, FFTs are best known for getting accurate values of the frequency components in a waveform, but the amplitudes of these components is usually not as accurate. The waveform from a PME experiment will have very pure frequency components because of the resonant nature of the PEMs, so knowledge of the frequencies of these components is not needed. The need is for very accurate amplitudes of components of precisely known frequencies.

One method that can give very accurate values of the amplitudes is to perform a Fourier integration over the waveforms. For a one-modulator instrument where the phase of the light intensity has been adjusted to the phase of the PEM, the analysis of the intensity equation (6.27) is quite straightforward:

$$I_{dc} + I_Y J_0(A) = \frac{1}{2\pi M} \int_0^{MT} I(t)dt \tag{6.46a}$$

$$I_X - \delta_o I_Y = \frac{1}{2\pi M J_1(A)} \int_0^{MT} \sin(\omega t) I(t)dt \tag{6.46b}$$

$$I_Y + \delta_o I_X = \frac{1}{2\pi M J_2(A)} \int_0^{MT} \cos(2\omega t) I(t)dt, \tag{6.46c}$$

where the integration is performed over M periods of time duration T. The accuracy of this technique depends mainly on using a complete number of cycles. Since the waveform is digitized, it may not be possible to find a beginning and ending point such that an integer number of cycles is included. In this case, an interpolation correction can be employed to improve the accuracy.

6.5.4 Two-Modulator Systems

Waveform analysis for two-modulator PMEs is considerably more difficult than it is for single modulator PMEs. This is primarily because each PEM is a highly resonant device, with its own frequency and phase, which cannot be controlled by external electronics. (The electronic controls are used to control the modulator amplitude, but have practically no control of the PEM frequency or phase). This results in the complicated waveform, such as is shown in the lower left of Fig. 6.7.

As noted above, this waveform does not formally repeat itself. The repeat criterion is

$$m_0 T_0 = m_1 T_1 = T_r, \tag{6.47}$$

where m_0 and m_1 are integers and T_0 and T_1 are the periods for the two PEMs. If the phases of the two PEMs are ϕ_0 and ϕ_1 at time 0, then the repeat time T_r is the time when the two phases are again ϕ_0 and ϕ_1 This equality is seldom observed for two real PEMs. To show this, rewrite Eq. (6.47) to take into account phase mismatches:

$$(2m_0\pi + \phi_0)\, T_0 /(2\pi) = (2m_1\pi + \phi_1)\, T_1 /(2\pi) = T. \tag{6.48}$$

As an example, the periods for the two modulators used in the calculations shown in Fig. 6.7 are

$$T_0 = 1/50.2 \text{ kHz} = 19.920 \text{ μs}; \; T_1 = 1/60.3 \text{ kHz} = 16.583 \text{ μs}.$$

After five complete cycles of PEM 0 ($m_0 = 5$), $T = 99.60$ μs, and after six complete cycles of PEM 1 ($m_1 = 6$), $T = 99.50$ μs. If the repeat time is set to the average, $T = 99.55$ μs, then $\phi_0 = -0.0025$ rad $= -0.14°$ and $\phi_1 = 0.0030$ rad $= 0.17°$ after five cycles of PEM 0 and six cycles of PEM 1. These two modulators come very close to repeating every 99.55 μs, but there is a small walk-off in the relative phases for multiples of the time T.

One way of dealing with this problem is by triggering the waveform capture whenever the relative phases ϕ_0 and ϕ_1 are both very close to 0 (or some other value), but not exact. The number of times that the trigger will activate will depend upon the gate time; the smaller the gate time, the fewer triggers that will take place per unit time, but the more accurate the phase.

If the intensity is rewritten in terms of Fourier coefficients, it is given by

$$I(t) = I_{dc} + 2\sum_{k=1}^{\infty} R_k[\alpha_k \cos(\Omega_k t) + \beta_k \sin(\Omega_k t)] \tag{6.49}$$

The frequencies Ω_k are the harmonics of each of the modulators, as well as the sum and difference frequencies of each of the modulators. The summation in Eq. (6.49) is over all possible frequency combinations, and so is infinite. The quantities R_k are the coefficients to be determined, and are proportional to the coefficients I and the respective integer Bessel functions. Some of the more important terms are shown in Table 6-1.

Table 6-1. Some frequency components for the Fourier integral analysis of the time-dependent intensity expression shown in Eq. (6.49). The frequency components Ω_k are calculated in the Frequency column, assuming that ω_0 = 50.2 kHz and ω_1 = 60.3 kHz. The coefficients R_k are given in terms of the coefficients of Eq. (6.37) and the integer Bessel functions. The products of the integer Bessel functions are given in the last column for $A_0 = A_1 = 2.4048$.

k	Ω_k	Frequency (kHz)	R_k	$J_x(A)\,J_y(A)$
1	$\omega_0 - \omega_1$	10.1	$I_{X0X1}J_1(A_0)\,J_1(A_1)$	0.269
2	$2(\omega_0 - \omega_1)$	20.2	$I_{Y0Y1}J_2(A_0)\,J_2(A_1)$	0.187
3	$2\omega_0 - \omega_1$	40.1	$I_{Y0Y1}J_2(A_0)\,J_2(A_1)$	0.224
4	ω_0	50.2	$I_{X0}J_1(A_0)$	0.519
5	ω_1	60.3	$I_{X1}J_1(A_1)$	0.519
6	$\omega_0 - 2\omega_1$	70.4	$I_{X0Y1}J_1(A_0)\,J_2(A_1)$	0.224
7	$2\omega_0$	100.4	$I_{Y0}J_2(A_0)$	0.423
8	$\omega_0 + \omega_1$	110.5	$I_{X0X1}J_1(A_0)\,J_1(A_1)$	0.269
9	$2\omega_1$	120.6	$I_{Y1}J_2(A_1)$	0.423
10	$3\omega_0$	150.6	$I_{X0}J_3(A_0)$	0.199
11	$2\omega_0 + \omega_1$	160.7	$I_{Y0X1}J_2(A_0)\,J_1(A_1)$	0.224
12	$3\omega_1$	180.9	$I_{X1}J_3(A_1)$	0.199
13	$4\omega_0$	200.8	$I_{Y0}J_4(A_0)$	0.065
14	$2(\omega_0 + \omega_1)$	221.0	$I_{Y0Y1}J_2(A_0)\,J_2(A_1)$	0.187
15	$4\omega_1$	241.2	$I_{Y1}J_4(A_1)$	0.065

Fortunately, the higher order terms are also accompanied by higher integer Bessel functions, so they are naturally attenuated. This can be seen in Fig. 6.7, where there are no significant terms greater than ~400 kHz. The Fourier coefficients α_k and β_k are determined from the intensity waveform by performing Fourier integration, similar to that shown in Eqs. (6.46).

By exploring Table 6-1, it can be seen that each of the eight basis function coefficients can be measured using a different frequency component, and that several of the basis function coefficients can be measured using more than one frequency component. For example, the I_{Y0Y1} component can be measured either by $k = 2$ or by $k = 14$. Since the products of the integer Bessel functions are the same for both cases, the choice is a matter of convenience. However, the I_{Y0} component can be measured by $k = 7$ or by $k = 13$, but since the integer Bessel function value for $k = 13$ is significantly smaller than it is for $k = 7$, $k = 7$ is the preferred choice. The magnitude of each component is determined by $(\alpha_k^2 + \beta_k^2)^{1/2}$. The sign of the individual components is determined by the measured phase and by the relative modulator phases. The measured phase is given by $\phi_k = \arg(\alpha_k + i\,\beta_k)$, where the arg function is the principal value of the argument of the complex number $(\alpha_k + i\,\beta_k)$ between $-\pi$ and π. For example,

$$I_{Y0} = \text{sign}(\cos(\phi_m + \phi_0))\,(\alpha_7^2 + \beta_7^2)^{1/2}\,/(2J_2(A_0)). \qquad (6.50)$$

The other components can be calculated in a similar manner.

6.6 Calibration Procedures

In order for any PME to be able to take accurate measurements, it is extremely important that the instrument be properly calibrated. Because the magnitude of both the dynamic and static retardation of the PEM depends upon wavelength, the calibration of a PEM-based ellipsometer is somewhat more involved than the calibrations for rotating analyzer (RAE) or rotating polarizer ellipsometers (RPE), but similar in complexity to the calibrations required for rotating compensator ellipsometers. The difficulty in PME calibration is often cited as a reason for using RAEs or RPEs. However, these calibrations can now be done very accurately and, with computer control, with very little user interaction, so calibration difficulty is no longer a valid reason for preferring rotating element ellipsometers over PMEs.

Each polarizer-PEM pair in the ellipsometer requires four calibrations:

1) The azimuthal angle of the polarizer with respect to the PEM θ_b,
2) The modulator static strain δ_o,
3) The amplitude of modulation A (or equivalently, the drive voltage V_m required to give a specific value of A, such as 2.4048), and
4) The azimuthal angle of the PEM with respect to the plane of incidence θ_m.

The calibrations 1) and 4) are independent of wavelength, but calibrations 2) and 3) are not. Since the PEM is such a stable device, calibrations 1), 2), and 3) need only be performed occasionally, while calibration 4) needs to be performed for each sample with the sample in place, unless the sample stage is designed such that it is possible to position the sample precisely. As discussed in Section 6.2.1, it is often appropriate to express the wavelength dependence of V_m and δ_o using two or three terms of a Cauchy expansion.

In this section, calibration procedures of one-modulator and two-modulator PMEs will be discussed. For the purposes of this discussion, it will be assumed that the measured quantities are the coefficients of the basis functions (such as I_X and I_Y) and that appropriate techniques for deconvoluting the intensity waveform to get the coefficients I_X and I_Y have been applied.

6.6.1 One-Modulator PMEs

1) Determination of θ_b

The calibration of θ_b can be done in any configuration and azimuthal angle alignment, as long as either I_X or I_Y is large when θ_b is set to $\pm_b 45°$. However, this alignment can never be set directly to $\pm_b 45°$ with a single modulator PME, but rather must first be set to 0° or 90° and then rotated to $\pm_b 45°$ using a precision rotator (this limitation does not exist for two-modulator systems). For example, in the straight-through configuration,

$$I(t) = 1 + C_b C_{m-p} - \delta_o S_b S_{m-p} X + S_b S_{m-p} Y. \tag{6.51}$$

If S_{m-p} is set to ± 1 (resulting in $C_{m-p} = 0$), then $I_Y = S_b$. This coefficient can then be nulled to establish the 0° and 90° positions, and then a precision rotator can be used to set θ_b to $\pm_b 45°$.

2) Determination of the static strain δ_o

As can be seen from Eq. (6.34), above, the straight-through configuration can be used to measure the static strain, since $I_X = -S_b S_{m-p} \delta_o$. This measurement must be done as a function of wavelength, since the static strain of any modulator will be dispersive. The wavelength-dependent strain-induced retardation can be fit to the Cauchy expansion shown in Eq. (6.8a), obtaining one or two fitting parameters. Since the strain-induced retardation is usually small for most modern PEMs, no more than two parameters are required.

3) Determination of the modulator amplitude A

The most crucial and difficult calibration of PMEs is the measurement of the modulator amplitude and the subsequent calibration of the modulator control voltage. As with the strain-induced retardation, this calibration depends upon wavelength. There are at least two ways in which this calibration can be performed.

One calibration technique sets the ellipsometer in the straight-through configuration, where $S_b S_{m-p} = \pm 1$ and the intensity is given by [see Eq. (6.34)]:

$$I(t) = I_o \left[1 + S_b S_{m-p} J_0(A) - \delta_o S_b S_{m-p} X + S_b S_{m-p} Y\right]. \qquad (6.52)$$

If the modulator is turned off, then the dc component will be proportional to $1 + S_b S_{m-p}$; that is, the intensity will be $2 I_o$ for one azimuthal orientation of the PSA polarizer and 0 for the other orientation (which will be 90° with respect to the first orientation). By adjusting the modulator amplitude, a point can be found where the dc component of the intensity is just I_o. At this point the modulator amplitude $A = 2.4048$, since $J_0(A = 2.4048) = 0$.

Another calibration technique also utilizes Eq. (6.52). Here, the modulator is set such that $A \sim 2.4048$, and the azimuthal orientation of the PSA polarizer is alternated between $S_b S_{m-p} = \pm 1$. The modulator drive voltage is adjusted such that the dc component does not change when the PSA polarizer is changed by 90°.

4) Determination of the modulator azimuthal angle θ_m and the polarizer angle θ_p

The modulator azimuthal angle θ_m of the PSG (PSA) and the polarizer azimuthal angle of the PSA (PSG) must be determined with the sample in place, since the sample is required to establish the plane of incidence. From

Section 6.4.2, we find that the I_X and the I_Y coefficients are given by

$$I_X = S_b S_{p1} S, \tag{6.31b}$$

$$I_Y = S_b[-C_m S_{p1} C + S_m(-N + C_{p1})]. \tag{6.31c}$$

Therefore, nulling I_X will establish the zero of S_{p1} (where $\theta_{p1} = 0°$ or 90°) and nulling I_Y will establish the zero of S_m (where $\theta_m = 0°$, 180°), assuming that S and N are reasonably large. This calibration can be used for either the P-PEM-S-P or the P-S-PEM-P configuration. Once this calibration is performed, any other azimuthal angle of the PSG or the PSA can be set using accurate automatic rotators.

6.6.2 Two-Modulator PMEs

The calibration needs for two-modulator systems are essentially the same as those for the one-modulator systems, except that both polarizer-modulator pairs need to be calibrated. That is, the four calibrations mentioned above for one-modulator systems must be performed for each modulator. Actually, calibrations are *easier* for two modulator systems than they are for one-modulator systems, once the basis function coefficients have been determined.

As with the one-modulator systems, the most useful calibration configuration is the straight-through configuration, where the sample is free space. In this configuration, the intensity waveform can be written as

$$\begin{aligned} I(t) = I_{dc}\,[1 &+ I_{X0}X_0 + I_{Y0}Y_0 + I_{X1}X_1 + I_{Y1}Y_1 \\ &+ I_{X0X1}X_0X_1 + I_{X0Y1}X_0Y_1 + I_{Y0X1}Y_0X_1 + I_{Y0Y1}Y_0Y_1]. \end{aligned} \tag{6.53}$$

If it is assumed that the two polarizers are set to $\pm_{b0}45°$ and $\pm_{b1}45°$, then

$$I_{X0} = 0, \tag{6.54a}$$

$$I_{Y0} = \pm_{b0}\pm_{b1}[J_0(A_1)\cos(2\theta_m) - 2\sin(2\theta_m)\,\varepsilon_{b1}], \tag{6.54b}$$

$$I_{X1} = 0, \tag{6.54c}$$

$$I_{Y1} = \pm_{b0}\pm_{b1}[J_0(A_0)\cos(2\theta_m) + 2\sin(2\theta_m)\,\varepsilon_{b0}], \tag{6.54d}$$

$$I_{X0X1} = -\pm_{b0}\pm_{b1}, \tag{6.54e}$$

$$I_{X0Y1} = -\pm_{b0}\pm_{b1}[\delta_1 + \delta_0 \cos(2\theta_m)], \tag{6.54f}$$

$$I_{Y0X1} = -\pm_{b0}\pm_{b1}[\delta_0 + \delta_1 \cos(2\theta_m)], \tag{6.54g}$$

$$I_{Y0Y1} = \pm_{b0}\pm_{b1}\cos(2\theta_m). \tag{6.54h}$$

In Equations (6.54), the azimuthal angle θ_m is the relative angle of the PSA with respect to the PSG, $J_0(A_0)$ and $J_0(A_1)$ are the 0th order integer Bessel functions near $A_0 \sim A_1 \sim 2.4048$, ε_{b0} and ε_{b1} are the errors in the azimuthal angles of the polarizers with respect to the PEMs, and δ_0 and δ_1 are the static retardations of the two PEMs.

As can be seen from the equation for the intensity given above, the first three calibrations can be performed in the straight-through configuration. If the PSG is set at ±45° with respect to the PSA (such that $\sin(2\theta_m) = \pm 1$ and $\cos(2\theta_m) = 0$), then the errors in the alignment of the two polarizers with respect to the PEMs can be set by nulling I_{Y0} and I_{Y1}. This is actually an easier alignment than for a one-modulator PME, since the polarizers can be set directly to ±45° rather than set to 0° or 90° and then moved to ±45° using a precision rotator. This also means that the rotation stage need not be calibrated or linear, so a cheaper rotation stage can be used. While the 2-MGE is set in this configuration, both static strain-induced retardations can be measured from the basis function coefficients I_{X0Y1} and I_{Y0X1}.

If the PSG is aligned with the PSA (such that $\sin(2\theta_m) = 0$ and $\cos(2\theta_m) = \pm 1$), then it is possible to measure $J_0(A_0)$ and $J_0(A_1)$, which are close to 0. If these functions are close enough to 0 such that a linear approximation can be made, then

$$A_i = 2.4048 - J_0(A_i)/0.5196,\ i = 0,1. \tag{6.55}$$

The modulator drive voltage required to give $A_i = 2.4048$ can then be determined, since it is proportional to A_i to first order [see Eq. (6.5)].

Therefore, the straight-through configuration can be used to determine the azimuthal angle of the polarizer with respect to the PEM, and the wavelength-dependent modulator calibrations of the required modulator drive voltage to give $A = 2.4048$ and a measure of the modulator static strain. These calibrations can be performed simultaneously for both modulators.

A separate calibration must be performed to determine the azimuthal angles of the PSG and the PSA with respect to the plane of incidence. Since the plane of incidence is only defined when the sample is in place, this calibration must be performed in the ellipsometry configuration. Fortunately, again the calibration for the 2-MGE is easier than it is for one-modulator systems, but the assumption must be made that the sample is isotropic, or that the optic axis of the sample is in the plane of incidence

making the cross-polarization terms 0. If this is the case, then the I_{Y0} and I_{Y1} coefficients can be used to calibrate the azimuthal angles. These coefficients are given by

$$I_{Y0} = -\pm_{b0} [N \sin(2\theta_{m0}) - \pm_{b1} C J_0(A_1)]$$
$$I_{Y1} = -\pm_{b1} [N \sin(2\theta_{m1}) - \pm_{b0} C J_0(A_0)],$$

where the coefficients N and C are the ellipsometry parameters introduced in Eq. 6.23. If it can be assumed that the modulators are calibrated (such that $J_0(A_0) = J_0(A_1) = 0$) and that N is large, then I_{Y0} and I_{Y1} can be nulled to give a precise position of $\theta_{m0} = \theta_{m1} = 0°$. Note that the azimuthal angles of both modulators can be calibrated simultaneously.

6.7 Errors

6.7.1 General Discussion

As with any measurement technique, it is very important to identify and characterize the source of errors in an ellipsometry experiment. Generally speaking, ellipsometry experiments are very accurate. In both rotating element and phase modulation spectroscopic ellipsometers, the various harmonics of the detected light intensity are normalized to the dc level. As a result, fluctuations in the lamp intensity or incomplete collection of the light from the sample do not result in an increase in the error.

Ellipsometry errors are either random or systematic. Random errors tend to be small, mainly because there generally is a lot of light in an ellipsometry experiment. These errors can be reduced by collecting more light, either by increasing the integration time or by increasing the amount of light on the sample. The analysis of random errors in ellipsometry experiments is the same as the analysis of random errors in any other experiments.

Systematic errors in ellipsometry experiments are probably the most important, but are the hardest to identify. Being systematic, simply taking more data does not improve the quality of the data. Therefore, it is important to identify as closely as possible the sources of these errors, minimize them when possible, and devise ways to take into account the error that remains.

In performing spectroscopic data analysis (see Chapter 3), it is critically important to include the measurement errors in fitting spectroscopic ellipsometry data to models. As a result, any ellipsometry experiment is incomplete until an analysis has been made of the sources of error and an attempt has been made to determine the spectroscopic errors for each data point. The random component is relatively easy to determine: Take several

measurements and calculate the mean and standard deviation. The systematic errors are considerably more difficult to quantify and to incorporate into fitting algorithms.

There are two major problems with interpreting spectroscopic ellipsometry data using fitting algorithms. First of all, systematic errors cannot be treated formally as random errors, but these errors are generally larger than random errors. A second problem is that the model equations used in spectroscopic ellipsometry are decidedly nonlinear. Both of these problems result in an uncertainty in interpreting the results of the fits. One way to address this dilemma is to estimate the systematic errors of the ellipsometry measurement as accurately as possible, and then combine them with the measured random errors. This will also involve relaxing the precise meaning of the various quantities obtained from the fitting algorithms. If the system is linear and only random errors are included, then the fitted parameters, their correlated and uncorrelated errors, cross-correlation coefficients, and the final χ^2 have precise meanings. The same formalism can be used for nonlinear systems and where some of the errors are systematic, but all the statistical parameters regarding the fit, such as the errors, cross-correlation coefficients, and χ^2 lose their precise meaning. This, of course, is not a unique problem of spectroscopic ellipsometry experiments, and the interested reader is referred to the further reading list of Chapter 3 (Section 3.8) for more information.

6.7.2 Systematic Errors of PMEs

Given that systematic errors can be so important in spectroscopic ellipsometry measurements, we will close this chapter with a list and discussion of errors and approximate magnitudes for phase modulated ellipsometers. Although all errors listed here apply to PMEs, many will also apply to rotating element ellipsometers.

1) *Misalignment of the polarizer with respect to the PEM (angle θ_b).* The alignment of the polarizer in a polarizer-PEM PSG or PSA can be set directly to ±45° very accurately in a two-modulator system, but only to 0° or 90° in a one-modulator system. Therefore, the error in θ_b in one-modulator systems will be due to two errors: the error in determining the 0° or 90° position and the error in positioning the polarizer at ±45° with respect to the PEM, once the 0° or 90° position is established. This last error depends largely on the quality of the polarizer rotator used in the instrument. In two-modulator systems, θ_b can be set directly to the ±45° position; this reduces the error in θ_b and reduces the accuracy requirements for the polarizer

rotator, decreasing the cost of the instrument without affecting the quality of the measurement. Generally, θ_b can be set to $\pm 0.01°$ to $0.02°$ if a null position can be used, while setting θ_b according to the scales of a polarizer holder can result in errors as large as $\pm 0.1°$, depending upon the system. The resultant error in θ_b is denoted as ε_b and will be ~0.0002–0.002 radians.

Note that 0.0002 radians corresponds to a tangential movement of 1 micron of a polarizer that is 1 cm in diameter! Any thermal distortions that change either the position of the polarizer or the position of the PEM can result in an error in θ_b. As can be seen in Eqs. (6.31a) and (6.32a), errors in θ_b propagate only into errors in I_{dc} for one modulator systems, but will propagate into harmonic components in two modulator systems. In all cases, the error factor is $2\varepsilon_b \times$ (sine or cosine of $2\theta_{m0}$ or $2\theta_{m1}$) $\times$ (N, S, or C), and is propagated into an error in the basis function coefficient (such as I_X or I_Y). Therefore the resultant error in the measured parameter due to errors in θ_b will be < 0.04 to 0.4%.

2) *Misalignment of the PEM with respect to the plane of incidence (θ_m).* The error in the angles θ_{m0} and θ_{m1} are denoted by ε_{m0} and ε_{m1}. As with the error ε_b discussed above, the errors ε_{m0} and ε_{m1} consist of contributions from two errors: finding the 0° or the 90° position using a nulling technique, and then re-positioning the polarizer-PEM pair using a precision rotator. Furthermore, the initial positioning of the polarizer-PEM pair is with respect to the plane of incidence. Therefore, any shift of the position of the sample can lead to an error in θ_m. Typically, this angle can be determined to $\pm 0.02°$ to $0.10°$ (0.0004 to 0.002 radians), depending upon the system. Similar to case 1, the error in the measured parameters will be < 0.04 to 0.4%.
3) *Calibration errors in the PEM amplitude*. The error in the PEM amplitude A can come from errors in the calibration procedure itself, or from drifts in the PEM over time. Furthermore, the voltage required to give a constant value of A depends on wavelength, so there is the possible error from the PEM dispersion. The error associated with the performance of the calibration procedure is highly dependent on which procedure is used. Generally, the calibration of two-modulator systems sets $J_0(A)$ to 0 ± 0.001, which corresponds to $A = 2.4048 \pm 0.0019$. One-modulator calibrations of A are less accurate. Of course, any drift of this calibration

due to temperature changes or other effects will increase this error. Errors in A contribute to errors in the measured quantities through the integer Bessel functions. If $A \neq 2.4048$, then $J_0(A) \neq 0$ and I_{dc} will contain an additional term $I_Y J_0(A)$ [see Eq. (6.44)]. Furthermore, there will be errors in $J_1(A)$ and $J_2(A)$ which will propagate directly into error in the measured quantities. Similar errors will also occur in two-modulator systems.

4) *Errors in the measurement of the PEM retardation due to static strain (δ).* The static retardation of a PEM can be measured to ~±0.001 using either one-modulator or two-modulator systems. This is a dispersive quantity, so it must also be measured as a function of wavelength. The static retardation mixes the measured parameters (see Eqs. 6.43, 6.44 for one-modulator systems) without the factor of 2 that must be included with errors in the azimuthal angles θ_m and θ_b, so an error in δ of 0.001 will result in an error of less than 0.1% in the measured quantities.
5) *Wavelength calibration errors.* Errors in the wavelength calibration of the monochromator or spectrograph can also be a source of error in any ellipsometry experiment. Even very small errors can become very important in regions of the spectrum where the measured quantities are changing very rapidly, such as might occur with a sample that has a very thick film. Wavelength calibration errors will cause an error in the registration between the experimental and calculated spectra. If the spectrum does not have regions where the measured quantities vary rapidly with wavelength, then generally this error is not as important.
6) *Wavelength spread errors.* Any monochromator or spectrograph selects a narrow range of wavelengths to present to the detector, which tends to be broadband. If this range becomes too large, then a significant error can occur due to the quasimonochromatic nature of the sampled light: each wavelength will "see" a different value of the sample parameters (such as $N(\lambda)$, $S(\lambda)$, and $C(\lambda)$), but the detector will "see" the integrated average $N_m(\lambda)$, $S_m(\lambda)$, and $C_m(\lambda)$. This will result in a quasi-depolarization, and $[N_m(\lambda)^2 + S_m(\lambda)^2 + C_m(\lambda)_2] < 1$. This error matters most when the N, S, or C parameters are changing the most rapidly with wavelength, such as for samples with very thick films.
7) *Angle of incidence.* The angle of incidence is also a critical parameter in any spectroscopic ellipsometry measurement.

In many commercial applications, the angle of incidence can be easily set because the PSG and the PSA are set on precise rotators. However, the accuracy of the angle of incidence is usually not determined by the accuracy of the rotators but rather by the accuracy of the placement of the sample at the center of the rotators. Placement of the sample can be improved if a separate alignment laser is used, reflecting off the front of the sample. One measure of the angle of incidence that can easily be performed is to use a standard sample (such as silicon or SiO_2) where the optical properties are precisely known, perform the spectroscopic ellipsometry measurement, and then fit the resulting spectroscopic ellipsometry spectrum to the angle of incidence and the surface layer thickness (this is close to SiO_2 for silicon, and surface roughness for SiO_2). The resulting accuracy of the angle of incidence for this procedure is ~0.02°.

Another error associated with the angle of incidence comes from the fact that the incident light on the sample is not perfectly collimated. Therefore, the sample is illuminated with a distribution of angles of incidence. If very small spot sizes are required, then very short focal length lenses must be employed, increasing the numerical aperture at the sample, and increasing the spread of angles of incidence incident upon the sample. Again, this can lead to a quasi-depolarization of the light beam, and is worst when there are large changes in the N, S, and C parameters, such as might occur when the sample consists of a very thick film.

8) *Sample depolarization errors.* As we have seen, depolarization can occur from many elements in the system, including the sample itself. Depolarization errors are not particular to PMEs, but PMEs have a particular advantage in that they can measure the depolarization. If the depolarization is simple, then the fraction of polarized light is given by $\beta = (N^2 + S^2 + C^2)^{1/2} \leq 1$. Therefore, if N, S, and C are all measured, it is possible to determine β. Sample depolarization can result from sample nonuniformity, sample transparency, or sample roughness. Sample nonuniformity occurs when optical characteristics of the sample are different at one part of the illumination spot than they are at another part of the spot (such as when the sample has a film of nonuniform thickness). Sample transparency depolarization occurs when the sample is transparent. In this case, the light beam reflected from the back surface is also detected, but this beam is incoherent with the beam reflected from the front

surface. Rough surfaces can also cause depolarization or cross-polarization due to the complex interaction of the light beam with the rough surface.

6.8 Further Reading and References

6.8.1 Further Reading

1. P. S. Hauge, "Recent developments in instrumentation in ellipsometry," *Surf. Sci.,* **96**, 108–140 (1980).
2. R. A. Chipman, "Polarimetry" *Handbook of Optics, Vol. II*, (2nd edition, McGraw-Hill, New York, 1995), Chapter 22.
3. R. M. A. Azzam, "Ellipsometry" *Handbook of Optics, Vol. II*, (2nd edition, McGraw-Hill, New York, 1995), Chapter 27.
4. D. S. Kliger, J. W. Lewis, and C. E. Randall, *Polarized Light in Optics and Spectroscopy*, (Academic Press, New York) (1990).
5. M. Born and E. Wolf, *Principles of Optics, sixth edition* (Cambridge University Press, 1980).
6. B. Drévillon, "Phase Modulated Ellipsometry from the Ultraviolet to the Infrared: *in situ* application to the growth of semiconductors," *Progress in Crystal Growth and Characterization of Materials*, **27**, 1–87 (1993).

6.8.2 Numbered References

1. M. M. Decker and H. Mueller, "Transmitting Data by Light Modulation," *Control Engineering*, **4**, 63 (1957).
2. M. Billardon and J. Badoz, "Birefringence Modulator," *C. R. Acad. Sci*. Ser. **B 262**, 1672 (1966).
3. S. N. Jasperson and S. E. Schnatterly, "An Improved Method for High Reflectivity Ellipsometry Based on a New Polarization Modulation Technique," *Rev. Sci. Instrum*. **40**, 761 (1969).
4. J. C. Kemp, "Piezo-Optical Birefringence Modulators: New Use for a Long Known Effect," *J. Opt. Soc. Am*. **59**, 950 (1969).
5. L. F. Mollenauer, D. Downie, H. Engstrom, and W. B. Grant, "Stress Plate Optical Modulator for Circular Dichroism Measurements," *Appl. Opt*. **8**, 661 (1969).
6. G. Baldachini and L. F. Mollenauer, "Theoretical and Experimental Analysis of a Stress-Plate Modulator for Magnetic Circular Dichroism Measurements," *Appl. Opt*. **11**, 2676 (1972).
7. J.-C. Canit and J. Badoz, "New Design for a Photoelastic Modulator," *Appl. Opt*. **22**, 592 (1983).
8. J. C. Kemp, "Symmetric Birefringent Modulator," U.S. Patent #3,867,014, issued Feb. 18, 1975.
9. W. C. Johnson, "A Circular Dichroism Spectrometer for the Vacuum Ultraviolet" *Rev. Sci. Instrum*. **42**, 1283 (1971).

10. M. M. Russel, M. Billardon, and J. Badoz, "Circular and Linear Dichrometer for the Near Infrared," *Appl. Opt.* **11**, 2375 (1972).
11. "PEM-90 Head Assemblies," PEM-90 Data Bulletin, Published by Hinds Instruments, Inc.
12. F. A. Modine and G. E. Jellison, Jr., "Errors in Polarization Measurements Due to Static Retardation in Photoelastic Modulators," *Appl. Phys. Commun.* **12**, 121 (1993).
13. J. P. Badoz, M. P. Silverman, and J. C. Canit, "New model of a photoelastic modulator with distributed birefringence," in *Polarization Considerations for Optical Systems II,* R. A. Chipman, ed., *Proc. Soc. Photo-Opt. Instrum. Eng.* **1166**, 478–488 (1989); "Wave propagation through a medium with static and dynamic birefringence: theory of the photoelastic modulator," *J. Opt. Soc. Am. A* **7**, 672–682 (1990).
14. O. Archer, E. Bigan, and B. Drévillon, "Improvements of Phase-Modulated Ellipsometry," *Rev. Sci. Instrum.* **60**, 65 (1989).
15. G. E. Jellison, Jr. and F. A. Modine, "Accurate calibration of a photoelastic modulator in a polarization modulation ellipsometry experiment," in *Polarization Considerations for Optical Systems II*, R. A. Chipman, ed., Proc. Soc. Photo-Opt. Instrum. Eng. **1166**, 231–241 (1989).
16. V. M. Bermudez and V. H. Ritz, "Wavelength-Scanning polarization-modulation ellipsometry: some practical considerations," *Appl. Opt.* **17**, 542–552 (1978).
17. G. E. Jellison, Jr. and F. A. Modine, "Optical constants for silicon at 300 and 10 K determined from 1.64 to 4.73 eV by ellipsometry," *J. Appl. Phys.* **53**, 3745–3753 (1982).
18. B. Drévillon, J. Perrin, R. Marbot, A. Violet, and J. L. Dalby, "Fast polarization modulated ellipsometer using a microprocessor system for digital Fourier analysis," *Rev. Sci. Instrum.* **53**, 969–977 (1982).
19. UVISEL ellipsometer, available from ISA/J-Y.
20. W. M. Duncan and S. A. Henck, "Insitu spectral ellipsometry for real-time measurement and control," *Appl. Surf Sci.* **69**, 9–16 (1993).
21. B. Drévillon, J. Y. Parey, M. Stchaskovsky, R. Benferhat, Y. Josserand, and B. Schlayen, *SPIE Symp. Proc.,* **1188**, 174 (1990).
22. G. E. Jellison, Jr. and F. A. Modine, "Two-channel polarization modulation ellipsometer," *Appl. Opt.* **29**, 959–974 (1990).
23. R. Anderson, "Measurement of Mueller matrices," *Appl. Opt.* **31,** 11–13 (1992).
24. G. E. Jellison, Jr. and F. A. Modine, "Two-modulator generalized ellipsometry: experiment and calibration," *Appl. Opt.* **36**, 8184–8189 (1997); *ibid.* "Two-modulator generalized ellipsometry: theory," *Appl. Opt.* **36**, 8190–8198 (1997).
25. R. Barakat, "Bilinear constraints between elements of the 4×4 Mueller-Jones transfer matrix of polarization theory," *Opt. Commun.* **38**, 159 (1981).
26. G. E. Jellison, Jr. "Windows in Ellipsometry Measurements," *Appl. Opt.* **38**, 4784–4789 (1999).

7 Multichannel Ellipsometry

Robert W. Collins

Department of Physics and Astronomy
The University of Toledo

Ilsin An, Joungchel Lee, and Juan A. Zapien

Department of Physics and Materials Research Institute
The Pennsylvania State University

7.1 Introduction

The applications of engineered materials, including thin films, multilayered films, and modified surfaces, have multiplied in recent years, and concurrently the performance requirements of these materials in passive and active devices have become more demanding. As a result, optical diagnostics for real time analysis of materials fabrication and processing have become exceedingly important.[1] This is particularly true for multilayered thin film structures fabricated from dielectrics, semiconductors, and metals having optical, electronic, and magnetic applications. Consequently, multichannel ellipsometry spanning the range from the infrared to ultraviolet wavelengths has become an important tool for real time measurement, monitoring, and control.[2,3] In a multichannel ellipsometer, an array of detector elements is employed to capture full ellipsometric spectra on a sufficiently short time scale (e.g., every 15 ms ~1 s) to follow the changes of interest in the sample structure. The array detector employed in the multichannel ellipsometer provides the advantage of high speed over single-channel detectors used in conventional scanning spectroscopic ellipsometers. In the latter instruments, a monochromator or spectrometer must be stepped sequentially from one wavelength position to the next, a time consuming process that leads to typical acquisition times for full spectra on the order of minutes. Before instrumentation issues in high speed multichannel ellipsometry are described, a brief definition of ellipsometric measurement modes is appropriate here.

In situ ellipsometry refers to measurements performed with the sample, e.g., the substrate/film structure, enclosed in a controlled environment. A film growth or surface modification process may not necessarily be in progress. The additional instrumentation considerations in progressing from ex situ to in situ measurements involve either ellipsometer connection

around a sample chamber or sample cell insertion at the ellipsometer axis. One problem that must be addressed for in situ ellipsometry is the development of new ellipsometer and/or sample alignment procedures appropriate for this more complex situation. Another problem is the design of chamber or cell windows that do not perturb the incident and reflected beam polarization states.

Real time ellipsometry refers to in situ measurements performed during film growth or surface modification. Real time feedback to the operator is not a prerequisite for real time measurement, and the data may be stored sight-unseen to be examined by the operator at a later time. The additional instrumentation considerations in progressing from in situ to real time ellipsometry involve performing the measurements at a sufficiently high speed while retaining the precision and accuracy necessary to follow the dynamical processes of interest at the sample surface. Time scales of interest accessible to conventional ellipsometers range from microseconds, for example in studies of laser processing of surfaces, to hours, for example in studies of controlled oxidation or contamination of surfaces.

In *real time monitoring*, the data collected in real time are also displayed in real time in a form that is meaningful to the operator. However, the materials processing parameters are not modified in response to the real time measurements, and the operator does not act on the displayed information. Real time monitoring may require significant real time analysis to convert the data to a meaningful form. Thus, the additional considerations in progressing from real time measurement to real time monitoring involve efficient high-speed computation for data reduction and analysis.

In *real time control*, the data collected in real time are also interpreted in real time using a process model. The resulting information is then applied to control the process in order to achieve prespecified thin film or surface characteristics, preferably by computer without operator intervention. Real time control may be as straightforward as turning off the process at a prespecified film thickness (i.e., endpoint detection), or as involved as modifying one or more process variables such as gas partial pressure, gas flow, substrate temperature, or deposition rate on a continuous basis in order to achieve the prespecified characteristics. Thus, the additional considerations in progressing from real time monitoring to real time control involve the development of a feedback loop between the ellipsometric output and the control inputs of the process instrumentation, e.g., gas flow or temperature controllers.

Essentially, any of the automatic spectroscopic ellipsometer designs that can provide 100 to 300 point spectra in (ψ, Δ) over the near-infrared to near-ultraviolet wavelength range within an hour, i.e., for ex situ measurement of materials,[4] can also be adopted for real time ellipsometry at a single selectable wavelength. In this case, the wavelength is fixed and repetitive measurements are performed on a changing surface with a time

resolution ranging from microseconds to tens of seconds, depending on the instrument design and the time scale of the surface processes of interest. As a result, no significant instrumentation redesign is required in advancing from automatic spectroscopic ellipsometry performed ex situ to single-wavelength ellipsometry performed in real time. In fact, before the first development and application of real time spectroscopic ellipsometry in ~1983,[5] the most powerful real time ellipsometric approach combined real time single-wavelength ellipsometry during the process and in situ spectroscopic ellipsometry before and after the process.[6]

In fact, among the many "multiple-measurement" ellipsometric approaches, e.g., multiple wavelength, multiple incident polarization, multiple angle-of-incidence, multiple ambient, etc.,[4,7–9] the most powerful single one for materials and surface analysis is the multiple-wavelength or spectroscopic approach. In addition, the spectroscopic approach has no fundamental limitations in the extension to real time measurements. As a result, there has been a strong motivation over the years to develop spectroscopic ellipsometry as a real time probe. Significant instrument redesign is required in order to advance from ex situ spectroscopic ellipsometry to real time spectroscopic ellipsometry, however, and this important aspect of ellipsometric instrumentation is the focus of the present chapter.

In Section 7.2, the three successful instrument designs for real time spectroscopic ellipsometry are described in the order in which they were reported: self-compensating (1983),[5] rotating element (1990),[10] and phase modulation (1993).[11] Among these three instrument designs, the rotating-element variety is now used most widely,[3] and forms the basis for at least two commercial instrument product lines.[12] As a result, Section 7.3 focuses on the rotating-element instruments in greater detail, including data collection, data reduction, and calibration. Concurrent with the development and improvement of real time spectroscopic ellipsometers in the rotating-polarizer configuration,[10,13–16] described in Section 7.3.1, more complicated and powerful designs have been demonstrated also, including the single rotating-compensator[17] (Section 7.3.2) and the dual rotating-compensator[18] (Section 7.3.3) configurations.

7.2 Overview of Instrumentation

7.2.1 Self-Compensating Designs

The first real time spectroscopic ellipsometer was based on the self-compensating principle with Faraday cell modulation.[5,19–21] Schematics of the instrument and data acquisition circuitry are shown in Figs. 7.1 and 7.2. The key optical elements include: (i) a high-pressure

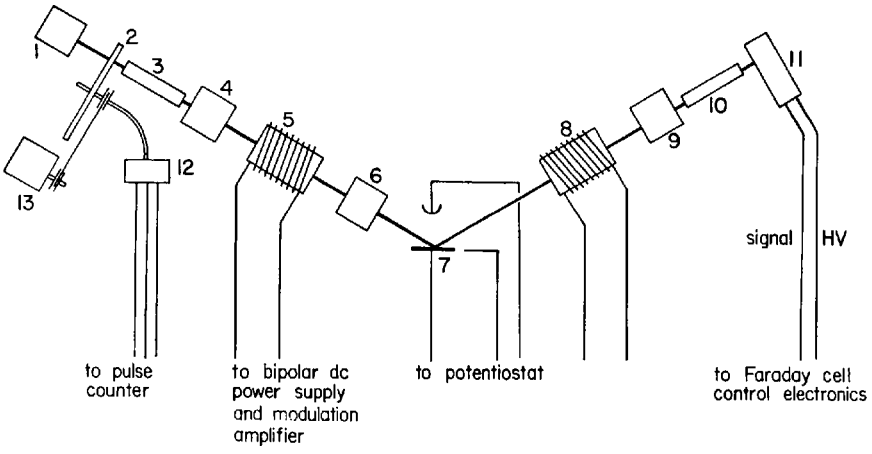


Figure 7.1 Schematic diagram of the optical elements for a self-compensating, rapid-scanning ellipsometer designed for real time spectroscopic measurements of electrochemical processes. The optical components include (1) a Xe source, (2) a rotating interference filter, (3) collimating optics, (4) a polarizer, (5) and (8) polarizer and analyzer Faraday cells, (6) an achromatic compensator, (7) a sample within an electrochemical cell, (9) an analyzer, (10) focusing optics, (11) a photomultiplier tube detector, (12) a digital encoder, and (13) the spectral scanning drive. The spectral range of this instrument spans from 370 nm (3.35 eV) to 720 nm (1.72 eV). The minimum full-spectrum acquisition time is ~3 s. [Reproduced from Ref. 19.]

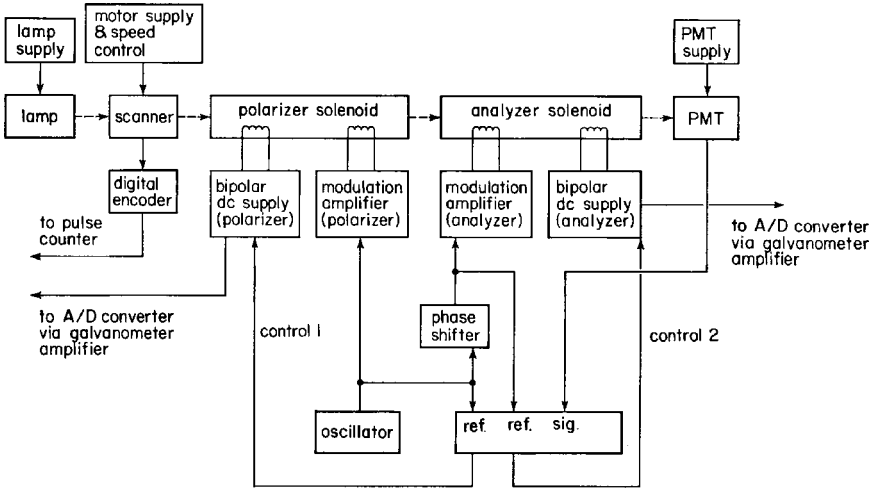


Figure 7.2 Schematic diagram of the control circuitry for the self-compensating, rapid-scanning ellipsometer of Fig. 7.1. This instrument is designed for real time spectroscopic measurements in electrochemical research. [Reproduced from Ref. 19.]

Xe lamp and collimating optics, (ii) a rotating interference filter, (iii) fixed polarizer, (iv) polarizer Faraday cell, (v) achromatic compensator, (vi) the reflecting sample (in this case inside an electrochemical cell), (vii) analyzer Faraday cell, (viii) fixed analyzer, (ix) focusing optics and photomultiplier tube detection system.[19] Operating an ellipsometer according to compensating principles involves directing a quasi-monochromatic light beam through the polarization generation and detection arms of the instrument and finding a minimum (or "null") in the detected irradiance as a function of the polarizer and analyzer azimuthal angles while fixing the compensator angle.[8] From the "nulling" values of the polarizer and analyzer angles measured with respect to the plane of incidence, the ellipsometric angles (ψ, Δ) can be determined. Typically for this measurement, the compensator is set for operation with quarterwave retardance ($\delta = 90°$) with the fast axis oriented at an azimuthal angle of $\pm 45°$ with respect to the plane of incidence. In the following two paragraphs, the basic developments in compensating ellipsometry are described that led to the advanced instrument concept of Figs. 7.1 and 7.2.

In compensating ellipsometry, the output of the detector is an approximately parabolic function of the polarizer and analyzer azimuthal angles in the vicinity of the null. Thus, the output of the detector is flat at the null, and the technique of azimuth modulation is applied in order to identify with precision the values of the polarizer and analyzer angles at the null.[8] For a manually operated instrument this means taking the average of the angular values on opposing sides of the null at which equal detector irradiance is obtained when varying either the polarizer or analyzer angle. As a result, a single compensating or so-called "null" ellipsometry measurement takes about 10 min. with a manual instrument. Automatic null or "self-compensating" instruments such as that in Figs. 7.1 and 7.2 have led to significant reductions in measurement time. A measurement time of 1 s (based on one modulation cycle) has been obtained using mechanical rotation of the polarizer and analyzer with stepping motors.[22]

Manual or automatic mechanical rotation of the polarizer and analyzer in compensating ellipsometry can be replaced by Faraday cell rotation, as in Figs. 7.1 and 7.2, that provides a much higher speed.[23] In the instrument of Figs. 7.1 and 7.2, the Faraday cells incorporate SF-6 glass cores. These cells are controlled by separate magnetically decoupled concentric windings that carry azimuth modulation currents at high frequency ($\omega_P = \omega_A = 10$ kHz) and nulling currents at lower frequency, the latter providing the angular readout from which (ψ, Δ) are computed.[19] Polarizer and analyzer modulations are separated by phase quadrature ($\phi_A - \phi_P = \pm\pi/2$), and error signals derived from

phase-sensitive detection drive the polarizer and analyzer Faraday cells to their null positions in about 1 ms, 10^3 times faster than is possible with the automatic instrument design that uses stepping motors. The actual measurement time for the self-compensating ellipsometer with Faraday cells depends on the changes in azimuth required since the rate at which the polarizer and analyzer are directed to the null is fixed at $\sim 1.4 \times 10^3$ deg/s.

The high speed with which a single-wavelength (ψ, Δ) measurement could be performed with the self-compensating instrument opened up the possibility of a real time spectroscopic ellipsometer based on this instrument principle.[19] To achieve this capability, the monochromator in the traditional ex situ spectroscopic ellipsometer design was replaced by a rotating interference filter as shown in the schematic of Fig. 7.1. With such a filter, the transmitted wavelength increases linearly with the angular orientation of the filter from about 370 nm (3.35 eV) to 720 nm (1.70 eV) during the first half turn of the filter and then decreases linearly back to 370 nm during the second half turn. Two measurement modes were developed. In the high speed mode, the spectral scanning rate is 114 nm/s (~0.16 Hz filter rotation frequency), leading to a full spectrum acquisition time of $t_s \sim 3$ s, with a potential repetition time of 3 s. In the low speed mode, the spectral scanning rate is 23 nm/s (0.03 Hz), leading to an acquisition time of $t_s \sim 15$ s, with a potential repetition time of ~15 s. An encoder on the rotating interference filter generates one index or trigger pulse and 400 clock pulses per rotation, leading to a possible 200 spectral positions in (ψ, Δ). In the time interval between one encoder clock pulse and the next (~15 and 75 ms for the fast and slow modes, respectively), a number of measurements of the null position can be collected and averaged for noise reduction.

The factor of ~200 difference between the single wavelength acquisition time ($T \sim 15$ or 75 ms) and the full spectrum acquisition time ($t_s \sim 3$ or 15 s) for the self-compensating spectroscopic ellipsometer system is characteristic of a serial wavelength scanning approach.[19] With such an approach $t_s = D_E T$, where D_E is the total number of spectral positions collected. Thus, for a given full spectrum acquisition time t_s, there is a trade-off between the number of spectral positions collected and the time available for signal averaging. Specifically, as the number of spectral positions increases, the signal-to-noise ratio or precision at a given wavelength decreases. The resolution in establishing the nulling values of the polarizer and analyzer angles with the self-compensating instrument of Figs. 7.1 and 7.2 limits the resolution in (ψ, Δ) to (0.025°, 0.05°);[21] however, no detailed characterization of the precision of the instrument in the different operating modes has been reported.

7.2.2 Rotating-Element Designs

An alternative design for real time spectroscopic ellipsometry is based on the rotating-polarizer principle.[10,24–26] A simplified schematic of the instrument is shown in Fig. 7.3. The key optical elements include: (i) a high-pressure Xe lamp and collimating optics, (ii) a polarizer mounted within the hollow shaft of a dc motor for continuous rotation, (iii) the reflecting sample, (iv) a fixed analyzer (within a stepping-motor-controlled rotator for calibration), (v) focusing optics and prism spectrograph, and (vi) a linear photodiode array detection system.[24] In contrast to the self-compensating ellipsometer, the rotating-polarizer ellipsometer and its symmetric counterpart, the rotating-analyzer ellipsometer, are both photometric instruments.

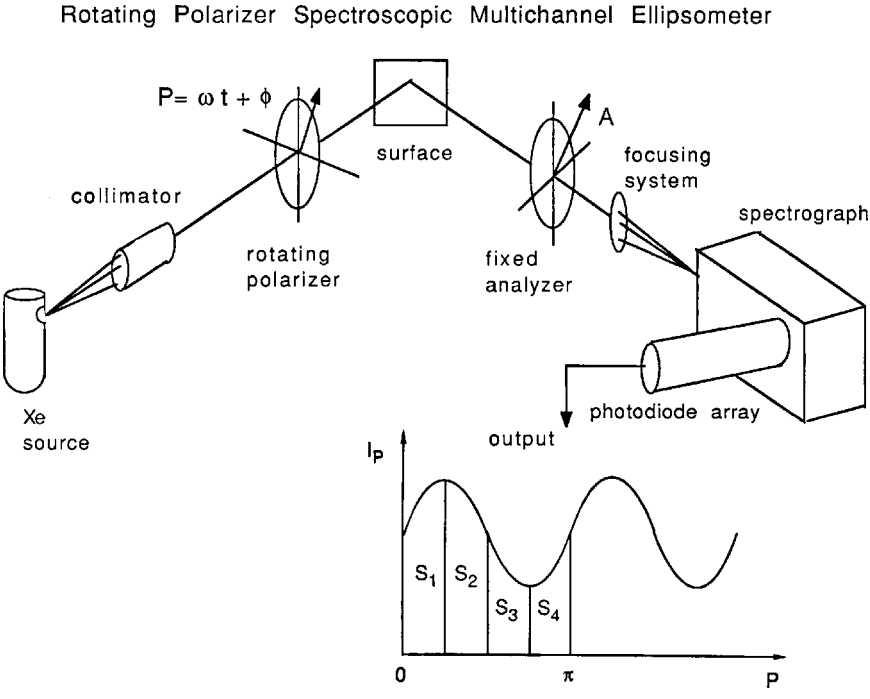


Figure 7.3 Simplified schematic of a rotating-polarizer multichannel ellipsometer designed for real time spectroscopic ellipsometry studies of thin film nucleation and growth. The source consists of a collimated Xe lamp, and the detection system includes a prism spectrograph and a linear array of silicon photodiodes. The typical spectral range for this instrument is 310 nm (4.0 eV) to 825 nm (1.5 eV), obtained in studies of the vacuum deposition of thin films. The minimum reported full-spectrum acquisition time for such an instrument is 16 ms. [Adapted from Ref. 25.]

This means that the ellipsometric angles are determined from detector readings obtained at specific values of the polarizer and analyzer azimuthal angles.[8] Operating an ellipsometer according to rotating-polarizer or analyzer principles typically involves directing a quasi-monochromatic light beam through the polarization generation and detection arms of the instrument and measuring the output of the detector at a series of equally spaced polarizer or analyzer angles, respectively, over a single half-rotation. From the angular values and the detector output values, the ellipsometric data (ψ, Δ) can be computed. In the following paragraphs, the various considerations that led to the optimum design of Fig. 7.3 are discussed.

First, it should be noted that an ellipsometric measurement at a single wavelength can be performed manually with a rotating-polarizer or analyzer instrument using the detector outputs from a minimum of three different angular settings[27] in a typical time of about 10 min. Automatic data acquisition along with data analysis by Fourier methods[28] have led to a significant reduction in the measurement time as described next.[29] In the automatic rotating-polarizer and analyzer instruments, an encoder with two outputs is attached to the motor shaft that continuously drives the rotating element.[30,31] The first output provides one pulse per rotation and is used to initiate overall data acquisition. The second output provides a series of equally spaced clock pulses (typically 10–100) over a 180° rotation of the optical element and is used to initiate sampling of the detector. A Fourier analysis of the resulting detector output as a function of the polarizer or analyzer azimuth provides the dc and 2ω Fourier coefficients of the irradiance waveform at the detector, where ω is the angular frequency of optical element rotation (either the polarizer or the analyzer). The two normalized ac Fourier coefficients, obtained by taking the ratios of the ac to the dc coefficients, can be used to determine the ellipsometric angles. For typical optical element rotation frequencies $\omega/2\pi$ of 10 to 100 Hz, a single pair of ellipsometric angles (ψ, Δ) can be obtained in a minimum time π/ω of 5 to 50 ms.

The high speed with which a single-wavelength (ψ, Δ) measurement could be performed with the rotating-polarizer and analyzer instruments ultimately led to real time spectroscopic ellipsometers based on such principles.[24,32,33] For ex situ spectroscopic ellipsometry measurements with an automatic instrument, the rotating analyzer is the preferred configuration. In this configuration, the source and monochromator are placed before the fixed polarizer, and an end-viewing photomultiplier tube detector is placed after the rotating analyzer.[31] As a result, this design completely eliminates potential errors due to source polarization, which occurs most strongly when a rotating polarizer is placed immediately after the monochromator.[34] In addition, this design minimizes potential errors due to detector polarization sensitivity which occurs most strongly when the rotating analyzer is placed immediately in front of a spectrometer, but is much reduced when the

rotating analyzer is placed immediately in front of an end-viewing photomultiplier tube.[35] One simple way to adapt the conventional ex situ rotating-analyzer spectroscopic ellipsometer to real time operation is to replace the conventional monochromator by a rotating interference filter, as was done for the self-compensating instrument.[5] With a rotating-analyzer frequency of $\omega_A/2\pi = 100$ Hz, this serial wavelength scanning approach would provide a minimum acquisition time of $t_s = 0.5$ s for 100 spectral points.

An improvement over high-speed serial wavelength scanning is provided by the novel multichannel design of Fig. 7.3 in which fully parallel spectroscopic data acquisition is possible.[24] In fact, the rotating-element ellipsometer designs are well-suited for the incorporation of multichannel detection systems owing to the similar time scales of the optical element rotation and the detector readout. To achieve this goal, the monochromator on the source side and the end-viewing photomultiplier tube on the detector side of the ex situ rotating-analyzer spectroscopic ellipsometer are replaced with a spectrograph and linear photodiode array, both at the detector side of the instrument. With this detection system design, the ellipsometer configuration is reversed as shown in Fig. 7.3. Specifically, the analyzer is fixed to eliminate potentially severe problems associated with the polarization sensitivity of the detection system, and the polarizer is rotated continuously. Thus, in this fully parallel design, source polarization associated with the Xe lamp and collimating optics remains an unavoidable error that requires careful minimization and/or correction.[34] Fully parallel spectroscopic data acquisition requires operating the photodiode array in the integrating mode, meaning that a single readout of the array provides spectral values proportional to the integral of the irradiance over the exposure time t_e, i.e., the time interval since the previous readout.[36] Using the encoder outputs from the rotating polarizer for timing, a minimum of three photodiode array readouts are acquired over a half-rotation of the rotating polarizer. From these readouts, spectra in the normalized $2\omega_P$ Fourier coefficients and can be computed; then the spectra in (ψ, Δ) can be determined.

For the acquisition of N photodiode array readouts per half rotation of the polarizer and for a typical readout time of $t_r \sim 5$ ms for a 1024 pixel photodiode array, the maximum polarizer angular rotation frequency is $\omega_P = \pi/(0.005N)\text{s}^{-1}$. For higher frequencies, the array readout will not be completed in the time interval between successive encoder clock pulses, and data acquisition will fail. Thus, the minimum acquisition time of $t_s \sim 15$ ms for full spectra in (ψ, Δ) is achieved with $N = 3$ and $\omega_P/2\pi \sim 33$ Hz. For the specific rotating-polarizer multichannel ellipsometer of Fig. 7.3, a 1024 pixel photodiode array was used in an 8x pixel grouping mode for 128 possible spectral positions and a readout time of $t_r = 4.5$ ms.[24] In addition, the polarizer rotation frequency

was set at $\omega_P/2\pi \sim 12.5$ Hz, and $N = 4$ readouts of the array were performed over a single half-rotation of the polarizer. As a result, the exposure time was $t_e = 10$ ms and the minimum full spectrum acquisition time was $t_s = 40$ ms. A prism spectrograph avoided the complications of higher order diffractions associated with grating spectrographs and provided a relatively wide spectral range of 310 nm (4.0 eV) to 825 nm (1.5 eV) in studies of vacuum deposition of thin films. As described in Section 7.2.4, recent efforts have led to a significant extension of this range deeper into the ultraviolet.[16]

In the first applications of the rotating-polarizer multichannel ellipsometer, the photodiode array outputs over 80 successive half-rotations of the rotating polarizer were averaged to improve the precision, for a total acquisition time of $t_s = 3.2$ s.[24] The repetition time between successive spectra was set at 7.5 s. In this time, the ellipsometric angles (ψ, Δ) were computed, displayed in graphical form, and stored on disc. The 7.5 s repetition time also included an intentional delay to better match the repetition time to the monolayer growth time for the specific applications. With signal-averaging over 80 half-cycles, the standard deviations in (ψ, Δ) for successive measurements on a stable metal surface were reported to be (0.003°, 0.007°) at 500 nm.[13] These values are a factor of ~8 lower than the resolution in (ψ, Δ) provided by the self-compensating real time spectroscopic ellipsometer. The later applications of the rotating-polarizer instrument exploited its high-speed capabilities.[13] For example, by setting the polarizer rotation frequency at $\omega_P/2\pi = 31$ Hz, again with $N = 4$, and by increasing the pixel grouping to 16x (for 64 possible spectral positions and $t_r = 3.5$ ms), a full spectrum acquisition time of $t_s = 16$ ms and a repetition time of 32 ms were achieved. Under these conditions, the standard deviations in (ψ, Δ) at 500 nm were (0.04°, 0.12°) for a semiconductor surface during etching. For such high speed acquisition, the waveform integrations ($N = 4$ per half-cycle of the polarizer) were stored in the memory of the photodiode array controller and no real time graphical output was provided.

It should be emphasized that for this ellipsometer design, all pixel groups of the photodiode array are simultaneously integrating the incident irradiance during the exposure time t_e. As a result, the single-wavelength (ψ, Δ) data and full spectra acquisition times are equal, i.e., $t_s = T = Nt_e$. This is in contrast to serial scanning instruments in which $t_s = D_E T$. As noted in Section 7.2.4, a fully parallel real time spectroscopic ellipsometer achieves a higher precision in (ψ, Δ) by a factor of $D_E^{1/2}$ relative to the serial scanning ellipsometer, assuming that both detectors have similar quantum efficiencies and operate within a shot-noise limited regime (i.e., at relatively high light levels where the dark current is a small fraction of the photon-generated current).

7.2.3 Phase-Modulation Designs

A third type of real time spectroscopic ellipsometer is based on the phase-modulation principle.[11,37] A simplified schematic of such an instrument is presented in Fig. 7.4. The major system components include: (i) a Xe source with collimating optics, (ii) a fixed polarizer, (iii) a photoelastic modulator, (iv) the reflecting sample, (v) fixed analyzer, (vi) focusing and fiber optics with a grating spectrograph, and (vii) a linear photodiode array. The phase-modulation ellipsometer is a photometric instrument since its operation typically involves directing a quasi-monochromatic light beam through the polarization generation and detection arms of the instrument and measuring the output of the detector as a function of time over several periods of the modulator.[8] From the irradiance waveform characteristics, the ellipsometric angles (ψ, Δ) can be computed. In the next paragraph, the phase-modulation approach is described with additional details.

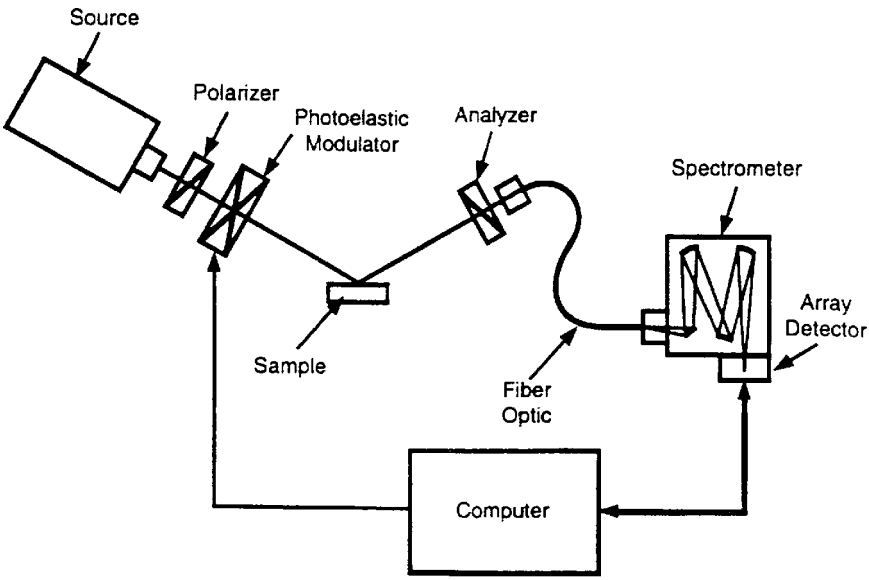


Figure 7.4 Simplified schematic of a multichannel ellipsometer based on the phase modulation principle. The source consists of a collimated Xe lamp, and the detection system includes a grating spectrograph and a linear array of silicon photodiodes. This instrument was designed for real time monitoring and control of semiconductor materials processing, spanning a spectral range from 450 nm (2.75 eV) to 750 nm (1.65 eV). The minimum reported full-spectrum acquisition time for this instrument is ~75 ms. [Reproduced from Ref. 37.]

A single measurement of (ψ, Δ) can be performed with the phase-modulation ellipsometer using lock-in techniques that provide the amplitudes of the ω_M and $2\omega_M$ components of the irradiance waveform at the detector, where ω_M is the modulator frequency, commonly 50 kHz. Lock-in detection is best suited for automatic ex situ spectroscopic ellipsometry.[39,40] An alternative approach takes advantage of the high modulation frequency, however, and yields high speed data acquisition, which is best suited for real time measurements at a single wavelength.[41] In this approach, the waveform is sampled at high rates, and the resulting data are analyzed by Fourier methods in order to provide (ψ, Δ) measurement times that, in principle, could be as short as 20 μs. Because of the short waveform sampling time, however, averages over more than 100 modulation cycles are required in order to attain a precision suitable for characterizing thin film and surface processes at the monolayer level of sensitivity. In the highest speed phase-modulation design, clock pulses generated at a rate of 12.8 MHz are used to trigger waveform sampling 256 times over a single modulator cycle.[41] These pulses are derived from the modulator reference signal using a frequency multiplication scheme. A Fourier analysis of the resulting detector output as a function of time provides the dc, ω_M, and $2\omega_M$ Fourier coefficients of the irradiance waveform. The two normalized ac Fourier coefficients, obtained by taking the ratios of the ac to the dc coefficients, are used to determine (ψ, Δ).

The high speed with which a single-wavelength (ψ, Δ) measurement could be performed with the phase-modulation ellipsometer ultimately led to real time spectroscopic instruments based on this design, as well.[11] Considering ex situ spectroscopic ellipsometers based on the phase-modulation design for the moment, the monochromator or spectrometer can be placed either at the polarization generation or polarization detection sides of the instrument, respectively. This flexibility arises because both the polarizer and analyzer are fixed, and thus errors due to source polarization or polarization sensitivity of the detection system are avoided. In fact, two approaches have been reported for real time ellipsometry at multiple wavelengths, based on the phase modulation design with a grating spectrograph placed after the fixed analyzer.[11,42] In the first approach, an array of as many as 16 fiber optic cables is mounted at the exit focal plane of the spectrograph, with each cable leading to a separate photomultiplier tube.[42] With such a small number of wavelengths available, this approach cannot be legitimately described as "spectroscopic" but rather "multi-wavelength". In the second approach, which is truly "spectroscopic", a photodiode array is attached to the spectrograph, allowing 40-point spectral coverage from 450 nm (2.75 eV) to 750 nm (1.65 eV).[11] Although this relatively narrow spectral range eliminates the need for order-sorting filters mounted over the window of the photodiode array element, it limits the overall system capabilities in the analysis of complex materials fabrication processes.

The adaptation of the photodiode array to the phase-modulation instrument does not lead to a fully parallel spectroscopic measurement because the typical readout time of the array (~1 ms) is much slower than the phase-modulator period (20 μs). As a result, the array cannot be operated in an integrating mode as is possible for the rotating-polarizer instrument. Instead, the pixel outputs are multiplexed into common signal processing circuitry.[11] The output of a given pixel is digitized 20 times per modulation period (i.e., at 1 MHz), and a fast Fourier transform provides the amplitudes of the dc, ω_M, and $2\omega_M$ frequency components of the irradiance waveform from which (ψ, Δ) can be determined. In principle, spectra in (ψ, Δ) with 40 spectral positions can be collected in 0.8 ms. In practice, however, averages over at least 100 modulation cycles are required to obtain sufficient precision in (ψ, Δ) for monolayer-level sensitivity in measurements of thin films and surfaces. No detailed characterization of the precision of the photodiode array-based phase-modulation ellipsometer has been reported under different operating conditions. As a result, it is difficult to compare directly the performances of the real time spectroscopic ellipsometers based on self-compensating, rotating-polarizer, and phase-modulation principles. In spite of this difficulty, in the next part a brief discussion is presented comparing the strengths and weaknesses of the three different instruments in general terms.

7.2.4 Design Comparisons

The first key issue in the comparison of real time spectroscopic ellipsometer designs concerns the data acquisition mode, which determines the measurement interval T and the ultimate precision. A study of the limiting precision accessible to the different ellipsometer designs has led to expressions of the form:[4,43]

$$\delta\Delta = c_1 \sqrt{\frac{1 + c_2\gamma + c_3(n_d/\langle n_i\rangle)}{N_i}}, \tag{7.1a}$$

$$\delta\psi = c_4\delta\Delta. \tag{7.1b}$$

In these expressions $N_i = \langle n_i\rangle T$, where N_i and $\langle n_i\rangle$ are the total number of events registering at the detector in the measurement time interval T and their average rate of arrival, respectively. In addition $n_d = i_d/e$, where i_d and e are the detector dark current and electron charge, and γ is the extinction ratio of the optical system, i.e., the ratio of the leakage at null to the maximum throughput of the optical system. (γ is non-zero due to optical component imperfections.) Finally, c_i ($i = 1, 2, 3, 4$) are instrument-dependent constants; for example, $1.58 \leq c_1 \leq 2.32$ and $0.4 < c_4 \leq 0.5$ for the three

types of instruments. For the rotating-polarizer and phase-modulation ellipsometers $c_2 = 1$ and $c_3 = 2$, whereas for the self-compensating ellipsometer $c_2 = c_3 = 4/(5A_m^2)$, where A_m is the analyzer azimuth modulation amplitude (in radians), which is assumed to be equal to P_m the polarizer azimuth modulation amplitude. If the irradiance is high enough and the extinction ratio is low enough so that the unit term under the square root in Eq. (7.1a) dominates, then the precision is limited by shot noise (i.e., photon statistics). For a fixed full spectrum acquisition time of t_s, the measurement interval T is shorter by a factor of D_E for a serially scanning instrument. As a result, the precision in (ψ, Δ) is poorer by a factor of $D_E^{1/2}$ and for the typical 100 spectral positions in (ψ, Δ), an order of magnitude improvement in precision is attainable with a parallel instrument design. Thus, as noted earlier, the rotating-polarizer multichannel ellipsometer maintains an important advantage since it is the only one that operates in a fully parallel mode.

The second key issue concerns the spectral range/resolution of the real time spectroscopic ellipsometer. Near the upper and lower limits of the spectral range of the instrument, the irradiance at the detector is very weak. As a result, $<n_i>$ can approach n_d in Eq. (7.1a) and the precision as well as the accuracy of the instrument are degraded. Furthermore, ambient and stray light sources as described in Section 7.2.5 in effect increase the extinction ratio γ in Eq. (7.1a), and also degrade precision and accuracy. Because the issue of spectral range/resolution is so system specific, a two paragraph digression on this issue will focus on the three instrument designs of Sections 7.2.1 through 7.2.3.

For the self-compensating instrument of Fig. 7.1, the spectral range [370–720 nm (1.70–3.35 eV)] is limited by the design of the rotating filter.[19] Certainly a broader range filter is possible, however, the prospect of extending the lower wavelength limit deeper into the ultraviolet is restricted by absorption within the ~15 cm long glass cores of the Faraday cells. Furthermore, the prospect of extending the upper limit further into the near-infrared is restricted due to decreases in the output of the Xe source and in the efficiency of the photomultiplier tube. In addition to restricting the spectral range, the rotating filter of Fig. 7.1 also limits the spectral resolution. The minimum possible wavelength bandwidth varies from ~6 nm (0.05 eV) at 370 nm to ~16 nm (0.04 eV) at 720 nm. In contrast, the rotating-polarizer and phase-modulation ellipsometers offer the possibility of wider spectral ranges provided suitable care is taken in the choice of optical components. In fact, it is easiest to implement a wide spectral range for the rotating-polarizer system since it is essentially achromatic, meaning that the optical elements, i.e., polarizer and analyzer in this case, do not have an intrinsic wavelength-dependent response.

Consequently, a rotating-polarizer multichannel ellipsometer has been developed recently to span the range from 190 nm (6.5 eV) to 825 nm

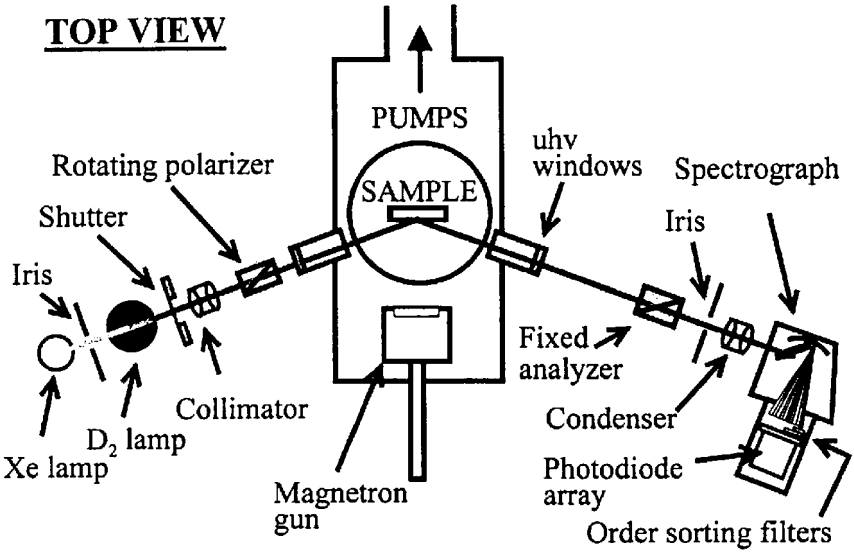


Figure 7.5 Schematic diagram of a rotating-polarizer multichannel ellipsometer designed for a wide spectral range from 190 nm (6.5 eV) to 825 nm (1.5 eV). The source consists of a Xe lamp and a see-through D_2 lamp in tandem. The detection system includes a grating spectrograph and a linear array of silicon photodiodes, the latter with order-sorting filters mounted on its surface. The minimum reported full-spectrum acquisition time is ~25 ms. [Reproduced from Ref. 16.]

(1.5 eV) with a minimum acquisition time of ~25 ms for full spectra in (ψ, Δ).[16] This instrument is shown schematically in Fig. 7.5. Three major modifications over previous instruments are incorporated into the new design. First, with the incorporation of a see-through deuterium (D_2) lamp, a tandem Xe/D_2 source configuration is adopted for a relatively flat spectral output as shown in Fig. 7.6. This figure demonstrates the advantage of the D_2 lamp in extending the lower spectral limit from 310 nm (4.0 eV) to 190 nm (6.5 eV), essentially doubling the photon energy range for the characterization of semiconductor materials and thin films. Second, MgF_2 Rochon polarizers are used for high transmission in the ultraviolet. Third, a spectrograph with a concave grating blazed at 250 nm serves to flatten the spectral response of the overall system, and two stages of internally mounted order-sorting filters serve to eliminate higher order grating diffractions. The positioning of these filters is shown at the top in Fig. 7.6, and their transmission spectra are shown in the inset. With these three modifications, 132-point (ψ, Δ) spectra have been collected in ~50 ms with wavelength detection bandwidths of 2 nm (0.08 eV) at 190 nm and 11 nm (0.02 eV) at 825 nm. Because this instrument incorporates a 1024-pixel

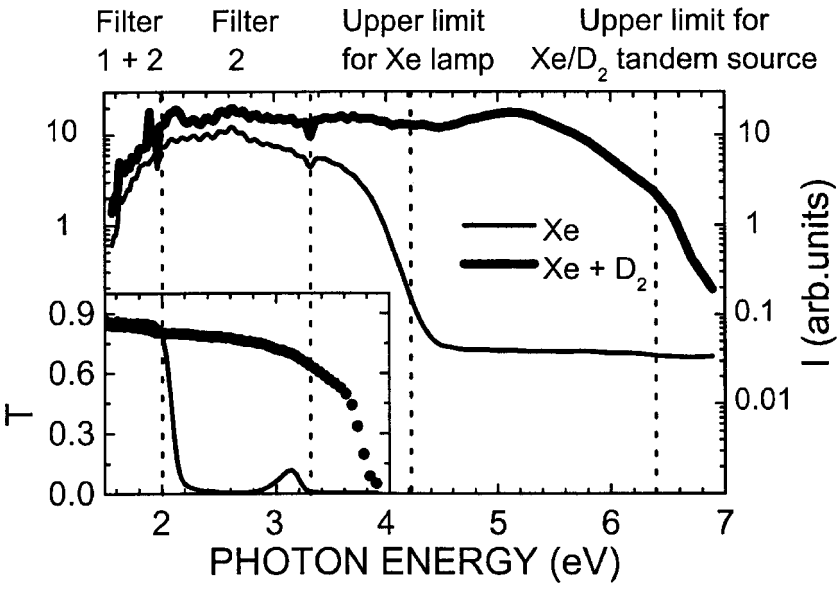


Figure 7.6 Unprocessed output of the linear photodiode array (in arbitrary units) for the rotating-polarizer multichannel ellipsometer of Fig. 7.5, plotted as a function of photon energy. These results were obtained in the straight-through $P_r(\omega_P)A$ configuration for a Xe lamp alone (light line) and for a tandem Xe-D_2 source (bold line). The inset shows the measured transmittance of order-sorting filters #1 (line) and #2 (points) mounted at the surface of the array so as to cover the energy ranges as shown at the top. [Reproduced from Ref. 16.]

linear photodiode array, it provides greater flexibility in controlling the spectral resolution. For example, the bandwidths reported here have been obtained with nonuniform pixel grouping, varying from 4x at short wavelengths to 16x at long wavelengths, in order to establish a more uniform photon energy resolution for the instrument. Thus, in principle, the wavelength bandwidths can be reduced by factors of 4 and 16 at short and long wavelengths, respectively.

The third key issue in the comparison of the three instrument designs of Figs. 7.1 through 7.4 is measurement accuracy. The advantage of the self-compensating ellipsometer design is that it retains high accuracy over the full ranges of (ψ, Δ) (i.e., $0 \leq \psi \leq 90°$; $-180° < \Delta \leq 180°$). By employing an achromatic compensator before the sample in Fig. 7.1, high accuracy is retained throughout the spectral range of the instrument, as well. The disadvantages of the two photometric instruments are their sensitivities to specific photodiode array detector errors such as nonlinearity, image lag (persistence), and stray light (see Section 7.2.5). Both

photometric instruments also exhibit ranges in (ψ, Δ) for which the sensitivity of the instrument is poor. Because the rotating-polarizer and analyzer instruments measure $\cos\Delta$, their accuracy and precision are poor when $\Delta \sim 0°$ and $\pm 180°$. As a result, this design leads to difficulties when measuring film growth on dielectric substrates for which $\Delta \sim 0°$ over a wide spectral range, or when measuring thick dielectric films on semiconductor or metal substrates for which Δ periodically crosses $0°$ and $\pm 180°$ as a function of wavelength. Because the phase-modulation ellipsometer of Fig. 7.4 measures either $\sin 2\psi$ for the fixed polarizer, modulator, and analyzer angles of 45°, 0°, and 45°, respectively (configuration II), or $\sin\Delta$ for the corresponding angles 0°, 45°, 45°, respectively (configuration III), then its accuracy and precision are poor either when $\psi \sim 45°$ (configuration II) or when $\Delta \sim \pm 90°$ (configuration III).[39] As a result, this design has difficulties in measuring film growth on metallic substrates, or in measuring thick dielectric films with strongly modulated (ψ, Δ) spectra.

Another disadvantage of the rotating polarizer and analyzer instruments mentioned earlier is their susceptibility to errors due to source polarization or detector polarization sensitivity that must be either eliminated or corrected in calibration and data reduction.[34,35] The advantage of these instruments is the simplicity of their designs compared with the self-compensating and phase-modulation instruments. The rotating polarizer and analyzer instruments incorporate no wavelength-dependent elements that would require separate calibrations to determine their characteristics. For the self-compensating ellipsometer, the Faraday cells as well as the compensator exhibit wavelength-dependent responses, whereas for the phase-modulation instrument, the modulator exhibits a wavelength-dependent response. Accuracy in determining these responses and their sensitivity to factors such as beam alignment and ambient temperature can limit the overall instrument accuracy. As will be described in Section 7.3, many of the disadvantages of the rotating polarizer and analyzer instruments can be eliminated by changing the optical configuration to rotating compensator.[17,44,45] As a result, the instrument complexity increases; however, it retains high accuracy over the full range of (ψ, Δ), and thus provides an advantage over the phase-modulation design.

7.2.5 Errors Unique to Multichannel Detection Systems

Parallel acquisition of (ψ, Δ) spectra at high speeds in spectroscopic ellipsometry requires the incorporation of a spectrograph and photodiode array into the instrument design. Thus, multichannel spectroscopic ellipsometry presents a unique set of challenges since the successive outputs

at each pixel, obtained in high-speed repetitive readouts of the array, must be analyzed photometrically to extract the waveform characteristics at each wavelength. No other applications of the photodiode array pose such challenges. As a result, detailed characterizations of spectrograph and photodiode array systems have been undertaken to assess their performance under real time ellipsometric measurement conditions. Here potential sources of errors are briefly described, and methods for their correction are outlined. Further details can be found in the references.[34,46,47]

Dark current and ambient light (or background). With a shutter blocking the incident beam to the reflecting sample, there are two sources of the remaining counts that register at the detector. These include counts from the detector dark current and counts from ambient light. Both sources can be measured before the onset of real time data acquisition by closing a shutter in front of the light source. As long as the two sources show no time dependence during the experiment, this background readout can be used to correct all successive detector readouts. In many cases, the narrow acceptance cone of the detection system along with the light rejection by the spectrograph lead to ambient light levels that are negligible compared to the dark current. In this case, a shutter is not needed and the dark current correction routine provided in commercially available photodiode array detection systems can be used. At the opposite extreme, if ambient light levels are very high, e.g., for thin film depositions at high substrate temperature or with strongly emitting excitation sources, then it may be necessary to perform alternate measurement cycles with a shutter open and closed. In this way, the detector readouts obtained with the shutter open can be corrected for variations in the background level as a function of time obtained with the shutter closed. Although this continuous background measurement significantly reduces the repetition rate at which (ψ, Δ) spectra can be obtained, it has been successful in a number of demanding thin film growth applications. For example, rotating-polarizer multichannel ellipsometry has been performed during the growth of diamond thin films at 900 °C using high power (500 W) microwave plasma excitation of carbon-containing gases.[48]

Detector nonlinearity. Any detection system used with a photometric ellipsometer must be linear to avoid experimental errors. Nonlinearity corrections are often difficult to establish uniquely for single-channel photomultiplier tube detection systems owing to the gain variations that are used when measuring different surfaces at different wavelengths.[31] For typical commercial photodiode array systems, the gain is fixed and nonlinearity corrections are straightforward to determine and implement. Nonlinearity can be assessed by illuminating the array with different light levels and reading out the array with different exposure times, while

applying the appropriate background corrections for all readouts. From such data, the measured count rate can be plotted as a function of the number of counts detected for the different pixels. These plots should be constant for each light level and for each pixel, independent of the number of counts. Any variations can be corrected with a factor that depends on the number of counts (and possibly on the pixel number), but not on the count rate. Nonlinearity effects have been found to vary extensively from one commercial photodiode array system to another.[46,47] Some systems have been found to require no linearity corrections.[16]

Detector lag (or image persistence). Image persistence occurs in successive readouts of the photodiode array when photon counts that register during one exposure remain to affect the next exposure. This effect can be characterized by blocking the beam incident on the array using a high speed shutter while continuously reading out the array, setting the same exposure conditions as in the ellipsometric measurements. In such an experiment, pixels for which the shutter is fully closed during the entire exposure time should read zero counts after background correction. In fact such an experiment reveals that some fraction of the counts persist from the previous readout to affect the subsequent readout. This fraction varies typically from 0.1 to 1% of the count level of the previous readout, depending on the photodiode array manufacturer. In one case, the image persistence has been found to be pixel-dependent, varying from 0.1 to 1% within a single array.[16] Although the image persistence effect can be difficult to characterize with high precision, it can be corrected with two terms according to $S_{j,c} = S_{j,u} + C_{IP}(S_{j,u}) - C_{IP}(S_{(j-1),u})$. Here $S_{j,c}$ and $S_{j,u}$ represent the corrected (c) and uncorrected (u) count levels for the jth readout in the succession, and $C_{IP}(x)$ is the image persistence correction factor for a count level of x (typically a constant multiplied by x). $C_{IP}(S_{j,u})$ represent the unread counts that remain to affect $S_{(j+1),u}$, whereas $C_{IP}(S_{(j-1),u})$ represent the counts that remain from $S_{(j-1),u}$. Figure 7.7 shows an example of a pixel-dependent image persistence correction factor for the array detector of the rotating-polarizer multichannel ellipsometer of Figs. 7.5 and 7.6.

Read-time errors. Pixel read-time errors arise because the photon-registering capability of the photodiode array is disrupted during the pixel read time. The pixel read time t_x can vary from 5 to 35 μs depending on the detector control system. If one assumes that the pixel is insensitive to photons incident during its read time and if this effect is not taken into account in the data analysis, then errors in the measured count levels S_j on the order of $t_x/t_e \sim 10^{-3}$ can be generated, where t_e is the exposure time. First-order corrections have been derived to include this effect in the analysis of the incident irradiance waveform at each pixel.[34,46,47]

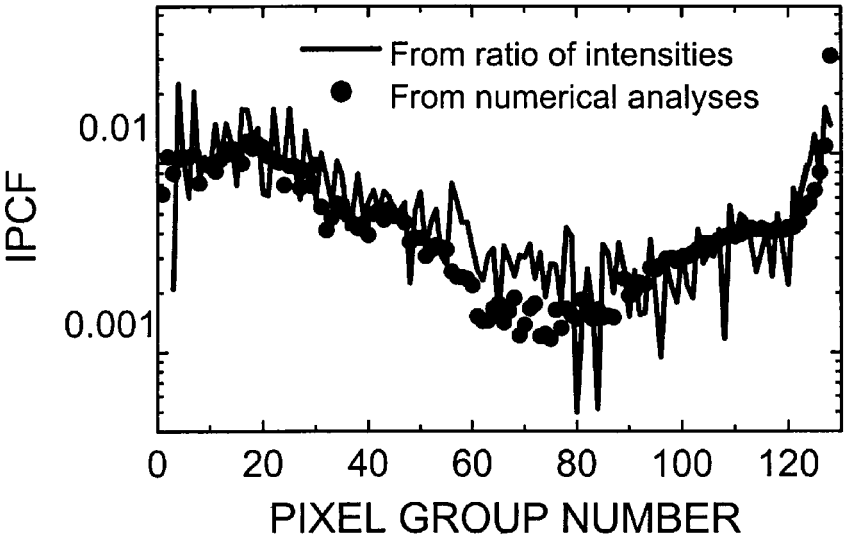


Figure 7.7 Image persistence correction factor (IPCF) for the rotating-polarizer multichannel ellipsometer of Figs. 7.5 and 7.6. These results were measured independently from (i) the ratio of the integrated irradiances between two successive readouts as a fast shutter is closed to block the beam incident on the detection system (solid line) and from (ii) the residual function under the assumption that its departure from zero in the straight-through configuration is solely due to image persistence (points). [Reproduced from Ref. 16.]

Stray light. Stray light is the term used to describe the light that reaches the photodiode array detector not having followed the designed optical path. Such light can be generated at three stages in the beam path: (i) via scattering from imperfections in the mirrors and grating within the spectrometer/detector enclosure, (ii) via multiple reflections between the surfaces of the detector window(s), and (iii) via the cross-talk among pixel groups. The latter two sources are identifiable by features in the stray light distribution, and are best minimized by system design, e.g., by limiting the number of interfaces that the incident beam crosses before being absorbed by the Si element of the array. If possible, the detector window should be removed and the spectrograph housing should be purged. In addition, coatings on the surface of the array should be avoided since they can lead to waveguiding of the light and cross-talk. The scattered stray light source (i) can be corrected at the raw data level as it is nearly uniform over the photodiode array. To characterize this source, a long (wavelength) pass optical filter can be mounted before the entrance slit of the spectrograph to ensure that no light associated with a range of pixels at short

wavelengths enters the spectrograph. Thus, any counts detected over these blocked pixels arise from stray light. To determine an approximate scattered stray light correction, an empirical correlation is obtained between the average stray light counts per blocked pixel and the sum of the counts over all exposed pixels under a variety of measurement conditions. Such a correlation has been found to be linear for two different detection systems and can be used to correct the photodiode array readouts obtained during the real time measurement.[46,47]

Higher order grating reflections. When a grating spectrograph is used as the dispersing component of a multichannel detection system, one must be aware of the possibility that higher order diffractions from the grating will corrupt the data. If the wavelength range of spectroscopic measurement is narrow, as in the phase-modulation instrument of Fig. 7.4, then a filter can be placed externally in order to block the higher order wavelengths from entering the input slits of the spectrograph. For real time instruments operating in parallel over a wide spectral range, one can use two different approaches. In the first approach, a prism is employed instead of a grating for dispersion. In the second approach, order-sorting filters are mounted internal to the grating spectrograph over the surface of the photodiode array. The latter approach is preferable since it permits use of a number of high-quality, commercially available spectrographs with flat focal planes designed for linear photodiode array detectors. To minimize stray light problems and distortions caused by filter mounting over the array surface, very thin (~0.1 mm) plastic color filters are used.[16] Figure 7.6 shows an example of the positioning of two order-sorting filters so as to obtain a spectral range from 190 to 825 nm with good rejection of diffractions from second to fourth order.

Finally, it should be noted that some commercial photodiode array detection systems have provided the capability of pixel grouping in the readout hardware. This allows one to sacrifice spectral resolution in favor of enhanced signal averaging and a shorter detector readout time. A shorter detector readout time in turn allows a higher rotation frequency for the rotating-polarizer multichannel ellipsometer, and thus shorter minimum acquisition and repetition times. In addition, nonuniform pixel grouping is also possible that can provide a more uniform resolution in photon energy when a grating spectrograph is used. For example, in a typical scheme with a prism spectrograph, uniform 8x grouping is used which effectively reduces a 1024 pixel photodiode array to a 128 pixel array, leading to a reduction in readout time from 16 ms to 4.5 ms.[24] Owing to the nonlinear prism dispersion with wavelength, a relatively uniform resolution in photon energy is retained by uniform grouping. In a typical scheme with a grating spectrograph, nonuniform grouping is used which effectively reduces a 1024 pixel photodiode array to a 138 pixel array with a read time of 5 ms.[16]

At the shortest wavelengths (~200 nm) 4x grouping is used, and with increasing wavelength this is stepped up, first to 8x and then to 16x, for a more uniform resolution in photon energy for studies of wide bandgap semiconductor film growth. When grouping in the readout hardware is used, care must be taken to perform the detector assessments (e.g., nonlinearity and image persistence) using the same grouping scheme as in the ellipsometric measurements. The grouping scheme adopted for ellipsometric measurements also impacts how certain corrections must be performed (e.g., read time errors and stray light). Other commercial systems provide a 5 ms readout time for the full 1024 pixels of the photodiode array without the capability of hardware grouping. In this case, all detector assessments and corrections are performed in the ungrouped mode. Pixel grouping can then be applied to the corrected data through manipulations in the software in order to achieve additional signal averaging.

7.3 Rotating-Element Designs

Figure 7.8 shows three different multichannel ellipsometer designs based on rotating-element principles, that have been successfully developed and applied in a wide variety of problems in thin film and surface characterization. These designs include (a) the rotating-polarizer multichannel ellipsometer in the $P_r(\omega_P)SA$ optical configuration used for real time (three-parameter) spectroscopic ellipsometry, (b) the single rotating-compensator multichannel ellipsometer in the $PSC_r(\omega_C)A$ configuration used for real time unnormalized (four-parameter) Stokes vector spectroscopy, and (c) the dual rotating-compensator multichannel ellipsometer in the $PC_{1r}(5\omega_B)SC_{2r}(3\omega_B)A$ configuration used for high-speed unnormalized (16-parameter) Mueller matrix spectroscopy. In this notation P, C, S, and A, designate the polarizer, compensator, sample, and analyzer, respectively. The subscript 'r' indicates the rotating element(s) with the rotational angular frequency denoted in parentheses. In the following sections, a theoretical description is presented, along with the methods of data collection, data reduction, and calibration for each of the three instrument configurations.

7.3.1 Rotating Polarizer

7.3.1.1 Theoretical Description

The following equations describe the time-dependent irradiance waveform $I(t)$ predicted at each pixel (or pixel group) of the photodiode array detector in terms of the sample parameters $\{(\psi, \Delta), R(A')\}$, i.e., the

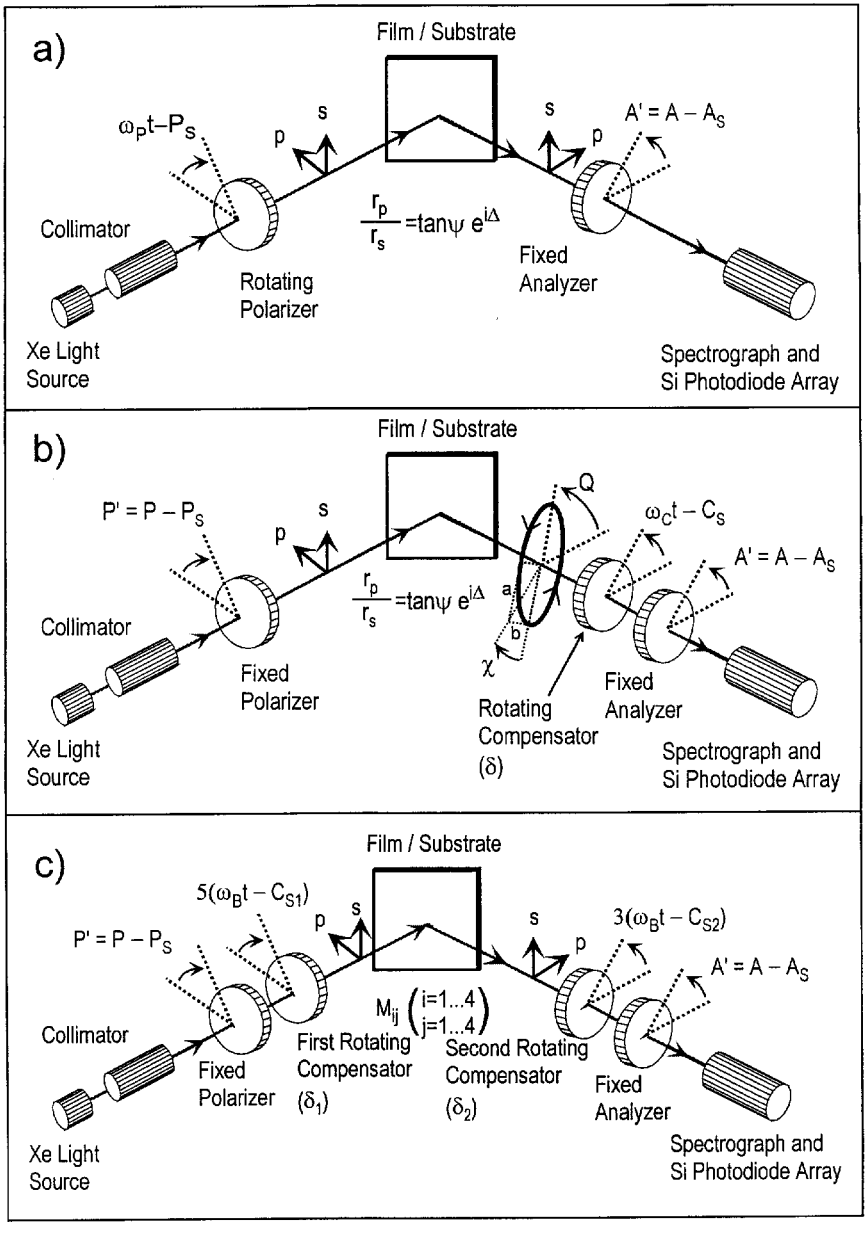


Figure 7.8 Three different multichannel ellipsometer designs based on rotating-element principles, including (a) the rotating-polarizer multichannel ellipsometer in the $P_r(\omega_P)SA$ optical configuration, (b) the single rotating-compensator multichannel ellipsometer in the $PSC_r(\omega_C)A$ configuration, and (c) the dual rotating-compensator multichannel ellipsometer in the $PC_{1r}(5\omega_B)SC_{2r}(3\omega_B)A$ configuration.

ellipsometric angles and the polarized reflectance, for an ideal isotropic sample [14,34]

$$I(t) = I_0\{1 + \alpha\cos 2(\omega_P t - P_S) + \beta\sin 2(\omega_P t - P_S)\}, \tag{7.2a}$$

$$I_0 = I_{00}R(A') = I_{00}(|r_p|^2\cos^2 A' + |r_s|^2\sin^2 A'), \tag{7.2b}$$

$$B = \alpha + i\beta = \frac{[\tan^2\psi - \tan^2 A'] + i[2\tan\psi\tan A'\cos\Delta]}{[\tan^2\psi + \tan^2 A']}. \tag{7.2c}$$

The ellipsometric angles (ψ, Δ) in Eq. (7.2c) are related to r_p and r_s, the complex amplitude reflection coefficients of the sample for light polarized linearly along the p (parallel to the plane of incidence) and s (perpendicular to the plane of incidence) directions, in accordance with $\tan\psi\exp(i\Delta) = r_p/r_s$. The reflectance of the sample $R(A')$ is related to r_p and r_s, and the analyzer angle A', through the second equality in Eq. (7.2b). In Eqs. (7.2), $\{I_0, (\alpha, \beta)\}$ are the theoretically predicted average irradiance and the dc-normalized $2\omega_P$ cosine and sine Fourier coefficients of the irradiance waveform, respectively. The theoretically predicted Fourier coefficients in these equations are explicitly corrected for the phase angle of the polarizer through P_S. From Eq. (7.2a), it is clear that $-P_S$ is the angle of the polarizer transmission axis at time zero of the data collection process for the given pixel. This angle is measured with respect to the plane of incidence in a counter-clockwise positive sense, looking opposite to the beam direction. In Eqs. (7.2b) and (7.2c), $A' = A - A_S$ is the true angle of the analyzer transmission axis with respect to the plane of incidence, where A is the analyzer scale reading and A_S is the correction to the scale reading. In addition, in Eq. (7.2b), I_{00} describes the spectral response function of the ellipsometer. Finally, in Eq. (7.2c), the complex quantity B is used to describe both Fourier coefficients in terms of the sample properties (ψ, Δ) and the analyzer angle A'. To determine the individual Fourier coefficients, the real and imaginary parts of this equation are equated. The reason for using this shorthand notation will become clear as the other rotating-element configurations are discussed.

7.3.1.2 Data Collection

For an error-free system in the rotating polarizer configuration, the irradiance waveform at any given pixel (or pixel group) obeys the following experimental expression:[14,34]

$$I'(t) = I'_0(1 + \alpha'\cos 2\omega_P t + \beta'\sin 2\omega_P t), \tag{7.3a}$$

where ω_P is the angular frequency of the rotating polarizer and α' and β' are dc-normalized Fourier coefficients to be determined experimentally.

This experimental expression differs from the theoretical one of Eq. (7.2a) in that the former waveform does not include the polarizer phase angle P_S. This term is omitted from the phase of the experimental waveform because it cannot be determined until after the calibration is performed.

Conventional Fourier analysis relies on sampling of the detector output at regular intervals of the polarizer angle. This approach cannot be employed when the photodiode array detector is operating in an integrating mode for fully parallel data acquisition (unless $t_e \ll \pi/\omega_P$). Thus, if the array operating in an integrating mode is read out N times per half rotation of the rotating polarizer, leading to N equal exposure times $t_e = \pi/N\omega_P$, then N spectra are generated having the following form:

$$S_j = I'_0 \int_{(j-1)\pi/N\omega_P}^{j\pi/N\omega_P} (1 + \alpha' \cos 2\omega_P t + \beta' \sin 2\omega_P t)dt\,; \quad (j = 1, \ldots, N), \tag{7.3b}$$

$$= \frac{\pi I'_0}{N\omega_P} + \frac{I'_0}{\omega_P}\left(\sin\frac{\pi}{N}\right)\left\{\alpha' \cos\frac{(2j-1)\pi}{N} + \beta' \sin\frac{(2j-1)\pi}{N}\right\}. \tag{7.3c}$$

For each pixel of the photodiode array, Eq. (7.3c) represents a system of N equations in three unknowns $\{I'_0, (\alpha', \beta')\}$. For $N = 3$, these equations can be readily solved to yield the spectra in $\{I'_0, (\alpha', \beta')\}$. For $N > 3$, $\{I'_0, (\alpha', \beta')\}$ are overdetermined, and an effective method of solution must be established to utilize the full data set $\{S_j, j = 1, \ldots, N\}$ in the determination of $\{I'_0, (\alpha', \beta')\}$ as described in the next section.

7.3.1.3 Data Reduction

In this part, the approach will be described to convert the raw spectra in $\{S_j, j = 1, \ldots, N\}$, i.e., the waveform integrations, to spectra in the ellipsometric angles and reflectance $\{(\psi, \Delta), R(A')\}$. For the rotating polarizer multichannel ellipsometer, a common choice is $N = 4$, which allows the experimental dc and normalized $2\omega_P$ Fourier coefficients to be determined according to the following expressions:[36]

$$\alpha' = \frac{\pi}{2}\frac{(S_1 - S_2 - S_3 + S_4)}{(S_1 + S_2 + S_3 + S_4)}, \tag{7.4a}$$

$$\beta' = \frac{\pi}{2}\frac{(S_1 + S_2 - S_3 - S_4)}{(S_1 + S_2 + S_3 + S_4)}, \tag{7.4b}$$

$$I'_0 = \frac{1}{\pi}(S_1 + S_2 + S_3 + S_4). \tag{7.4c}$$

Measurement of the four waveform integrals provides a fourth equation that can be written as:

$$\beta'_4 = \frac{\pi}{2}\frac{(S_1 - S_2 + S_3 - S_4)}{(S_1 + S_2 + S_3 + S_4)}, \tag{7.4d}$$

where β'_4 is the experimentally determined sine $4\omega_P$ Fourier coefficient. In theory this coefficient should vanish [see Eq. (7.2a)]; however, in the presence of specific experimental errors, e.g., source polarization or detector nonlinearity, small non-zero values are obtained.[34] Thus, Eq. (7.4d) serves as a consistency check that allows one to assess and even to minimize the errors in some cases. For example, the spectrum in β'_4 can be monitored in real time as the Xe lamp position is adjusted. This can result in an alignment that minimizes errors due to source polarization.

As an example, Fig. 7.9 (upper panel) shows the waveform sum with alternating signs from the numerator of Eq. (7.4d) as well as the total sum from the denominator. Figure 7.9 (lower panel) shows the ratio of these two

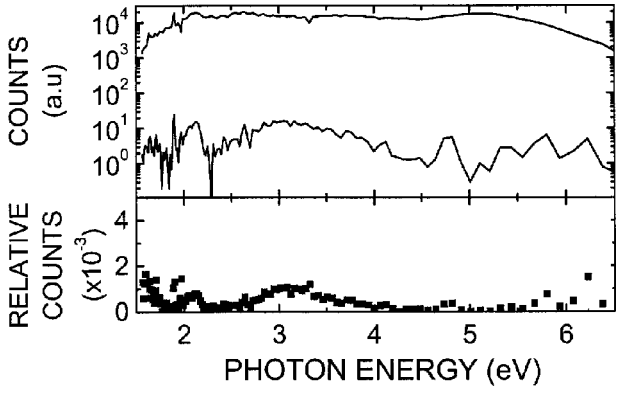


Figure 7.9 A consistency check for the rotating-polarizer multichannel ellipsometer of Figs. 7.5–7.7 (upper panel), consisting of alternating differences and sums of the four waveform integrals, $S_1 - S_2 + S_3 - S_4$, (lower data) along with the total sum, $S_1 + S_2 + S_3 + S_4$ (upper data). Also shown is the ratio of the two results from the upper panel, yielding the quantity $2\beta'_4/\pi$ in Eq. (7.4d) (lower panel). The quantity $2\beta'_4/\pi$ should vanish in the absence of errors such as detector nonlinearity and source polarization. [Unpublished data, J. A. Zapien, Pennsylvania State Univ. (2000).]

sums, which is given by $2\beta'_4/\pi$ from Eq. (7.4d). These results show that β'_4 is less than 0.002 over the full range of photon energy from 1.5 to 6.5 eV. In the highest sensitivity range of the instrument, the values are more than an order of magnitude lower. An alternative approach to Eqs. (7.4) results from using $N = 5$ in Eqs. (7.3b) and (7.3c). From five integrals, it is possible to obtain both the cosine and sine $4\omega_P$ coefficients, α'_4 and β'_4. In this case the amplitude $|B'_4|^2 \equiv \alpha'^2_4 + \beta'^2_4 = 0$ provides an improved consistency check over Eq. (7.4d). When $|B'_4|^2$ vanishes, one is assured that both $4\omega_P$ coefficients vanish. Another consistency check of interest is to ensure the equality of the (α', β') spectra measured when two successive optical cycles vanish. Although these two pairs of spectra are optically equivalent, they are mechanically inequivalent. As a result, this approach can be used to check for optical component misalignments and imperfections.

Once the spectra in the experimental waveform parameters $\{I'_0, (\alpha', \beta')\}$ have been determined, they need to be converted to the theoretical parameters $\{I_0, (\alpha, \beta)\}$ so that the spectra in $\{(\psi, \Delta), R(A')\}$ can be computed. Equating Eqs. (7.2a) and (7.3a) yields the following transformation:

$$I_0 = I'_0, \tag{7.5a}$$

$$\alpha = \alpha' \cos 2P_S + \beta' \sin 2P_S, \tag{7.5b}$$

$$\beta = -\alpha' \sin 2P_S + \beta' \cos 2P_S. \tag{7.5c}$$

Alternatively, Eqs. (7.5b) and (7.5c) can be written as

$$(\alpha \quad \beta)^{\mathrm{T}} = \Re(2P_S)\,(\alpha' \quad \beta')^{\mathrm{T}}. \tag{7.5d}$$

In this latter compact expression, $\Re(2P_S)$ represents the 2×2 rotation transformation matrix with elements $r_{11} = r_{22} = \cos 2P_S$ and $r_{12} = -r_{21} = \sin 2P_S$, and the superscript 'T' represents the transpose operation that converts the two row vectors to column vectors.

Now that the transformed coefficients are available, the theoretical expressions given in Eqs. (7.2b–c) can be inverted to obtain information on the sample:

$$\tan\psi = \sqrt{\frac{1+\alpha}{1-\alpha}}\ \left|\tan(A - A_S)\right|, \tag{7.6a}$$

$$\cos\Delta = \frac{\beta}{\left|\sqrt{1-\alpha^2}\right|}, \tag{7.6b}$$

$$R(A') = \frac{I_0}{I_{00}}. \tag{7.6c}$$

Here it is assumed that $0 < A' < 90°$ so that the sign of $\cos\Delta$ is carried by β. Thus, conversion of raw spectra in $\{S_j, j = 1, \ldots, 4\}$ to $\{(\psi, \Delta), R(A')\}$ is ostensibly straightforward; however, the quantities P_S in Eqs. (7.5b) and (7.5c), A_S in Eq. (7.6a), and I_{00} in Eq. (7.6c) all have yet to be determined. Furthermore $A' = A - A_S$ must also be known in order to model the reflectance spectra deduced from Eq. (7.6c). In the next section, the issues of calibration that provide $\{P_S, A_S, I_{00}\}$ will be treated.

It should be emphasized here that the sign of Δ is not accessible in the rotating-polarizer or analyzer ellipsometer configurations. Prior knowledge of the sample properties or a detailed analysis of the ellipsometric spectra of the sample can usually provide this information. For example, if the sample is known to present a single reflecting interface to the ambient, then $0° \leq \Delta \leq 180°$ within the so-called "optics" ($e^{i\omega t}$) convention of time-varying electric fields adopted here and in Ref. 8. [In contrast, the so-called "physics" ($e^{-i\omega t}$) convention is adopted in Ref. 4.]

Figure 7.10 provides an example of ellipsometric spectra in (ψ, Δ) spanning the range from 1.5 to 6.5 eV collected with the rotating-polarizer

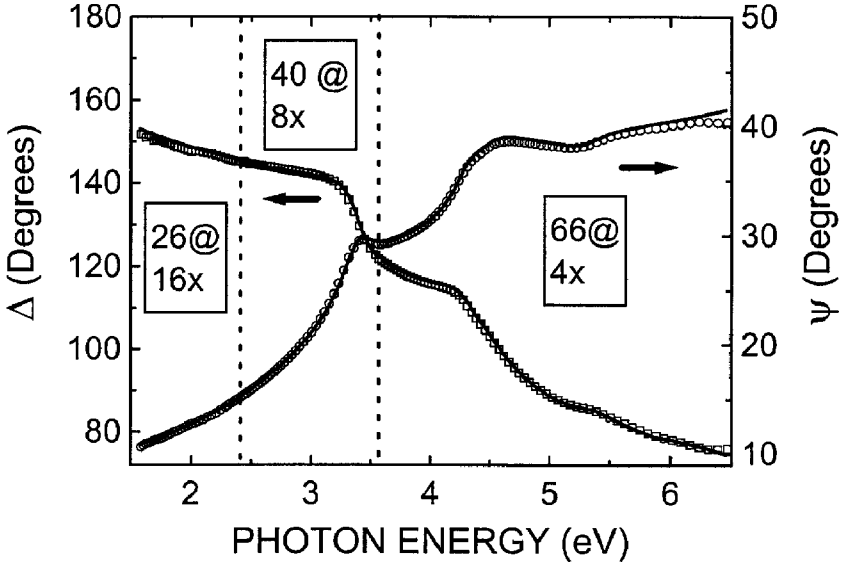


Figure 7.10 Ellipsometric spectra in (ψ, Δ) collected with the rotating-polarizer multichannel ellipsometer of Fig. 7.5 for a crystalline silicon wafer with a ~100 Å thermally grown oxide layer on its surface (points). Nonuniform pixel grouping was performed in accordance with the description in the boxes, yielding 132 spectral points. The total acquisition time was 1.96 s. The solid line represents the best fit two-parameter simulation, yielding the oxide thickness (120.7 ± 0.6 Å) and the angle of incidence (69.94 ± 0.09°). [Adapted from Ref. 16.]

multichannel ellipsometer of Fig. 7.5. For this example, the specularly reflecting sample consists of a crystalline silicon wafer with a ~100 Å thermally grown oxide layer on its surface. [In this case, $\Delta > 0°$ within the optics ($e^{i\omega t}$) convention.] The data were obtained as an average of waveform integrals over 80 optical cycles of 24.5 ms each; thus, the total acquisition time was 1.96 s. Nonuniform pixel grouping was performed in accordance with the description in the boxes in Fig. 7.10, yielding 132 spectral points and improved photon energy resolution for energies greater than 3.5 eV.

Figure 7.11 provides another illustrative example of ellipsometric spectra collected with a rotating-polarizer multichannel instrument. In this example, the instrument of Fig. 7.3 was used, which provided a spectral range from 1.4 to 4.0 eV. In contrast to the spectra of Fig. 7.10, the results of Fig. 7.11 were obtained in real time, namely during the growth of a 320 Å silver thin film by magnetron sputtering onto a native oxide-covered crystalline silicon wafer held at room temperature. Each pair of spectra in (ψ, Δ) has been converted to spectra in the real and imaginary parts of the pseudo-dielectric function ($<\varepsilon_1>$, $<\varepsilon_2>$) in Fig. 7.11, using the Fresnel equations for a hypothetical single interface. In addition, each reflectance spectrum has been normalized by that of the oxide-covered silicon substrate. Consequently, the relative reflectance spectra are all close to unity throughout the first 8 min when the target was presputtered; a shutter prevented the resulting atomic flux from reaching the substrate. The three three-dimensional surfaces in Fig. 7.11 are constructed from 240 triplets of spectra collected versus time, each having 85 spectral points from 1.4 to 4.0 eV; thus the entire plot contains 6.12×10^4 data points. Spectra in the four waveform integrals S_j ($j = 1, \ldots, 4$), which ultimately provided the three spectra of Fig. 7.11 at a given time, were collected in 0.8 s, corresponding to an average over 10 mechanical rotations of the rotating polarizer ($\omega_P/2\pi = 12.5$ Hz). The silver film was deposited at an average rate of 0.15 Å/s; thus, ~0.1 Å accumulated during the acquisition time for the four averaged waveform spectra. The ability to measure all three spectra in Fig. 7.11 allows one to solve the classic problem in ellipsometry, in this case determining the true dielectric function and thickness as continuous functions of time in the initial stages during silver film nucleation when the discontinuous film can be approximated as a single layer.[14]

7.3.1.4 Calibration

Two different methods have been described for the determination of the offset and phase angles, i.e., P_S and A_S, for rotating-polarizer and analyzer ellipsometers. It is important to design such calibration methods

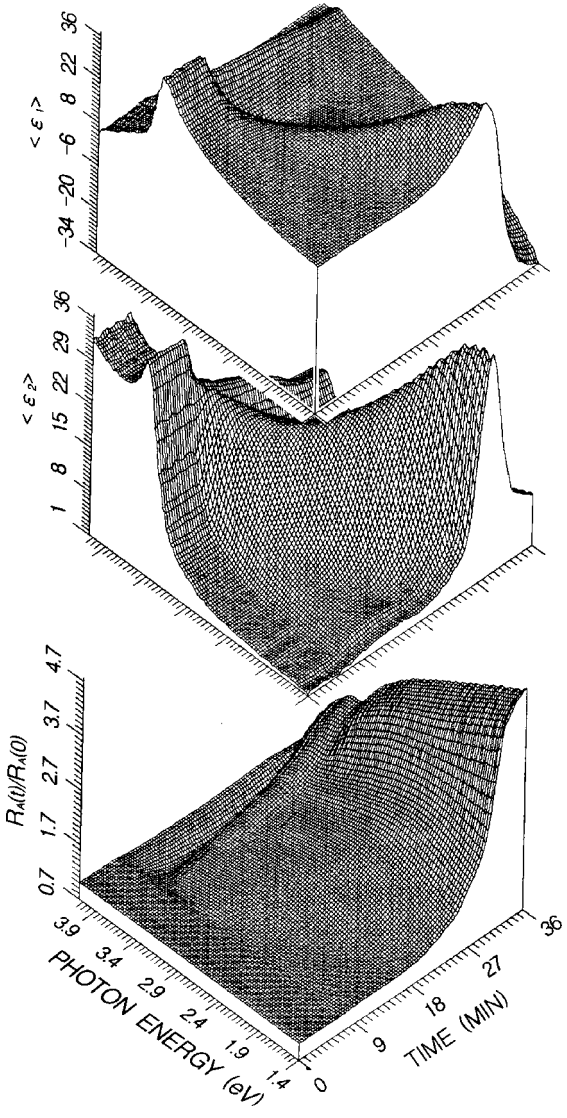


Figure 7.11 Processed spectra, including the pseudo-dielectric function and the sample reflectance relative to the substrate reflectance, from three-parameter real time spectroscopic ellipsometry performed using the instrument of Fig. 7.3. These results were deduced from spectra in the waveform integrals collected during the growth of a 320 Å silver film by magnetron sputtering onto a native oxide-covered silicon wafer held at room temperature. To construct the three-dimensional plots, 85 points were collected versus photon energy and 240 triplets were collected versus time. [Reproduced from Ref. 15.]

so that they can be performed just prior to sample measurement, with the sample in place, using the same configuration and alignment as is used in the measurement. The first method has been described by Aspnes[50] and involves determining specific functions of the measured Fourier coefficients at a given wavelength, e.g., the residual and phase functions, versus the fixed element angular reading in the neighborhood of the p-direction [i.e., $A \approx A_S$, for the rotating-polarizer instrument] and/or in the neighborhood of the s-direction [i.e., $A \approx A_S + (\pi/2)$]. These functions are then fit with polynomial expressions to obtain P_S and A_S. No other information on the sample is required in the fitting. In an alternative method described by Johs,[51] the measured Fourier coefficients (α', β') at a given wavelength are obtained as essentially continuous functions of the fixed element angular reading from the *p* to the *s*-direction. These data are complicated functions of the parameters (ψ, Δ), P_S, and A_S, as can be seen by the substitution of Eq. (7.2c) into the inverted forms of Eqs. (7.5b) and (7.5c). These functions can be fitted to deduce all four parameters. Thus, in this case, the ellipsometric angles are determined along with the calibration angles. In both calibration methods, specific error parameters have also been extracted. For example, in the first demonstration of the method described by Aspnes, the optical activity parameters of the quartz Rochon polarizer and analyzer were determined for a single-channel rotating-analyzer system.[50] In the first demonstration of the method by Johs, parameters that describe the polarization sensitivity of the detection system were determined.[51] In both demonstrations, the ac/dc gain ratio of the detection system electronics was also determined.

In this part, the focus will be on the Aspnes method of calibration, and different approaches based on this general framework will be described for the specific case of the rotating-polarizer multichannel ellipsometer. These approaches include the residual/phase function calibration,[50] the zone-difference-phase/phase function calibration,[52] and the irradiance/phase function calibration.[53] In addition, the extension of the original single-wavelength calibration mode to a fully parallel spectroscopic calibration mode will be emphasized. Finally, this section will conclude with a discussion of the calibration procedure described by An et al.[14] to deduce the spectral response function I_{00}.

Offset and Phase Angle Calibrations

Residual-Function/Phase-Function Calibration

The residual function calibration method for rotating-analyzer ellipsometers relies on the fact that a light wave reflected from the surface of an isotropic sample will be linearly polarized if the incident wave is linearly

polarized along the orthogonal *p*- or *s*-directions.[50] For an absorbing sample, the polarization state of the reflected light beam gradually increases in ellipticity as the linear polarization state of the incident beam is rotated away from these directions. The same principle also holds for the rotating-polarizer instrument configuration which is of primary interest here. In the absence of errors then, the experimental amplitude function $|B'|^2 \equiv \alpha'^2 + \beta'^2$ for the rotating-polarizer instrument should reach its maximum value of unity for $A' = 0$ or $\pi/2$ radians, i.e., along the orthogonal *p*- and *s*-directions. The residual function F' is defined as the deviation of the amplitude function from its maximum, which is minimized in the absence of errors when $A' = 0$ or $\pi/2$, or alternatively, when $A = A_S$ or $A_S + \pi/2$.

The following expressions for the residual function form the basis for the determination of A_S:

$$F'(A) = 1 - |B'|^2 \tag{7.7a}$$

$$= 1 - |B|^2 \tag{7.7b}$$

$$\cong 4(A - A_S)^2 \cot^2\psi \sin^2\Delta; \quad A \approx A_S, \tag{7.7c}$$

$$\cong 4\{A - [A_S + (\pi/2)]\}^2 \tan^2\psi \sin^2\Delta; \quad A \approx A_S + (\pi/2). \tag{7.7d}$$

Equation (7.7a) provides an expression into which the experimental spectra (α', β') collected at different analyzer angles are substituted. In Eq. (7.7b), $|B|^2 = \alpha^2 + \beta^2$; thus the residual function F' is invariant in the transformation of Eqs. (7.5b) and (7.5c) from the experimental to the theoretical Fourier coefficients. Equations (7.7c) and (7.7d) can be derived from Eq. (7.7b) by applying Eq. (7.2c); these two derivations are valid to second order in the small quantities $(A-A_S)$ and $\{A-[A_S + (\pi/2)]\}$, respectively. Thus, by fitting the experimental results for $F'(A)$ at each pixel (or pixel group) to a second-order polynomial, e.g., for $A \approx A_S$, $F'(A) = c_0 + c_1A + c_2A^2$, then A_S or $A_S + (\pi/2)$ is determined as the value of A at which the minimum in the best fit polynomial occurs, e.g., for $A \approx A_S$, $A_S = -c_1/2c_2$. As a result, A_S can be determined either from Eq. (7.7c) or from Eq. (7.7d) as a function of pixel number or wavelength.

A similar development is needed to determine the phase angle, P_S. In this case, the phase function is defined by:

$$\Theta'(A) = \frac{\arctan(\beta'/\alpha')}{2} \tag{7.8a}$$

$$= P_S + \frac{\arctan(\beta/\alpha)}{2} \tag{7.8b}$$

$$\Theta(A') \cong P_S + (A - A_S)\cot\psi\cos\Delta; \quad A \approx A_S, \tag{7.8c}$$

$$\cong P_S + \{A - [A_S + (\pi/2)]\}\tan\psi\cos\Delta; \quad A \approx A_S + (\pi/2). \tag{7.8d}$$

Again, Eq. (7.8a) provides an expression into which the experimental spectra (α', β') collected at different analyzer angles are substituted. In addition, Eqs. (7.8c) and (7.8d) can be derived from Eq. (7.8b) by applying Eq. (7.2c); these two derivations are valid to second order in the small quantities $(A - A_S)$ and $\{A - [A_S + (\pi/2)]\}$, respectively. Thus, calibration proceeds by fitting the experimental results at each pixel to a linear function, e.g., for $A \approx A_S$, $\Theta'(A) = c_0 + c_1 A$; then P_S is determined as the value of $\Theta'(A_S)$ or $\Theta'[A_S + (\pi/2)]$, e.g., for $A \approx A_S$, $P_S = c_0 + c_1 A_S$. As a result, P_S can be determined from Eq. (7.8c) or from Eq. (7.8d) as a function of pixel number or wavelength.

Lastly, it should be noted that if the inverse tangent in Eq. (7.8a) is confined by computation to the range $-\pi/2 < 2\Theta'(A) \leq \pi/2$, then P_S similarly must lie within the range $-\pi/4 < P_S \leq \pi/4$ for the equations to be valid. If P_S lies outside this range, then quadrant corrections of $\pm\pi/2$ radians will be needed on the right hand sides of Eqs. (7.8c) and (7.8d). In order to identify the required quadrant correction, it is sometimes helpful to pre-align the transmission axis of the polarizer with respect to the motor shaft in the straight-through configuration. In this procedure, P' is set to a value near zero at the onset of data collection ($t = 0$) for a selected reference pixel (or pixel group), thus allowing the resolution of any possible quadrant ambiguities. Such a procedure becomes even more important for the rotating-compensator instruments of Sections 7.3.2 and 7.3.3.

Limitations on the residual-function/phase-function calibration arise because there are certain ranges in (ψ, Δ) over which the calibration fails. Failure of the method occurs when Δ is near 0° or ±180° because the two second-order polynomials of Eqs. (7.7c) and (7.7d) vanish, and the minima that define A_S disappear. Since A_S cannot be determined under these conditions, P_S cannot be determined either. Thus, A_S is determined most accurately via Eqs. (7.7c) and (7.7d) when the curvature of the parabola is maximized; P_S is most accurately determined via Eqs. (7.8c) and (7.8d) when the slope of the linear function is minimized. These optimum conditions occur for Eqs. (7.7c) and (7.8d) when $|\Delta| \sim 90°$ and $\psi \sim 0°$, and for Eqs. (7.7d) and (7.8c) when $|\Delta| \sim 90°$ and $\psi \sim 90°$. Thus, if Δ is sufficiently far away from 0° or ±180° for the residual-function/phase-function approach to be effective, i.e., if $30° < |\Delta| < 150°$, then Eqs. (7.7c) and (7.8d) are preferable when $\psi < 45°$ and Eqs. (7.7d) and (7.8c) are preferable when $\psi > 45°$. In practice, however, if $|\Delta| < 30°$ or $|\Delta| > 150°$, an alternative approach is needed for accurate offset and phase angle calibration. For many samples, the broad spectral range of

multichannel ellipsometer calibration ensures that, over some range(s), the A_S and P_S values from the residual-function/phase-function calibration will be accurate. For samples such as transparent glasses, however, the alternate approach is unavoidable.

Zone-Difference-Phase-Function/Phase-Function Calibration

The zone-difference phase function provides an alternative method to obtain A_S:[52]

$$\Phi'(A) = \Theta'(A) - \Theta'[A + (\pi/2)] \tag{7.9a}$$

$$= \frac{\arctan(\beta'/\alpha')|_A - \arctan(\beta'/\alpha')|_{A+(\pi/2)}}{2}, \tag{7.9b}$$

$$\cong 2(A - A_S)\cot 2\psi \cos\Delta; \qquad A \approx A_S. \tag{7.9c}$$

Equation (7.9b) provides an expression into which the experimental spectra (α', β') collected at different pairs of analyzer angles $[A, A + (\pi/2)]$ can be substituted. The derivation leading to Eq. (7.9c) is valid to second order in the small quantity $(A - A_S)$. A quadrant correction may be required in Eq. (7.9b) in order to apply Eq. (7.9c). Thus, calibration proceeds by fitting the experimental results at each pixel to a linear function $\Phi'(A) = c_0 + c_1 A$; then A_S is determined as the value of A for which $\Phi' = 0$, i.e., $A_S = -c_0/c_1$. As a result, A_S can be determined from Eq. (7.9c) as a function of pixel number or wavelength. Once A_S is determined by this method, P_S can be determined from the phase function as described by Eq. (7.8c) or (7.8d).

The offset angle A_S can be determined most accurately from Eq. (7.9c) when the slope of the linear relationship is maximized. This occurs when $|\Delta| = 0°$ or 180° and $\psi = 0°$ or 90°. This calibration procedure fails when the slope of the relationship in Eq. (7.9c) vanishes, i.e., when $|\Delta| = 90°$ or $\psi = 45°$. Thus, the zone-difference phase function method of calibration is complementary to the residual function method. As a general rule, the residual-function/phase-function calibration is suitable for metals and for semiconductors excited at photon energies above the lowest direct bandgap, whereas the zone-difference-phase-function/phase-function calibration is suitable for dielectrics and for semiconductors excited at photon energies below the lowest direct bandgap.

Irradiance-Function/Phase-Function Calibration

The previously described calibration approaches in rotating-polarizer multichannel ellipsometry involve acquiring spectra in the dc-normalized $2\omega_P$ Fourier coefficients (α', β') of the irradiance waveform at the detector

as a function of the analyzer angular reading A. In these approaches, the angular dependence of the dc Fourier coefficient, i.e., the average irradiance in the reflected beam, is used only for normalization of the $2\omega_P$ Fourier coefficients. Thus, potentially useful information in I'_0 is discarded. Here it is shown how this information can be employed in instrument calibration. The basic equations from which A_S can be determined from I'_0 are as follows:[53]

$$I'_0(A) = I_{00}\left\{|r_p|^2 + \left(|r_s|^2 - |r_p|^2\right)(A - A_S)^2\right\}; \quad A \approx A_S, \tag{7.10a}$$

$$\begin{aligned} &= I_{00}\left\{|r_s|^2 - \left(|r_s|^2 - |r_p|^2\right)\{A - [A_S + \pi/2]\}^2\right\}; \\ &\quad A \approx A_S + (\pi/2). \end{aligned} \tag{7.10b}$$

These expressions are valid to second order in the small quantities $(A - A_S)$ and $\{A - [A_S + (\pi/2)]\}$. Thus, by fitting the experimental results for I'_0 at each pixel to a second-order polynomial, then A_S or $A_S + (\pi/2)$ is determined as the value of A at which the extremum in the polynomial occurs. As a result, A_S can be determined either from Eq. (7.10a) or from Eq. (7.10b) as a function of pixel number or wavelength. Both Eqs. (7.10a) and (7.10b) provide the same sensitivity for the determination of A_S. The highest sensitivity occurs when the difference between the reflectances for linearly polarized light in the p and s directions is maximized; the procedure fails when $|r_p|^2 = |r_s|^2$ or when $\psi \sim 45°$. The irradiance function calibration has an advantage over the more widely used residual function and the zone-difference phase function approaches because there are no constraints on the value of Δ for accurate calibration. Once A_S is determined according to Eq. (7.10a) or (7.10b), P_S can be deduced as usual from the phase functions in Eq. (7.8c) or (7.8d).

To demonstrate the irradiance-function/phase-function calibration procedure, test calibrations will be presented here for a Si wafer substrate covered with its native oxide, undertaken using a rotating-polarizer multichannel ellipsometer. These calibrations were run just prior to real time measurements performed during thin film deposition on the Si substrate. Figure 7.12 (top) shows the measured irradiance I'_0 for the 85th pixel group (out of 128 groups in an 8x grouping scheme for a 1024 pixel photodiode array), plotted as a function of the analyzer scale reading A. This pixel group corresponds to a photon energy of 1.98 eV. A high signal-to-noise ratio is obtained, and a fit to a parabolic function as described above yields a well-defined minimum at $A = A_S = 0.84°$. In the lower panel of Fig. 7.12, $\Theta'(A)$ is plotted for the same three-parameter data set. Fitting this curve to a linear relationship near $A = A_S$ and evaluating the resulting fit at $A = A_S$, yields $P_S = 87.56°$.

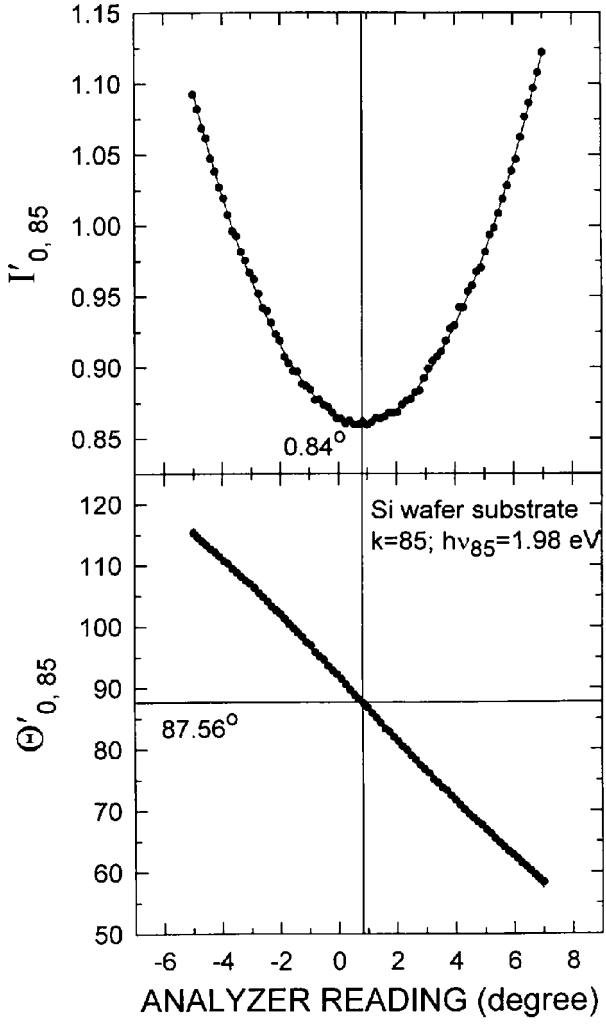


Figure 7.12 The dc Fourier coefficient I'_0 (top) and phase function Θ' (bottom) plotted as a function of the analyzer angular reading A near $A = A_S$ for the calibration of a rotating-polarizer multichannel ellipsometer using a silicon wafer sample. Parabolic and linear fits to these two functions yield the analyzer scale correction A_S and the polarizer phase angle P_S for the pixel group of interest. These results were obtained at pixel group $k = 85$ (from a total of 128 pixel groups established by an 8x grouping scheme for this particular detector), corresponding to a photon energy of 1.98 eV. [Adapted from Ref. 53.]

Figure 7.13(a) shows A_S versus photon energy from ~90 pixel groups covering the spectral range from 1.5 to 3.0 eV (the range that yields the most reliable results), obtained using the irradiance function approach of Fig. 7.12 (top). A_S is found to be constant versus energy with a standard deviation of ±0.02° about an average value of 0.852°. Also plotted in Fig. 7.13(b) are the results of the zone-difference phase function approach to extract A_S. In this approach, the phase function Θ' obtained experimentally for the settings of A in the neighborhood of $A_S + (\pi/2)$ is subtracted from that obtained for the settings of A in the neighborhood of A_S. The zero crossing of the resulting function defines A_S. In this case also, A_S is nearly constant with photon energy from 1.5 to 3.0 eV, as shown in Fig. 7.13(b), with a standard deviation of ±0.03° about an average value of 0.849°. The constancy of A_S and the near coincidence of the values obtained from the irradiance and zone-difference phase functions show that the instrument is free of chromatic artifacts. Figure 7.13(c) shows P_S versus pixel group number over the same range as for the data of Fig. 7.13(a–b). P_S is linear with pixel group number k, having a value of 104.93° at the first group and a slope of $\Delta P_S/\Delta k = -0.206°$. (Thus, the polarizer was not pre-aligned relative to the motor shaft in this case.) The linear dependence in Fig. 7.13(c) arises from the fact that although data acquisition with the multichannel detector is fully parallel, readout of the data is performed serially. In fact for the detector of Figs. 7.12 and 7.13, the time required to read one pixel group is 35 μs. Thus, time zero signifying the start of the S_1 integration in Eq. (7.3b) shifts by 35 μs for each successive pixel group. In this time, the polarizer rotates by 0.206° (at the polarizer rotation frequency of 16.35 Hz for the instrument) and this effect accounts for the observed slope in P_S.

Spectral Response Calibration

The determination of the polarized reflectance $R(A')$ opens up numerous possibilities for the characterization of non-ideal samples, e.g., samples with macroscopic roughness. The critical step in the measurement of $R(A')$ is the determination of the ellipsometer spectral response function I_{00}. In theory, I_{00} can be obtained by removing the sample and setting the ellipsometer in straight-through. However, such a procedure is incompatible with real time data acquisition in the oblique reflection measurement configuration. In the reflection configuration, it is best to employ the substrate or the starting surface just prior to film growth or materials processing as a calibration standard. Alternatively, in cases for which the substrate or starting surface structure is too complex, a test sample such as an ultrathin crystalline Si wafer can be mounted on top of the sample of interest, but only partially eclipsing the sample surface. Then the

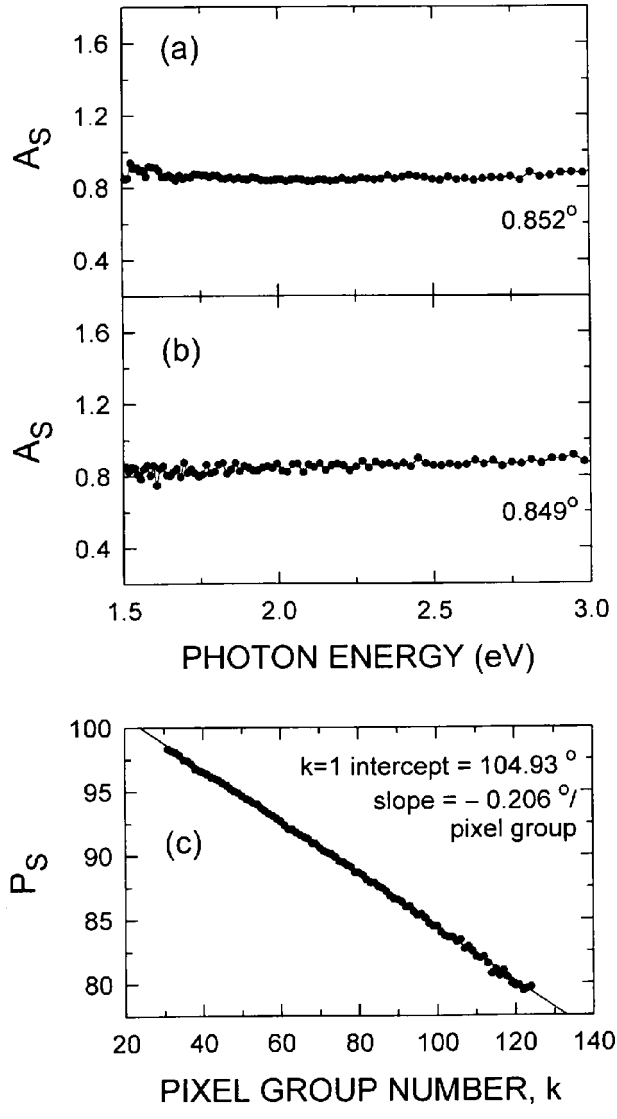


Figure 7.13 Calibration results for a rotating-polarizer multichannel ellipsometer, including the analyzer angular offset A_S versus photon energy obtained from (a) the dc Fourier coefficient I'_0 as in Fig. 7.12 and (b) the zone-difference phase function Φ'; (c) rotating-polarizer phase angle P_S obtained in the $I'_0 - \Theta'$ calibration of (a), plotted versus the pixel group number k. The values indicated in (a) and (b) are the average offset angles, whereas the values in (c) indicate the best fit $k = 1$ intercept and the slope. The pixel group range of the experimental data in (c) corresponds to 1.5–3.0 eV. [Adapted from Ref. 53.]

sample holder can be translated so that the incident beam probes first the static calibration sample surface and then the starting sample surface of interest to be used for the real time measurements. Thus in either calibration situation, $r_{p,\mathrm{cal}}$ and $r_{s,\mathrm{cal}}$ can be computed based on a knowledge of the thickness of any layers on the calibration ('cal') sample surface as well as the optical properties of the component materials of this sample. From this information, one can determine the instrument response function I_{00} from the measurement of I'_0 for the calibration surface according to:

$$I_{00} = \frac{I'_{0,\mathrm{cal}}}{\left\{|r_{p,\mathrm{cal}}|^2 \cos^2(A-A_S) + |r_{s,\mathrm{cal}}|^2 \sin^2(A-A_S)\right\}}. \tag{7.11}$$

Here, the subscript 'cal' is used to denote data characteristic of the calibration sample, either the substrate or starting surface (if it is sufficiently well known), or otherwise a Si wafer that partially eclipses the substrate. In the latter situation, it is assumed that I_{00} is unchanged upon translating the sample holder from viewing the Si wafer to viewing the substrate or starting surface. Potential alignment errors can be minimized by mounting two Si wafers on the substrate surface (e.g., at the top and bottom of the substrate for a horizontal plane of incidence) and performing an appropriate average of the measured $I'_{0,\mathrm{cal}}$ spectra.

7.3.1.5 Case Study in Error Analysis: Source Polarization

Imperfections in the polarization generation and detection stages of the rotating-polarizer multichannel ellipsometer can lead to errors in the experimental results obtained in calibration and data reduction. Furthermore, sample imperfections can be present, meaning that the sample characteristics in actuality are different from those assumed in the simulations of Sections 7.3.1.1 through 7.3.1.4. These can also lead to erroneous results in calibration and data reduction. Here the effects of weak source polarization on the calibration and data reduction for a rotating-polarizer multichannel ellipsometer will be outlined briefly. This treatment demonstrates a general approach that can be applied to correct for a variety of instrument and sample errors as long as they can be simulated in the calibration and data reduction equations.

Source polarization can be described in terms of a normalized Stokes vector given by

$$(1 \quad v_{s2} \quad v_{s3} \quad v_{s4})^{\mathrm{T}} = (1 \quad p_s \cos 2Q_s \cos 2\chi_s \quad p_s \sin 2Q_s \cos 2\chi_s \quad p_s \sin 2\chi_s)^{\mathrm{T}}$$

for the collimated light beam entering the rotating polarizer. Here, Q_s is the tilt angle (measured with respect to the plane of incidence) for the polarization

ellipse associated with the polarized part of the light beam; $\chi_s = \arctan e_s$ is the ellipticity angle of the polarization ellipse with e_s being the ellipticity; and p_s ($\ll 1$) is the degree of polarization. In the absence of the source polarization error $p_s = 0$, and the error-free equations of Section 7.3.1.3 are applicable. The presence of source polarization can be noted from the appearance of $4\omega_P$ Fourier coefficients in the detected waveform of Eq. (7.3a). Another approach is to set the ellipsometer in straight-through without a sample and to remove the analyzer. In this configuration, the presence of source polarization is noted from the appearance of $2\omega_P$ Fourier coefficients in the waveform as measured at a polarization-insensitive detector. In both cases, the Fourier coefficients depend sensitively on the source orientation through the angle Q_s in the source beam Stokes vector.

First, the data reduction equations will be described for rotating-polarizer multichannel ellipsometry with a source polarization error. The first step in data reduction involves applying the transformations of Eqs. (7.5a) through (7.5c) to the measured dc and $2\omega_P$ Fourier coefficients, where P_S must be determined in calibration equations modified by source-polarization error correction as described shortly. The second step involves correcting the transformed dc and $2\omega_P$ Fourier coefficients $\{I_0, (\alpha, \beta)\}$ according to:

$$\alpha_c = \frac{\alpha - v_{s2} + (1/2)v_{s3}\,(v_{s2}\beta - v_{s3}\alpha)}{1 - (1/2)(v_{s2}\alpha + v_{s3}\beta)}, \tag{7.12a}$$

$$\beta_c = \frac{\beta - v_{s3} + (1/2)v_{s2}\,(v_{s3}\alpha - v_{s2}\beta)}{1 - (1/2)(v_{s3}\beta + v_{s2}\alpha)}, \tag{7.12b}$$

$$I_{0c} = \frac{I_0}{1 + (1/2)(v_{s2}\alpha_c + v_{s3}\beta_c)}. \tag{7.12c}$$

These equations are exact, and in the third step, the resulting corrected Fourier coefficients $\{I_{0c}, (\alpha_c, \beta_c)\}$ are substituted into Eqs. (7.6a) through (7.6c) to extract the corrected ellipsometry angles and polarized reflectance. In order to perform these steps in practice, the calibration data I_{00}, P_S, and A_S are required, as are the two error parameters v_{s2} and v_{s3}. The determination of I_{00} is straightforward and parallels the detailed discussion of Section 7.3.1.4. In particular, $I_{00} = I_{0c,\mathrm{cal}}/R_{\mathrm{cal}}(A')$ where $I_{0c,\mathrm{cal}}$ is the dc Fourier coefficient of the calibration sample, corrected according to Eq. (7.12c), and $R_{\mathrm{cal}}(A')$ is the reflectance spectrum of the calibration sample.

First, equations are presented that describe the source polarization corrections to A_S and P_S as determined from the residual-function/phase-function calibration of Eqs. (7.7c), (7.7d), (7.8c), and (7.8d). These expressions are given to first order in the source polarization parameters v_{s2} and v_{s3} by:

Residual function calibration for $A \approx A_S$:

$$A_{Sc} = A_{S,0} - [(\tan\psi \cos\Delta)/(4 \sin^2\Delta)]v_{s3}; \tag{7.13a}$$

Residual function calibration for $A \approx A_S + (\pi/2)$:

$$A_{Sc} = A_{S,\pi/2} + [(\cot\psi \cos\Delta)/(4 \sin^2\Delta)]v_{s3}; \tag{7.13b}$$

Phase/residual function calib. for $A \approx A_S$:

$$P_{Sc} = P_{S,0} - \{1 + [(\cot^2\Delta)/2]\}(v_{s3}/2); \tag{7.13c}$$

Phase/residual function calib for $A \approx A_S + (\pi/2)$:

$$P_{Sc} = P_{S,\pi/2} + \{1 + [(\cot^2\Delta)/2]\}(v_{s3}/2). \tag{7.13d}$$

The corresponding equations that describe the source polarization corrections to A_S and P_S, as determined from the zone-difference-phase-function/phase-function calibration of Eqs. (7.8c) and (7.9c), are given by:

Zone-difference-phase-function calibration:

$$A_{Sc} = A_{S,0} + [(\tan 2\psi)/(2 \cos\Delta)]v_{s3}; \tag{7.13e}$$

Phase/zone-difference-phase-function calibration:

$$P_{Sc} = P_{S,0} + [(\sec 2\psi)/2]v_{s3}. \tag{7.13f}$$

The equations that describe the source polarization corrections to A_S and P_S, as determined from the irradiance-function/phase-function calibration of Eqs. (7.8c), (7.8d), (7.10a), and (7.10b), are given by:

Irradiance function calibration for $A \approx A_S$:

$$A_{Sc} = A_{S,0} + [(\tan 2\psi \cos\Delta)/4]v_{s3}; \tag{7.13g}$$

Irradiance function calibration for $A \approx A_S + (\pi/2)$:

$$A_{Sc} = A_{S,\pi/2} + [(\tan 2\psi \cos\Delta)/4]v_{s3}; \tag{7.13h}$$

Phase/irradiance function calibration for $A \approx A_S$:

$$P_{Sc} = P_{S,0} - \{1 - [(\cos^2\psi \cos^2\Delta)/\cos 2\psi]\}(v_{s3}/2); \tag{7.13i}$$

Phase/irradiance function calibration for $A \approx A_S + (\pi/2)$:

$$P_{Sc} = P_{S,\pi/2} + \{1 + [(\sin^2\psi \cos^2\Delta)/\cos 2\psi]\}(v_{s3}/2). \tag{7.13j}$$

In Eqs. (7.13a) through (7.13j), A_{Sc} and P_{Sc} are the offset and phase angles, including the first-order corrections for the source polarization error. If the error corrections are performed properly and no other errors are present, then these angles should not depend either on which calibration procedure is adopted or on whether the procedure is performed near $A' = 0$ or near $A' = \pi/2$. In contrast, $A_{S,0}$, $P_{S,0}$, $A_{S,\pi/2}$, $P_{S,\pi/2}$ are the analyzer offset and polarizer phase angles obtained near $A' = 0$ (subscript '0') and near $A' = \pi/2$ (subscript '$\pi/2$') from the earlier calibration procedures of Section 7.3.1.4, disregarding the effects of the source polarization error on the Fourier coefficients. Owing to the presence of the source polarization error, $A_{S,0}$ is no longer equal to $A_{S,\pi/2}$ and $P_{S,0}$ is no longer equal to $P_{S,\pi/2}$, in general. In addition, the corresponding angles obtained by the different calibration procedures differ, as well. For example, the values of $A_{S,0}$ obtained by the residual function procedure [Eq. (7.13a)] and by the irradiance function procedure [Eq. (7.13g)] are not equal. These values could be distinguished by a third subscript; however, for simplicity this method of differentiation is omitted. Finally, the values of (ψ, Δ) in the correction terms of Eqs. (7.13a) through (7.13j) are the "zero-order" values obtained as in Section 7.3.1.3, disregarding the source polarization error. It should be noted that in the determination of the correction terms, the sign of Δ is not required.

So far, no direction has been provided on the determination of the small quantities v_{s2} and v_{s3}. In fact, Eqs. (7.13a) through (7.13d) and Eqs. (7.13g) and (7.13h) provide multiple expressions for A_{Sc} and P_{Sc} that include v_{s3}. Furthermore, the minimum values of the residual function can be shown to depend on v_{s2}. From these considerations, one can arrive at the following expressions:

All calibrations:

$$v_{s2} = (F'_{\min,\pi/2} - F'_{\min,0})/2; \tag{7.13k}$$

Residual/phase function calibration:

$$v_{s3} = (P_{S,0} - P_{S,\pi/2}) - [(A_{S,0} - A_{S,\pi/2}) \cos\Delta \sin 2\psi]; \tag{7.13l}$$

Irradiance/phase function calibration:

$$v_{s3} = \frac{(P_{S,0} - P_{S,\pi/2})}{1 - (1/2)\cos^2\Delta}. \tag{7.13m}$$

In Eq. (7.13k), $F'_{\min,\pi/2}$ and $F'_{\min,0}$ designate the values of the residual function at the minima which occur near $A' = \pi/2$ and $A' = 0$. In the absence of the source polarization error, these two values are equal and in fact

vanish; otherwise they are non-zero and are well-defined irrespective of (ψ, Δ) for the sample. Equations (7.13k) and (7.13l) are applied under conditions for which the residual-function/phase-function approach provides P_{Sc} and A_{Sc} with high sensitivity, e.g., when $30 < |\Delta| < 150°$. In contrast, Eqs. (7.13k) and (7.13m) can be applied without restrictions on Δ.

It has been pointed out that weak polarization of the Xe source output is one of the most difficult systematic errors to eliminate in rotating-polarizer multichannel ellipsometry. It is then appropriate to reassess whether the situation can be improved by conversion to a rotating-analyzer multichannel ellipsometer configuration. As noted in Section 7.2.2, if the spectrograph and multichannel array detection system are placed immediately after a rotating analyzer, the errors generated by the polarization sensitivity of this detection system are likely to exceed any errors associated with source polarization in the rotating-polarizer configuration. In contrast, if a sufficient length of fiber optic cable or other effective depolarizing element is placed between the rotating analyzer and the spectrograph, then the errors associated with polarization sensitivity of the detection system can be reduced to tolerable levels. In order to correct remaining errors due to polarization sensitivity of the detection system for a rotating-analyzer multichannel ellipsometer, adapted forms of Eqs. (7.12) and (7.13) can be applied. Owing to the symmetry of the two configurations, all polarizer phase angles for the rotating-polarizer configuration in Eqs. (7.13) are replaced by analyzer phase angles for the rotating-analyzer; in addition, all fixed analyzer offset angles are replaced by polarizer offset angles. Furthermore, v_{s2} and v_{s3} for the rotating-polarizer configuration in Eqs. (7.12) and (7.13) are replaced by v_{d2} and v_{d3} for the rotating-analyzer configuration, given by:

$$v_{d2} = m_{d12} \cos 2Q_d - m_{d13} \sin 2Q_d, \tag{7.14a}$$

$$v_{d3} = m_{d12} \sin 2Q_d + m_{d13} \cos 2Q_d. \tag{7.14b}$$

In these expressions, m_{d12} and m_{d13} are the normalized Mueller matrix elements associated with the polarization modifying components of the detection system, and Q_d is the angle of the frame of reference of this Mueller matrix, measured with respect to the plane of incidence of the ellipsometer.

7.3.2 Single Rotating Compensator

As noted in Section 7.2.4, the single rotating-compensator multichannel ellipsometer in the $PSC_r(\omega_C)A$ configuration shown in Fig. 7.8(b) provides several advantages over the rotating-polarizer

multichannel ellipsometer in the $P_r(\omega_P)SA$ configuration.[17,44,45] These advantages arise from the fact that the rotating-compensator instrument provides all four Stokes vector components for the light beam reflected from the sample surface. In contrast, the rotating-polarizer instrument provides only three such components. With respect to Stokes vector measurements, the $PSC_r(\omega_C)A$ configuration can provide the sign of the ellipticity for the polarization state generated upon reflection from the sample. Thus, the sign of Δ can be determined, in contrast to the $P_r(\omega_P)SA$ configuration. In addition, the $PSC_r(\omega_C)A$ configuration provides accurate measurements of small ellipticity angles for the reflected beam polarization state. Thus, accurate measurements of Δ near 0° and ±180° can be obtained, again in contrast to the $P_r(\omega_P)SA$ configuration. Furthermore, the rotating-compensator instrument can provide the degree of polarization of the reflected beam, which can be useful for situations in which the sample depolarizes the incident light beam. Finally, this instrument is immune to source polarization and detector polarization sensitivity. The disadvantage of the $PSC_r(\omega_C)A$ configuration is the additional complexity resulting from two additional calibration parameters for the rotating compensator, the wavelength-dependent slow-to-fast axis phase shift, i.e., the retardance spectrum, of the compensator and the rotating-compensator phase angle.

7.3.2.1 Theoretical Description

For the rotating-compensator multichannel ellipsometer, the time-dependent waveform predicted at each pixel (or pixel group) of the photodiode array detector can be expressed in terms of the parameters of the Stokes vector for the beam reflected from the sample. This Stokes vector is given by $I_r(1 \quad p\cos 2Q\cos 2\chi \quad p\sin 2Q\cos 2\chi \quad p\sin 2\chi)^T$.[8] The parameters associated with the reflected beam polarization state include: (i) the tilt angle of the ellipse Q ($-90° < Q \leq 90°$), measured from the plane of incidence in a counterclockwise-positive sense, looking opposite to the beam direction; (ii) the ellipticity angle of the ellipse $\chi = \arctan e$ ($-45° \leq \chi \leq 45°$; $-1 \leq e \leq 1$), where e is the ellipticity, i.e., the semiminor-to-semimajor axis ratio of the polarization ellipse; (iii) the irradiance in the reflected beam I_r; and (iv) the degree of polarization p ($0 \leq p \leq 1$). The Stokes vector parameters $\{(Q, \chi), p, I_r/I_i\}$ (where the reflected beam irradiance I_r is normalized by the incident beam irradiance I_i) are employed rather than the sample parameters $\{(\psi, \Delta), R(P')\}$ in the theoretical development. In this way, one can fully characterize samples that not only modify the incident linear polarization, but also randomly depolarize some fraction of the incident beam. In the sample

parameter set, $R(P')$ is the specular beam reflectance of the sample for an incident beam linearly polarized at an angle P' with respect to the plane of incidence.

Based on this motivation, the following expressions can be derived for the multichannel ellipsometer in the $PSC_r(\omega_C)A$ configuration:[17]

$$I(t) = I_0\{1 + \alpha_2 \cos 2(\omega_C t - C_S) + \beta_2 \sin 2(\omega_C t - C_S) + \alpha_4 \cos 4(\omega_C t - C_S) + \beta_4 \sin 4(\omega_C t - C_S)\}, \tag{7.15a}$$

$$I_0 = I_{00} a_0 (I_r/I_i), \tag{7.15b}$$

$$a_0 = 1 + \{p \cos^2(\delta/2) \cos 2\chi \cos 2(A' - Q)\}, \tag{7.16}$$

$$B_2 = \alpha_2 + i\beta_2 = \frac{-ip \sin\delta \sin 2\chi \exp(2i A')}{a_0}, \tag{7.17}$$

$$B_4 = \alpha_4 + i\beta_4 = \frac{p \sin^2(\delta/2) \cos 2\chi \exp[2i (A' + Q)]}{a_0}. \tag{7.18}$$

The Mueller matrix for the compensator given in Ref. 8 was used in this derivation. In Eq. (7.15a), $\{I_0, (\alpha_{2n}, \beta_{2n}); n = 1,2\}$ are the theoretically predicted average irradiance and $2n\omega_C$ cosine and sine Fourier coefficients of the waveform, respectively, where ω_C is the compensator angular rotation frequency. The theoretically predicted Fourier coefficients in these equations are explicitly corrected for the phase angle of the rotating compensator through C_S. From Eq. (7.15a), it is clear that $-C_S$ is the angle of the compensator fast axis at time zero of the data collection process for the given pixel (or pixel group). This angle is measured with respect to the plane of incidence in a counterclockwise-positive sense, looking opposite to the beam direction. In Eq. (7.15b), I_{00} is the spectral response function of the ellipsometer. In Eqs. (7.16) through (7.18), δ is the slow-to-fast axis phase shift (i.e., the retardance) of the compensator, given by $\delta = 2\pi d(n_S - n_F)/\lambda$ for a single-plate compensator of thickness d, where n_S and n_F are the indices of refraction for propagation with electric fields vibrating along the slow (S) and fast axes (F). (Because $n_S > n_F$, then $\delta > 0$.) In addition, $A' = A - A_S$ is the true angle of the analyzer transmission axis with respect to the plane of incidence, where A is the analyzer scale reading and A_S is the correction to the scale reading. In Eqs. (7.17) and (7.18), the complex quantity B_{2n} is used to describe Fourier coefficients in terms of the Stokes vector parameters, the compensator retardance, and the true analyzer angle A'. To determine the individual Fourier coefficients, the real and imaginary parts of these equations are equated.

Next, theoretical expressions are provided that allow one to calculate the beam parameters in Eqs. (7.15) through (7.18) $\{(Q, \chi), p, I_r/I_i\}$ from the sample parameters $\{(\psi, \Delta), D, R(P')\}$. Here it is assumed that the interaction of the incident beam with the sample consists of polarization modification along with completely random depolarization due to the collection of near-specular, multiply scattered light. Thus, in the sample parameter set, the polarization transfer factor D has been added. This factor characterizes the depolarizing effect of the sample on any incident beam polarization state. In order to convert sample parameters to beam (or Stokes vector) parameters, the matrix equation $V_r = MV_i$ is applied, where

$$V_i = I_i \,(1 \quad \cos 2P' \quad \sin 2P' \quad 0)^{\mathrm{T}}$$

and

$$V_r = I_r \,(1 \quad p\cos 2Q \cos 2\chi \quad p\sin 2Q \cos 2\chi \quad p\sin 2\chi)^{\mathrm{T}}$$

are the incident and reflected beam Stokes vectors, and M is the Mueller matrix of the sample.[8] In the expression for V_i, P' is the true angle of the polarizer transmission axis measured with respect to the plane of incidence. For a sample that generates a randomly depolarized component in the detected beam as a result of multiple scattering, the non-zero unnormalized 4×4 Mueller matrix elements are given by:[54,55]

$$M_{11} = R_u D, \tag{7.19a}$$

$$M_{12} = M_{21} = -R_u \cos 2\psi, \tag{7.19b}$$

$$M_{22} = R_u, \tag{7.19c}$$

$$M_{33} = M_{44} = R_u \sin 2\psi \cos \Delta, \tag{7.19d}$$

$$M_{34} = -M_{43} = R_u \sin 2\psi \sin \Delta. \tag{7.19e}$$

The remaining Mueller matrix elements are zero. In Eqs. (7.19), R_u is the reflectance of the sample for unpolarized light, defined as the ratio of the reflected to incident irradiances for the specular beam, in the absence of near-specular scattering (i.e., when $D = 1$). As an example, the general approach described in this paragraph may be applicable for the analysis of thin film nucleation at low density on a smooth substrate (i.e., such that the nuclei spacing or size is on the order of, or greater than, the probe wavelength) or for the analysis of film growth on macroscopically rough substrates. In both cases, some fraction of near-specular, multiply-scattered light is collected by the detection system.[55] For the simple

Mueller matrix of Eqs. (7.19) to be valid, the detected near-specular scattering must be incoherent and unpolarized, and the p and s incident polarizations must be scattered with equal probability. In the absence of such near-specular scattering, $D = 1$. Finally, this same mathematical description can also be applied for the inclusion of instrument errors due to spectrograph stray light. Under certain conditions, the effect of stray light is identical to that of a randomly depolarizing sample. Thus by including D in the analysis, the stray light errors can be corrected.

Using the Mueller matrix of Eqs. (7.19), the following beam-parameter/sample-parameter relationships can be derived:

$$\frac{I_r}{I_i} = R(P')\frac{D - \cos 2\psi \cos 2P'}{1 - \cos 2\psi \cos 2P'}, \tag{7.20a}$$

$$p = \frac{1 - \cos 2\psi \cos 2P'}{D - \cos 2\psi \cos 2P'}, \tag{7.20b}$$

$$Q = [(1/2)\arctan(v_{r3}/v_{r2})] + [(\pi/2)u(-v_{r2})\mathrm{sgn}(v_{r3})], \tag{7.20c}$$

$$\chi = (1/2)\arctan\left\{\frac{v_{r4}}{|\sqrt{v_{r2}^2 + v_{r3}^2}|}\right\}, \tag{7.20d}$$

where the polarized specular reflectance $R(P')$ is defined as the ratio of the incident to reflected irradiances in the absence of near-specular scattering (i.e., when $D = 1$). In Eqs. (7.20c) and (7.20d), v_{r2}, v_{r3}, and v_{r4} are the Stokes vector components of the beam reflected from the surface normalized by the first component V_{r1}, i.e., $v_{rj} = V_{rj}/V_{r1}$ ($j = 2, 3, 4$). These normalized components are defined in terms of the sample parameters by:

$$v_{r2} = \frac{\cos 2P' - \cos 2\psi}{D - \cos 2\psi \cos 2P'} \tag{7.20e}$$

$$v_{r3} = \frac{\sin 2P' \sin 2\psi \cos \Delta}{D - \cos 2\psi \cos 2P'} \tag{7.20f}$$

$$v_{r4} = \frac{-\sin 2P' \sin 2\psi \sin \Delta}{D - \cos 2\psi \cos 2P'} \tag{7.20g}$$

Finally, in the term on the right of Eq. (7.20c), $u(x) = 0$ for $x < 0$; $u(x) = +1$ for $x \geq 0$; $\mathrm{sgn}(x) = -1$ for $x < 0$; and $\mathrm{sgn}(x) = +1$ for $x \geq 0$. This term is a quadrant correction based on the assumption that the inverse tangent computation yields a result between $-\pi/2$ and $+\pi/2$.

7.3.2.2 Data Collection

For an error-free system in the rotating compensator configuration, the irradiance on any given pixel (or pixel group) obeys the following experimental expression:

$$I'(t) = I'_0(1 + \alpha'_2 \cos 2\omega_C t + \beta'_2 \sin 2\omega_C t + \alpha'_4 \cos 4\omega_C t + \beta'_4 \sin 4\omega_C t), \tag{7.21}$$

where ω_C is the angular frequency of the rotating compensator and $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1, 2\}$ are dc and normalized Fourier coefficients to be determined experimentally. Equation (7.21) differs from the theoretical expression Eq. (7.15a) in that the waveform does not include the compensator phase angle C_S. This term is dropped from the phase of the experimental waveform because it cannot be determined until the calibration is performed.

If the detector is operated in an integrating mode for fully parallel data acquisition and the array is read out N times per half rotation of the rotating compensator, leading to N equal exposure times $t_e = \pi/N\omega_C$, then N spectra are generated having the following form:

$$S_j = I'_0 \int_{(j-1)\pi/N\omega_C}^{j\pi/N\omega_C} \left\{1 + \sum_{n=1}^{2} [\alpha'_{2n} \cos 2n\omega_C t + \beta'_{2n} \sin 2n\omega_C t]\right\} dt; \quad (j = 1,\ldots, N), \tag{7.22a}$$

$$= \frac{\pi I'_0}{N\omega_C} + \sum_{n=1}^{2} \frac{I'_0}{n\omega_C} \left[\sin\left(\frac{n\pi}{N}\right)\right] \times \left[\alpha'_{2n} \cos\frac{(2j-1)n\pi}{N} + \beta'_{2n} \sin\frac{(2j-1)n\pi}{N}\right]. \tag{7.22b}$$

For each pixel of the photodiode array, Eq. (7.22b) represents a system of N equations in five unknowns $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1,2\}$. Thus, for $N = 5$, the unknowns are exactly determined. For $N > 5$, however, the unknowns are overdetermined, and an effective method of solution must be established to utilize the full data set $\{S_j, j = 1, \ldots, N\}$ in the determination of $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1,2\}$.

7.3.2.3 Data Reduction

Next, the approach will be described to convert the raw spectra in $\{S_j, j = 1, \ldots, N\}$, i.e., the waveform integrations, first to spectra in the

Fourier coefficients $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1,2\}$, then to spectra in the Stokes vector parameters $\{(Q, \chi), p, I_r/I_i\}$, and finally to the sample parameters $\{(\psi, \Delta), D, R(P')\}$. For the rotating-compensator multichannel ellipsometer with $N = 5$ the experimental dc and normalized $2\omega_C$ and $4\omega_C$ Fourier coefficients can be determined according to the following equations:[17]

$$\alpha'_2 = 0.534480 + \{[1.195133(S_1 - S_2 - S_4 + S_5) - 2.672398\, S_3]/\pi I'_0\}, \tag{7.23a}$$

$$\beta'_2 = [1.256637\,(S_1 - S_5) + 2.033282\,(S_2 - S_4)]/\pi I'_0, \tag{7.23b}$$

$$\alpha'_4 = [0.816612\,(S_1 + S_5) - 2.137919\,(S_2 + S_4) + 2.642613\,S_3]/\pi I'_0\,, \tag{7.23c}$$

$$\beta'_4 = [2.513274\,(S_1 - S_5) - 1.553289\,(S_2 - S_4)]/\pi I'_0\,, \tag{7.23d}$$

$$I'_0 = (1/\pi)(S_1 + S_2 + S_3 + S_4 + S_5)\,. \tag{7.23e}$$

The disadvantage of this approach is that there are no consistency checks to assess the higher order Fourier coefficients, which should vanish. As an alternative, $N = 8$ can be chosen in which case the following simpler equations apply:

$$\alpha'_2 = (S_1 + S_2 - S_3 - S_4 - S_5 - S_6 + S_7 + S_8)/2I'_0, \tag{7.24a}$$

$$\beta'_2 = (S_1 + S_2 + S_3 + S_4 - S_5 - S_6 - S_7 - S_8)/2I'_0, \tag{7.24b}$$

$$\alpha'_4 = (S_1 - S_2 - S_3 + S_4 + S_5 - S_6 - S_7 + S_8)/2I'_0, \tag{7.24c}$$

$$\beta'_4 = (S_1 + S_2 - S_3 - S_4 + S_5 + S_6 - S_7 - S_8)/2I'_0, \tag{7.24d}$$

$$I'_0 = (1/\pi)(S_1 + S_2 + S_3 + S_4 + S_5 + S_6 + S_7 + S_8). \tag{7.24e}$$

For example, an expression analogous to Eq. (7.4d) is:

$$\beta'_8 = (S_1 - S_2 + S_3 - S_4 + S_5 - S_6 + S_7 - S_8)/2I'_0, \tag{7.25}$$

which must vanish in the absence of errors. Two other linearly independent consistency checks can be derived from the requirement that the $6\omega_C$ Fourier coefficients must vanish in the absence of errors.

Once the spectra in the experimental waveform parameters $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1,2\}$ have been determined, they need to be converted to the theoretical

parameters $\{I_0, (\alpha_{2n}, \beta_{2n}); n = 1,2\}$ so that the spectra in the beam and sample parameters can be deduced. In analogy with Eqs. (7.5a) and (7.5d), the following compact expressions can be applied:

$$I_0 = I'_0\,, \tag{7.26a}$$

$$(\alpha_{2n} \quad \beta_{2n})^{\mathrm{T}} = \Re(2nC_S)(\alpha'_{2n} \quad \beta'_{2n})^{\mathrm{T}};\ (n = 1,2), \tag{7.26b}$$

where $\Re(2nC_S)$ represents the 2×2 rotation transformation by the angle $2nC_S$ ($n = 1, 2$). As usual, the superscript 'T' represents the transpose operation that converts the two row vectors to column vectors.

Now that the transformed coefficients are available, the theoretical expressions given in Eqs. (7.15b) and (7.16) through (7.18) can be inverted to obtain information on the reflected beam Stokes vector:[17]

$$Q = \frac{1}{2}\arctan\left(\frac{\beta_4}{\alpha_4}\right) - A' + \left[\frac{\pi}{2}u(-\alpha_4)\mathrm{sgn}(\beta_4)\right]; \tag{7.27}$$

$$\begin{aligned}\chi &= \frac{1}{2}\arctan\left\{\frac{\alpha_2\cos 2(A'+Q)\tan(\delta/2)}{2\alpha_4\sin 2A'}\right\} \\ &= \frac{1}{2}\arctan\left\{\frac{-\beta_2\sin 2(A'+Q)\tan(\delta/2)}{2\beta_4\cos 2A'}\right\}\end{aligned} \tag{7.28a}$$

$$= \frac{1}{2}\arctan\left\{\frac{\sqrt{\alpha_2^2+\beta_2^2}}{2\sqrt{\alpha_4^2+\beta_4^2}}\tan(\delta/2)\right\}; \tag{7.28b}$$

$$p = \frac{\Re_Q}{\cos 2\chi\cos 2(A'-Q)\{1-(1+\Re_Q)\cos^2(\delta/2)\}}, \tag{7.29a}$$

$$\Re_Q = \alpha_4\cos 4Q + \beta_4\sin 4Q; \tag{7.29b}$$

$$\frac{I_r}{I_i} = \frac{I_0}{I_{00}\{1+p\cos^2(\delta/2)\cos 2\chi\cos 2(A'-Q)\}}. \tag{7.30}$$

The quadrant correction at the right in Eq. (7.27) ensures that the range of $Q + A'$ is given by $-90° < Q + A' \leq 90°$. However, because Q also spans the range of $-90° < Q \leq 90°$, an additional correction of $\pm\pi$ may be required in Eq. (7.27) depending on the values of Q and A'. In addition, χ spans the range $-45° \leq \chi \leq 45°$. As a result, either of the relationships in Eq. (7.28a) can provide the sign of χ, whereas Eq. (7.28b) is best for determining the magnitude of χ.

In adopting a Mueller matrix of the form given by Eqs. (7.19), it is assumed that polarization modification is accompanied by random depolarization owing to the collection of near-specular, multiply-scattered light. For this situation, inversion of Eqs. (7.20) provides the sample parameters $\{(\psi, \Delta), D, R(P')\}$ in terms of the beam parameters $\{(Q, \chi), p, I_r/I_i\}$:

$$\tan\psi = |\tan P'|\{sy + |\sqrt{y^2 + 1}|\}; \tag{7.31a}$$

$$s = \text{sgn}(\psi - |P'|); \tag{7.31b}$$

$$y = \frac{|\cos 2Q|}{|\sqrt{\tan^2 2\chi + \sin^2 2Q}|}, \tag{7.31c}$$

$$\tan\Delta = -\frac{\tan 2\chi}{\sin 2Q}, \tag{7.32}$$

$$D = \frac{\{1 - (1 - p)\cos 2\psi \cos 2P'\}}{p}, \tag{7.33}$$

$$R(P') = p\frac{I_r}{I_i} = |r_p|^2 \cos^2 P' + |r_s|^2 \sin^2 P'. \tag{7.34}$$

Because ψ spans the range of $0° \le \psi \le 90°$, then Eqs. (7.31) define ψ uniquely. However, because Δ spans the range $-180° \le \Delta \le 180°$, then expressions for $\sin\Delta$ and $\cos\Delta$ are also needed to determine the correct quadrant of Δ. These expressions are given by:

$$\cos\Delta = \frac{\cos 2\chi \sin 2Q\,(1 - \cos 2P' \cos 2\psi)}{\sin 2P' \sin 2\psi}, \tag{7.35a}$$

$$\sin\Delta = \frac{-\sin 2\chi\,(1 - \cos 2P' \cos 2\psi)}{\sin 2P' \sin 2\psi}. \tag{7.35b}$$

For a bulk substrate material presenting a single interface to the ambient, it should be recalled that $0° \le \Delta \le 180°$ within the optics ($e^{i\omega t}$) convention used here and in Ref. 8.

As an example of the capabilities of the rotating-compensator multichannel ellipsometer, Fig. 7.14 depicts results obtained in real time during the growth of a 4000 Å thick diamond film on a silicon wafer substrate by microwave plasma-enhanced chemical vapor deposition.[17] These results include the time evolution of the spectra in the tilt and ellipticity angles and the degree of polarization $\{(Q, \chi), p\}$ that define the normalized

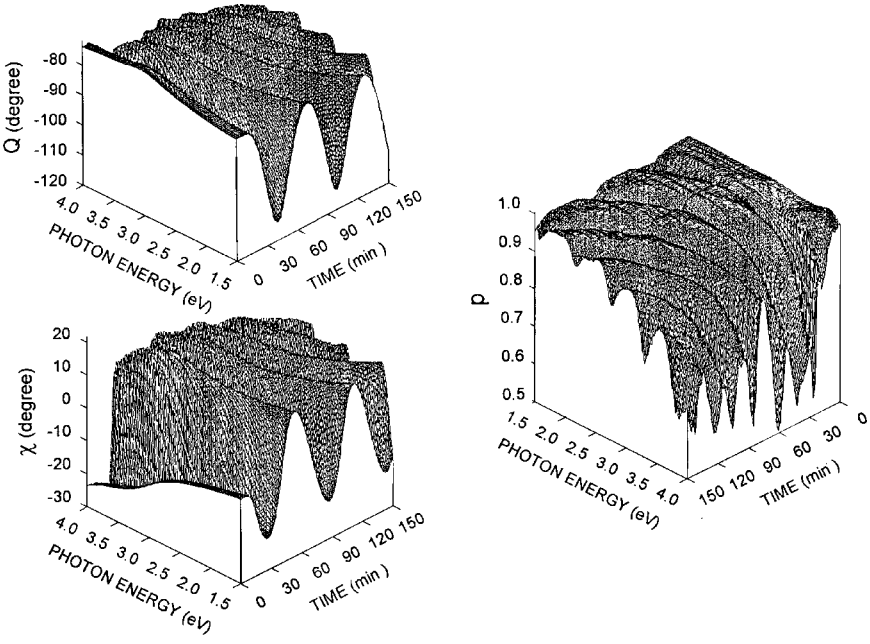


Figure 7.14 Time evolution of the spectra in the tilt and ellipticity angles and the degree of polarization $\{(Q, \chi), p\}$ that define the normalized Stokes vector, collected with a rotating-compensator multichannel ellipsometer using acquisition and repetition times of 10 and 35 s, respectively, during the growth of a 4000 Å thick diamond film on a silicon wafer substrate. This film was prepared by microwave plasma-enhanced chemical vapor deposition at a substrate temperature of 812°C. To construct the three-dimensional plots, 104 points were collected versus photon energy and 292 triplets were collected versus time. For clarity of presentation, positive values of Q ($0° \leq Q \leq 90°$) are plotted as $Q-180°$. [Reproduced from Ref. 17.]

Stokes vector. The acquisition time for a single triplet of spectra was 10 s and the repetition time was 35 s. During these times, 4.5 and 16 Å of diamond are deposited in the bulk film growth regime (after nucleation and coalescence). The longer acquisition and repetition times in this case include both "light" and "dark" data cycles, each an average over 80 mechanical rotations with a shutter open and closed, respectively. The waveform integrals from the dark cycle are required in order to correct those from the light cycle for the effects of radiative emission by the plasma and thermal emission by the substrate (which is heated to 812°C). To construct the plots of Fig. 7.14, 104 points were collected versus photon energy and 292 triplets were collected versus time for a total number of data points of ~9.1×10^4.

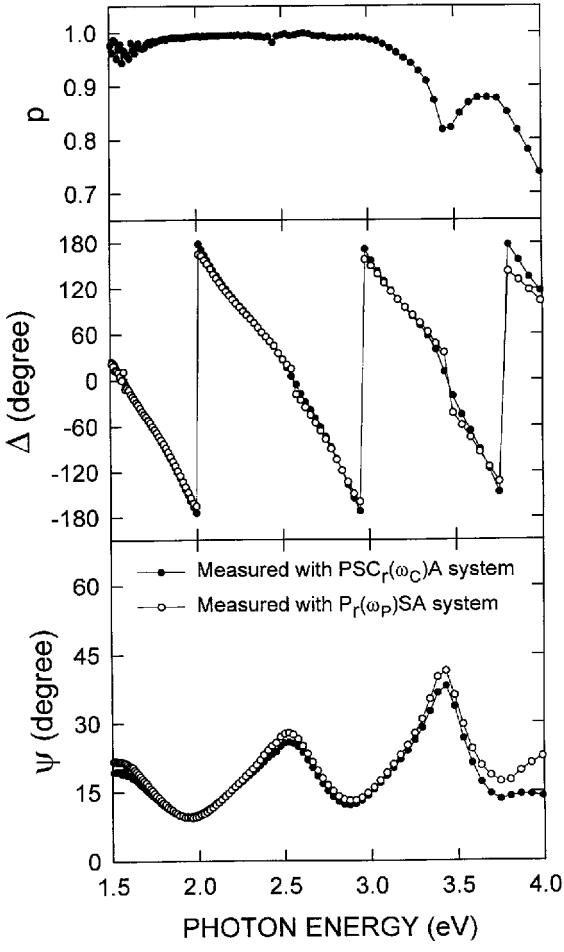


Figure 7.15 Ellipsometric spectra (ψ, Δ) for a 2000 Å thick diamond film deposited by microwave plasma-enhanced chemical vapor deposition on a crystalline silicon wafer substrate and measured ex situ at room temperature. Spectra obtained using rotating-polarizer [P$_r$(ω_P)SA] and rotating-compensator [PSC$_r$(ω_C)A] multichannel ellipsometers are shown for comparison. The degree of polarization p of the reflected beam is shown at the top, accessible only in the rotating-compensator configuration. [Reproduced from Ref. 17.]

Figure 7.15 shows (ψ, Δ) spectra from 1.5 to 4.0 eV for a 2000 Å thick diamond film deposited by microwave plasma-enhanced chemical vapor deposition on a crystalline silicon wafer substrate.[17] Spectra obtained ex situ using the rotating-polarizer [P$_r$(ω_P)SA] and rotating-compensator [PSC$_r$(ω_C)A] multichannel ellipsometer configurations are shown for

comparison. The spectra in the rotating-compensator configuration were obtained by applying Eqs. (7.31) and (7.32) in the data reduction process with $P' = 45°$. The degree of polarization p of the reflected beam is shown at the top, accessible only in the rotating-compensator configuration. The variations in p can be attributed to (i) pseudo-depolarization by the sample caused by film thickness nonuniformity over the area probed by the incident beam and (ii) depolarization caused by multiple scattering from diamond crystallites. The rotating-compensator configuration is advantageous for the measurement of Δ in such applications because (i) the sign is determined; (ii) higher sensitivity is obtained in the neighborhood of $\Delta = 0°$ and $\pm 180°$ (since $\tan \Delta$ is measured rather than $\cos \Delta$); and (iii) errors are avoided that arise from a misinterpretation of the depolarization as an increase in ellipticity angle in the $P_r(\omega_P)SA$ configuration. For ease of comparison in Fig. 7.15, the signs of Δ in the $P_r(\omega_P)SA$ configuration are set to match those obtained in the $PSC_r(\omega_C)A$ configuration for the corresponding photon energy. Advantages (ii) and (iii) noted above for the rotating-compensator configuration lead to an obvious improvement in the continuity in the Δ spectra near 0° where the reflected beam polarization is nearly linear ($\chi = 0°$).

7.3.2.4 Calibration

Ellipsometers in the rotating-compensator configuration $PSC_r(\omega_C)A$ are used less widely than their counterparts in the rotating-polarizer $P_r(\omega_P)SA$ and analyzer $PSA_r(\omega_A)$ configurations. As a result, fewer descriptions of calibration procedures are available.[17,56,57] Furthermore, the calibration procedures are much more complex than those of the rotating-polarizer and analyzer instruments. Not only must one determine the polarizer and analyzer offset angles (P_S, A_S) and the ellipsometer spectral response function I_{00}, but also the compensator retardance spectrum and its rotational phase (δ, C_S). In the rotating-compensator configuration, however, partial self-calibration is possible since there are five measured waveform parameters $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1,2\}$, but only four deduced sample parameters $\{(\psi, \Delta), D, R(P')\}$. Thus, in principle, it is possible to extract one calibration parameter directly from the real time sample measurements, i.e., without the need for a separate calibration step.

The conventional sequence for calibration of the rotating-compensator multichannel ellipsometer involves (i) measurement of the retardance spectra δ, (ii) determination of the optical element offset and phase angles (P_S, C_S, A_S), and (iii) measurement of the spectral response function I_{00}. According to (i), δ is determined once in the initial development of the ellipsometer (or subsequently in major system realignments) using a

straight-through ellipsometer configuration without a sample. The offset and phase angles are determined in a separate step after sample alignment and just prior to the real time measurements. Although the calibration can be performed for any specular starting sample, the sample properties (i.e., the ψ and Δ values) determine the details of the procedure, as in Section 7.3.1.4. In general, the spectral response function is also determined in a separate calibration step just prior to the real time measurements. In this case, however, the starting characteristics of the sample must be known so that its spectroscopic reflectance can be predicted. Essentially, this latter calibration is performed in the same way as for the rotating-polarizer instrument (Section 7.3.1.4), with the only difference being the calibration equation.

An alternative calibration sequence has been proposed recently in order to account for possible effects of sample alignment on the beam path through the rotating compensator.[57] Such effects lead to a weak sample alignment-dependence in the retardance spectrum for the compensator that, in turn, can lead to irreproducibilities in data reduction from run to run in real time measurements. To avoid such irreproducibilities, the retardance spectrum is measured with the sample in place after the offset and phase angle calibration, but before the spectral response function calibration.

Retardance Calibration

Here three possible approaches for determining the spectrum in δ are summarized. The first uses a separate single-channel ellipsometer in the $PCA_r(\omega_A)$ rotating-analyzer configuration and the second uses the $PC_r(\omega_C)A$ rotating-compensator multichannel ellipsometer configuration, both in straight-through. The third uses the $PSC_r(\omega_C)A$ sequence in the actual sample measurement configuration. The latter measurement can be performed on a run-to-run basis after mounting and aligning each sample, e.g., the starting substrate in a real time measurement. As noted above, this measurement is recommended in order to account for small run-to-run variations in δ that may arise due to variable alignments of the reflected beam with respect to the compensator rotation axis.

In the first approach, the compensator is mounted between the polarizer and analyzer of a separate single-channel, rotating-analyzer ellipsometer set for straight-through.[17] Initially, the fast-axis angle of the compensator C' is fixed, and a residual function calibration is performed near the quarter wavelength of the compensator with P' in the vicinity of C'. This involves measuring the residual function $F'(P) = 1 - |B'|^2$ [see Eq. (7.7a)] at several closely spaced polarizer angular readings P such that the polarizer transmission axis is nearly parallel to the compensator

fast axis. In the absence of errors, $F'(P)$ exhibits a minimum when $P' - C' = 0$. With this minimum identified, the polarizer is rotated to $P' = 45° + C'$ and spectra in (α', β') are collected. These are interpreted using the same equations as for rotating-analyzer ellipsometry in the reflection geometry [i.e., Eqs. (7.5) and (7.6) with (P_S, A, A_S) replaced by (A_S, P, P_S)], in which case (ψ, Δ) are given by $\psi = \psi_C = 45°$ (independent of energy) and $|\Delta(E)| = \delta(E)$.

In the second approach, the actual rotating-compensator multichannel ellipsometer is used, but in the straight-through $PC_r(\omega_C)A$ configuration. Initially, the transmission-axis angle of the analyzer is fixed, and an irradiance function calibration is performed with P' in the vicinity of A' and with the rotating compensator removed from the instrument. This involves measuring the dc irradiance level at the detector $I'_0(P)$ for several closely spaced polarizer angles P such that the transmission axes of the polarizer and analyzer are nearly parallel. In the absence of errors, $I'_0(P)$ exhibits a maximum when $P' - A' = 0$ (via Malus's Law). Next, the polarizer is fixed at a known value relative to the analyzer angle, i.e., a known value of $P' - A'$. Finally, the rotating compensator is added while maintaining the straight-through configuration. In this configuration, the $2\omega_C$ ac Fourier coefficients vanish, whereas the $4\omega_C$ ac Fourier coefficients yield the compensator retardance in accordance with the following expression:

$$\delta = 2 \arccos \sqrt{\frac{1 - |B'_4|}{1 + |B'_4| \cos 2(P' - A')}}, \tag{7.36a}$$

where $|B'_4| = \sqrt{\alpha'^2_4 + \beta'^2_4}$.

As an example of this second approach, the main part of Fig. 7.16 shows the retardance spectrum for a MgF_2 biplate compensator obtained in the straight-through $PC_r(\omega_C)A$ configuration using Eq. (7.36a). This type of compensator is preferred in rotating-compensator ellipsometers owing to its broad range of transmission from the near-infrared to the vacuum ultraviolet, as well as to its relative ease of mounting and alignment.[58] The solid line in Fig. 7.16 shows a best fit to the measured retardance using the expression

$$\delta(E) = 360d \frac{E}{1239.8} \sum_{k=1}^{5} c_k E^{k-1}.$$

This expression is derived assuming that the birefringence of MgF_2 $(n_S - n_F)$ is a fourth-order polynomial function of photon energy E. Here $(c_k; k = 1, \ldots, 5)$ are polynomial fitting coefficients, and d is now the thickness difference between the two MgF_2 plates. Thus, the units of δ, d, and E are degrees, nanometers, and eV, respectively. The thickness difference is

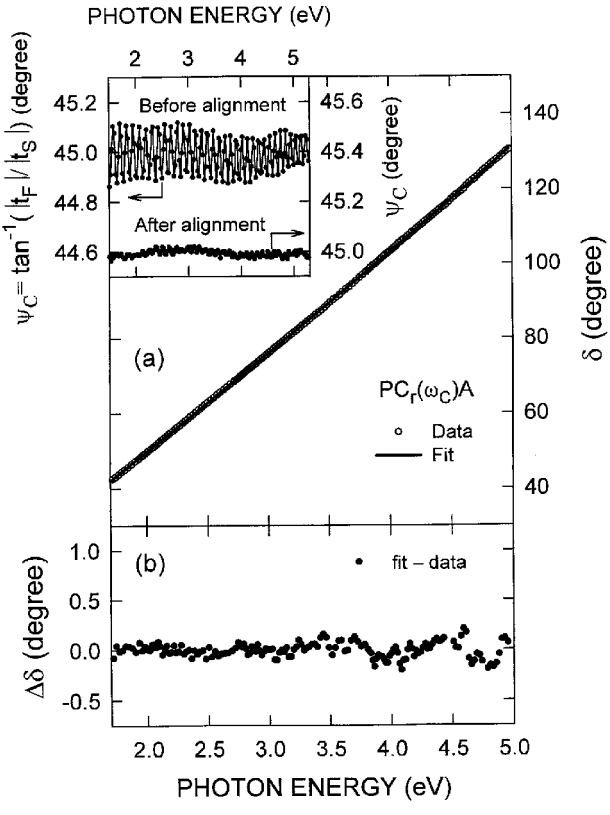


Figure 7.16 (a) Measured retardance spectrum δ (points) and its best fit (line) applying a polynomial in the birefringence versus photon energy for a MgF_2 biplate compensator obtained using a rotating-compensator multichannel ellipsometer in the straight-through $PC_r(\omega_C)A$ configuration. In (b), the difference between the experimental and best fit spectra in δ is shown. The inset in (a) shows the ellipsometric parameter $\psi_C = \tan^{-1}|t_F|/|t_S|$ of the compensator measured using a rotating-analyzer, single-channel ellipsometer in the straight-through $PCA_r(\omega_A)$ configuration, where t_F and t_S are the complex transmission coefficients of the compensator along the fast (F) and slow (S) axes. The two spectra in ψ_C in the inset were obtained before and after azimuthal alignment of the two plates using a specially designed alignment device. [Unpublished data from J. Lee, Pennsylvania State Univ. (2000).]

determined independently of the other parameters by assuming reference data for the birefringence of MgF_2, and fitting δ to extract d. In a second iteration, d is fixed at this best fit value and δ is fit using (c_k; $k = 1, \ldots, 5$) as free parameters. In Fig. 7.16, the best fit values obtained according to this fitting procedure include $d = 7161$ nm, $c_1 = 0.010605$ deg/(nm eV), $c_2 = 1.4944 \times 10^{-3}$ deg/(nm eV^2), and $c_3 = -6.2421 \times 10^{-4}$ deg/(nm eV^3),

$c_4 = 1.2879 \times 10^{-4}$ deg/(nm eV^4), and $c_5 = -9.5258 \times 10^{-6}$ deg/(nm eV^5), with a resulting quarterwave value of $\lambda_Q = 351.8$ nm ($E_Q = 3.524$ eV). The lower panel in Fig. 7.16 shows the difference between the data and the best fit spectra in δ. The systematic deviations of $\pm 0.1°$ observed here are found to be reproducible when measuring and fitting δ using both approaches described so far. This observation suggests that the deviations are true characteristics of the retardance. Thus, oscillatory terms in δ of unknown origin appear to be necessary in order to reduce the fitting errors below $\pm 0.1°$.

In the third approach for the calibration of δ, the $PSC_r(\omega_C)A$ ellipsometer in the actual reflection configuration is used, and the calibration angle P_S is required.[57] Thus, it is most convenient to perform this particular procedure after the offset and phase angle calibration that provides (P_S, C_S, A_S). Once P_S is determined (see, e.g., the procedure in the following subsection that employs $|B'_2|$, the amplitude of the $2\omega_C$ Fourier coefficients), then the polarizer is set at $P' = 0$ or $\pi/2$. This generates linear polarization upon reflection from an isotropic sample. In this experimental situation, one can write:

$$|B'_4|^{-1} = \gamma_0 + \gamma_2 \cos 2A + \eta_2 \sin 2A, \tag{7.36b}$$

where

$$\gamma_0 = [\sin^2(\delta/2)]^{-1}, \tag{7.36c}$$

$$\gamma_2 = \pm \cos 2A_S \cot^2(\delta/2), \tag{7.36d}$$

$$\eta_2 = \pm \sin 2A_S \cot^2(\delta/2). \tag{7.36e}$$

In these expressions, the coefficients of the Fourier series expansion in terms of the analyzer angular setting A are given for the values of 0 (upper sign) and $\pi/2$ (lower sign) for the true polarizer angle. Thus, the spectra in the coefficients (γ_0, γ_2, η_2) can be determined by stepping the analyzer over the range from 0 to π and Fourier analyzing the resulting spectra in $|B'_4|^{-1}$. The compensator retardance spectrum can be deduced from these coefficients according to:

$$\delta = 2 \arctan\{(\gamma_2^2 + \eta_2^2)^{-1/4}\}. \tag{7.36f}$$

The only disadvantage of this approach is that it requires stepping motors on both the polarizer and analyzer, whereas the general calibration approaches described below do not require a stepping motor on the analyzer.

Owing to the widespread use of MgF_2 biplate compensators in rotating-compensator ellipsometry, this part concludes with a brief discussion of a pre-calibration step required specifically for biplate compensators. This step is best illustrated with a recent example, as shown in the inset of Fig. 7.16, namely, internal alignment of the individual plates of the MgF_2 compensator in order to ensure that their fast and slow axes are perpendicular. This alignment is performed using a separate single-channel, rotating-analyzer ellipsometer in the straight-through $PCA_r(\omega_A)$ configuration, as in the first approach for the measurement of δ as described above. The parameter plotted in the inset is $\psi_C = \arctan |t_F|/|t_S|$, where t_F and t_S are the complex transmission coefficients of the compensator along the fast (F) and slow (S) axes. The two spectra in the inset were obtained before and after azimuthal alignment of the two plates using a specially designed alignment device. This device fixes the housing of one plate, while allowing the housing of the second plate to be rotated using a precision rotator. The spectra in both $\psi = \psi_C$ and $|\Delta| = \delta$ obtained in the $PCA_r(\omega_A)$ configuration before alignment show high frequency oscillations. These oscillations are clearest in ψ_C as evidenced in Fig. 7.16 owing to the absence of a measurable spectral variation in this parameter. The rotational adjustment dampens these oscillations significantly, and the spectrum in ψ_C obtained after alignment shows deviations of no more than $\pm 0.03°$ from its average value of 44.98°. This result indicates that the compensator of Fig. 7.16 is essentially ideal, exhibiting no significant dichroic behavior.

Offset and Phase Angle Calibration

Strongly absorbing surfaces ($30° < |\Delta| < 150°$)

In order to determine P_S for the rotating-compensator multichannel ellipsometer, the following second-order expansions can be derived for P' in the vicinity of 0 and $\pi/2$ (assuming $p = 1$):

$$G'_2(P) \equiv I'^2_0 |B'_2|^2 \tag{7.37a}$$

$$\propto a_0^2 R^2(P') |B_2|^2 \tag{7.37b}$$

$$\propto (2|r_p||r_s|\sin\delta \sin\Delta)^2 (P - P_S)^2; \quad P \approx P_S, \tag{7.37c}$$

$$\propto (2|r_p||r_s|\sin\delta \sin\Delta)^2 \{P - [P_S + (\pi/2)]\}^2; \quad P \approx P_S + (\pi/2). \tag{7.37d}$$

Here, $B'_2 = \alpha'_2 + i\beta'_2$, and r_p and r_s are the complex amplitude reflection coefficients of the reflecting surface. Thus, in order to determine P_S,

experimental values for $G'_2(P)$ are plotted versus the polarizer setting P at closely spaced values encompassing either P_S [Eq. (7.37c)] or $P_S + (\pi/2)$ [Eq. (7.37d)]. The resulting data are fit to a parabola, and the value of P at which the minimum in the parabola occurs is P_S or $P_S + (\pi/2)$. The sensitivity of these procedures is greatest for $\delta = 90°$, and also for $\Delta = 90°$. All procedures based on Eqs. (7.37c) and (7.37d) fail when $\Delta = 0°$ or $\pm 180°$, as in the case of transparent substrates, and an alternative procedure will be described later for such situations.

With P_S known, the next calibration step is to determine both A_S and C_S, where $-C_S$ is the angle of the compensator fast axis at the onset of data acquisition (i.e., at the onset of the S_1 integration for the given pixel). In this calibration step, the following relationships can be used:

$$\Theta'_2 = [\arctan(\beta'_2/\alpha'_2)]/2 = C_S + (A - A_S) \pm (\pi/4), \tag{7.38}$$

$$\Theta'_4(P) = [\arctan(\beta'_4/\alpha'_4)]/2$$

$$\approx 2C_S + (A - A_S) + \cot\psi \cos\Delta\,(P - P_S); P \approx P_S, \tag{7.39a}$$

$$\approx 2C_S + (A - A_S) + \tan\psi \cos\Delta\,\{P - [P_S + (\pi/2)]\};$$
$$P \approx P_S + (\pi/2), \tag{7.39b}$$

where the expansions in Eqs. (7.39a) and (7.39b) are valid to second order in P in the vicinity of $P' = P - P_S = 0$ and $P' = \pi/2$, respectively. The $\pm(\pi/4)$ term at the right in Eq. (7.38) is equivalent to an uncertainty in distinguishing the fast and slow axes of the compensator. In Eqs. (7.38) and (7.39), if the computed range of the inverse tangent is assumed to be $-\pi/2 < 2\Theta'_n \leq \pi/2$, then additional quadrant corrections may be required depending on the required sign choice in Eq. (7.38) and on the quadrants of $2(C_S + A')$ and $2(2C_S + A' + Q)$.

In order to perform the calibration, Θ'_2 and Θ'_4 are obtained versus P in the vicinity of P_S [or $P_S + (\pi/2)$]. The resulting data for Θ'_2 (which are independent of P) are averaged to improve the precision, and the resulting data for Θ'_4 are fit by a linear function of P which is evaluated at P_S [or $P_S + (\pi/2)$] to yield $2C_S + (A - A_S)$. Thus, C_S and $A - A_S$ both can be deduced from Eqs. (7.38) and (7.39a), neglecting the possible quadrant corrections:

$$C_S = \Theta'_4(P_S) - \Theta'_2 \pm (\pi/4), \tag{7.40a}$$

$$A - A_S = 2\Theta'_2 - [\Theta'_4(P_S) \pm (\pi/2)]. \tag{7.40b}$$

Either the two top (+) or two bottom (−) sign choices are made to preserve the sign uncertainty in Eq. (7.38). The correct sign choices in Eqs. (7.38)

and (7.40) along with the appropriate quadrant corrections are difficult to resolve theoretically. As a result, it is important to pre-align the fast axis of the compensator with respect to the motor shaft to ensure that $C' = \omega_C t - C_S$ is near zero at the onset of data collection ($t = 0$) for a selected reference pixel (or pixel group). In this way, approximate values of C_S and $A - A_S$ are known in advance of the calibration. An additional check is to ensure that for a bulk, optically absorbing substrate material the correct sign of Δ is obtained [$\Delta > 0°$ within the optics ($e^{i\omega t}$) convention (used here and in Ref. 8) and $\Delta < 0°$ within the physics ($e^{-i\omega t}$) convention (used, for example, in Refs. 4, 31, and 50)]. Equations (7.40) are also valid if Θ'_2 and Θ'_4 are measured and analyzed in the vicinity of $P_S + (\pi/2)$. A calibration based on Eq. (7.39a) works best for surfaces with $\Delta = 90°$ and $\psi = 90°$. For surfaces in which $\psi < 45°$, Eq. (7.39b) gives improved results. It should be noted that problems develop in Eqs. (7.39) for Δ near 0° and ±180° only because of the inability to deduce an accurate value of P_S [or $P_S + (\pi/2)$]. If an accurate value of P_S is available, then Eqs. (7.39) will still provide satisfactory results for Θ'_4. Since both $2\omega_C$ Fourier coefficients vanish when $\Delta = 0°$ or ±180° (for which $\sin 2\chi = 0$), an alternative method to Eqs. (7.38) and (7.40) must be used under these conditions to decouple C_S and $A - A_S$ based on the $4\omega_C$ coefficients alone. Such a method will be described below in conjunction with that used to determine P_S when Δ is near 0° or ±180°.

Figure 7.17(a) shows the results of the calibration of a rotating-compensator multichannel ellipsometer in which P_S is extracted based on Eq. (7.37c), using data from a Cr sample over the photon energy range from 1.8 to 3.1 eV. No chromatic errors in P_S are detected over this range, and the standard deviation is ~0.005°. Figures 7.17(b), (c), and (d) show the associated calibrations for $A - A_S$ and C_S based on Eqs. (7.40), for the same Cr sample in the same calibration procedure. In this case, the correct value of A_S can be identified to within ~±0.01°, in spite of fluctuations on the order of ±0.05°. In addition, C_S is found to be a linear function of pixel group number, and the relevant calibration information for the data of Fig. 7.17(c) is the equation that describes this relationship, namely, $C_S = -15.31° + 0.2054°\,(k-1)$. For this calibration and for the subsequent real time measurement, the 1024 pixel linear photodiode array detector was grouped uniformly by 8, and k represents the pixel group number which varies from 1 to 128 along the length of the array. The slope of the calibration equation for C_S represents the angle through which the compensator rotates in the time required to read out one pixel group. The observed slope is the consequence of a pixel group readout time of 35 μs and a compensator motor frequency of 16.30 Hz. Finally, Fig. 7.17(d) shows the deviations in C_S from the best fit linear relationship as a function of the

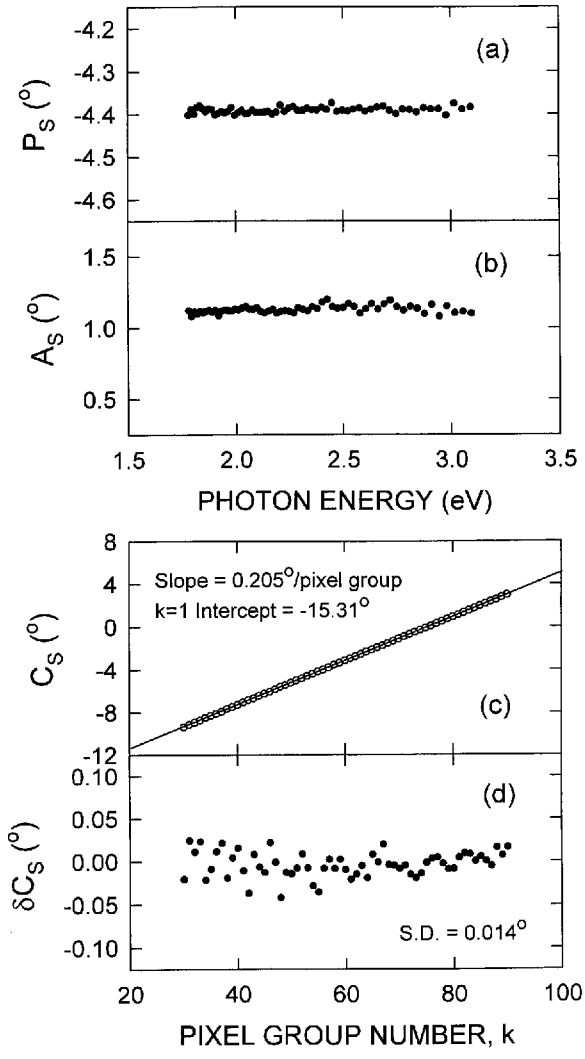


Figure 7.17 Calibration results for a rotating-compensator multichannel ellipsometer obtained in the PSC$_r(\omega_C)$A configuration in reflection from an opaque Cr thin film: (a) the polarizer angular offset P_S plotted versus photon energy from Eq. (7.37c); (b) the analyzer angular offset A_S plotted versus photon energy from Eq. (7.40b); (c) the rotating compensator phase angle C_S plotted versus pixel group number k from Eq. (7.40a); and (d) the deviation of C_S from the best fitting linear relationship versus pixel group number. In (d), the value given is the standard deviation of C_S from the linear relationship. [Adapted from Ref. 17.]

pixel group number. These deviations may result from motor instability over the time scale of the detector readout (4.5 ms) and also from chromatic errors in the optical system.

Weakly absorbing surfaces ($|\Delta| < 30°$; $|\Delta| > 150°$)

The calibration procedures of the previous paragraphs are dependent on having a relatively strongly absorbing surface for which $30° < \Delta < 150°$. Next, alternative procedures are outlined briefly that are applicable for weakly absorbing surfaces. In particular, such alternative procedures are a necessity for calibrations using glass, fused silica, or other dielectric surfaces.

By defining the following zone-difference phase function to second order in P':

$$\begin{aligned}\Phi'_4(P) &= \Theta'_4(A = \pi/2, P) - \Theta'_4[A = 0, P + (\pi/2)] \\ &\approx [2 \cot 2\psi \cos\Delta\,(P - P_S)]; \quad P \approx P_S,\end{aligned} \tag{7.41a}$$

it is evident that the zero-crossing of this function occurs at $P = P_S$. A quadrant correction of $\pm\pi/2$ may need to be incorporated, however, in order to apply Eq. (7.41a). Thus, an alternative calibration involves: (i) measuring pairs of spectra in $\Theta'_4 = (1/2)\arctan(\beta'_4/\alpha'_4)$ at several closely-spaced values of P near P_S with $A = \pi/2$ and near $P_S + (\pi/2)$ with $A = 0$; (ii) subtracting the two pairs of spectra according to Eq. (7.41a); (iii) fitting the results at each pixel (or pixel group) to a linear function of P; and (iv) identifying the zero-crossing of the best fit linear function as P_S. This method has the highest sensitivity for Δ near 0° and ±180° and for ψ near 0° and 90°. It fails when $\Delta = 90°$ or when $\psi = 45°$. From the resulting value of P_S, A_S can be deduced from:

$$\cos 2(A - A_S) = \frac{|B'_4(P_S)| - |B'_4(P_S + \pi/2)|}{2|B'_4(P_S)||B'_4(P_S + \pi/2)| - |B'_4(P_S + \pi/2)| - |B'_4(P_S)|}, \tag{7.41b}$$

where $|B'_4| = \sqrt{\alpha'^2_4 + \beta'^2_4}$. In this second procedure, $|B'_4|$ is determined at several closely spaced values of P near $P = P_S$ and $P = P_S + (\pi/2)$ with A near $A = A_S + (\pi/4)$ for optimum sensitivity in evaluating the inverse cosine. These data are fit to polynomial functions of P to improve the precision of the calibration. The polynomial functions are evaluated at $P = P_S$ and $P = P_S + (\pi/2)$ for substitution into Eq. (7.41b) to determine $A' = A - A_S$. The spectrum in C_S can then be extracted from Eqs. (7.39a) or (7.39b).

Alternative Approach for Analyzer Offset Angle

The calibration procedures of the previous discussion are performed by sweeping the polarizer setting P using a stepping motor on the polarizer rotator. Although a stepping motor on the analyzer rotator is also desirable [e.g., for the compensator retardance calibration in the $PSC_r(\omega_C)A$ reflection measurement geometry], it is not an absolute necessity for the success of the offset and phase angle calibration procedures. Experience has shown, however, that the analyzer offset angle A_S exhibits the largest experimental errors because the analyzer remains essentially fixed in all these calibrations.[17,56] In fact, a more accurate approach for A_S requires a stepping motor on the analyzer rotator and is based on the dc Fourier coefficient, given by Eqs. (7.15b) and (7.16). This procedure first requires knowledge of P_S, which is obtained using one of the two procedures described above depending on the value of Δ. Once P_S is known, then P is set equal to P_S or $P_S + (\pi/2)$ so that $P' = 0$ or $\pi/2$. When $P' = 0$ or $\pi/2$ and when A is in the neighborhood of A_S, then one arrives at the following two expressions

$$I'_0(A) \propto |r_p|^2\{[1 + \cos^2(\delta/2)] - [2\cos^2(\delta/2)\,(A - A_S)^2]\}; P' = 0; \quad A \approx A_S, \tag{7.41c}$$

$$\propto |r_s|^2\{[\sin^2(\delta/2)] + [2\cos^2(\delta/2)(A - A_S)^2]\}; \quad P' = \pi/2; \quad A \approx A_S. \tag{7.41d}$$

Thus, calibration is performed by ensuring that P' is 0 or $\pi/2$, and measuring the I'_0 spectra at several closely spaced values of A near A_S. I'_0 is fit to a parabolic function of A for each pixel, and A_S is the value of A at which the extremum in the fitted function occurs. Because the irradiance is used in this calibration, its sensitivity does not depend on the value of Δ. Equation (7.41c) provides higher sensitivity when $|r_p| > |r_s|$, i.e., when $\psi > 45°$; Eq. (7.41d) is used when $\psi < 45°$. Alternative expressions to those of Eqs. (7.41c) and (7.41d), but for $A \approx A_S + (\pi/2)$, are given as:

$$I'_0(A) \propto |r_p|^2\{[\sin^2(\delta/2)] + 2\cos^2(\delta/2)\{A - [A_S + (\pi/2)]\}^2\}; \quad P' = 0; \quad A \approx A_S + (\pi/2), \tag{7.41e}$$

$$\propto |r_s|^2\{[1 + \cos^2(\delta/2)] - 2\cos^2(\delta/2)\{A - [A_S + (\pi/2)]\}^2\}; \quad P' = \pi/2; \quad A \approx A_S + (\pi/2). \tag{7.41f}$$

Self-Calibration

The term "self-calibration" implies that the calibration data are extracted from the actual real time sample measurements in conjunction with the beam or sample parameters. This is in contrast to conventional calibration, which is performed as one or more separate steps prior to the real time sample measurements. Any multichannel instrument configuration provides the possibility of self-calibration when the number of non-zero Fourier coefficients exceeds the number of final sample parameters extracted in the data reduction. For the single rotating-compensator ellipsometer, for example, a total of five dc and ac Fourier coefficients are deduced, from which four beam parameters are extracted in accordance with Eqs. (7.27) through (7.30). Thus, one calibration parameter can be determined from the Fourier coefficients. The clear choice for self-calibration is C_S, which can be determined according to

$$C_S = (1/2)\{\arctan(\beta'_2/\alpha'_2) - [2(A - A_S) \pm (\pi/2)]\}. \qquad (7.41g)$$

This equation is useful as long as Δ is not near 0° or ±180° [so that $|B'_2| = \sqrt{\alpha'^2_2 + \beta'^2_2}$ is not near zero]. Equation (7.41g) is derived from Eq. (7.38), and the sign and quadrant uncertainties arise similarly. In fact, the ± in Eq. (7.41g) represents a fast-slow axis uncertainty for the compensator, which is most readily resolved either by the fast axis pre-alignment procedure noted earlier or by the measurement of a strongly absorbing substrate material such that $0 \leq \Delta \leq 180°$. With the self-calibration of Eq. (7.41g) for appropriate thin film deposition or materials processing, the long term stability in the compensator rotation frequency can be tracked through the changes in the slope of the relationship between C_S and the pixel group number k.

Spectral Response Calibration

As described in Section 7.3.1.4 on the calibration of the rotating-polarizer multichannel ellipsometer, the critical step in the measurement of the beam parameter I_r/I_i or the sample parameter $R(P')$ is the determination of the ellipsometer spectral response function I_{00}. As in the case of the rotating-polarizer multichannel ellipsometer, a non-depolarizing substrate or starting surface prior to film deposition or materials processing can be used as a calibration standard. Alternatively, a non-depolarizing test sample such as a crystalline Si wafer can be placed so as to partially eclipse the sample of interest, and the sample holder can be translated so that the incident beam can probe first the calibration sample and then the substrate of interest used for

the real time measurements. In either case, the relevant equation in the determination of I_{00} for the rotating-compensator multichannel ellipsometer is:

$$I_{00} = \frac{I'_{0,\mathrm{cal}}}{\{|r_{p,\mathrm{cal}}|^2 \cos^2 P' + |r_{s,\mathrm{cal}}|^2 \sin^2 P'\}\{1 + \cos^2(\delta/2)\cos 2\chi_{\mathrm{cal}} \cos 2(A' - Q_{\mathrm{cal}})\}} \tag{7.42}$$

Here again, the subscript 'cal' is used to denote data characteristic of the calibration sample, either the substrate or starting surface, if it is sufficiently well known, or otherwise an adjacent Si wafer.

7.3.3 Dual Rotating Compensator

The single rotating-compensator multichannel ellipsometer of Section 7.3.2 provides the reflected beam Stokes vector for a given incident linear polarization state, defined by the angle P' of the transmission axis of the fixed polarizer. This instrument is well-suited for the study of either strongly or weakly absorbing isotropic samples with certain simple types of heterogeneities. With this instrument, sample anisotropy or more complicated heterogeneities can be probed by performing measurements at different angles P' of the polarizer. Such a measurement strategy, however, is not suitable for real time studies of materials preparation. For real time studies of anisotropic samples or samples with complicated heterogeneities, the ultimate solution is to measure spectra in the complete Mueller matrix of the sample at high speeds (<1 s acquisition time). This can be done by modulating the Stokes vector of the beam incident on the sample using a fixed polarizer followed by a rotating compensator, and detecting the reflected beam Stokes vector using a rotating compensator followed by a fixed analyzer. Only specific frequency ratios of the two rotating compensators provide the full Mueller matrix, e.g., frequency ratios 5:n where n = 1, …, 4. Figure 7.8(c) shows such an instrument in the $\mathrm{PC}_{1\mathrm{r}}(\omega_{1C})\mathrm{SC}_{2\mathrm{r}}(\omega_{2C})\mathrm{A}$ configuration in which $\omega_{1C} = 5\omega_B$ and $\omega_{2C} = 3\omega_B$, where ω_B is the base mechanical frequency and $2\omega_B$ is the fundamental optical frequency. The rotation frequencies for the first and second compensators are selected for practical reasons. Specifically, ω_{1C} and ω_{2C} are maintained above the minimum rotation frequency possible for stable dc motor operation and below the maximum frequency possible for successful photodiode array scanning.

7.3.3.1 Theoretical Description

For the dual rotating-compensator multichannel ellipsometer with $\omega_{1C} = 5\omega_B$ and $\omega_{2C} = 3\omega_B$, the time-dependent waveform predicted at each

pixel of the photodiode array detector can be expressed in terms of the Mueller matrix of the sample.[59] The waveform is given by

$$I(t) = i_0\left\{a_0 + \sum_{n=1}^{16}[a_{2n}\cos(2n\omega_B t - \phi_{2n}) + b_{2n}\sin(2n\omega_B t - \phi_{2n})]\right\} \quad (7.43a)$$

$$= I_0\left\{1 + \sum_{n=1}^{16}[\alpha_{2n}\cos(2n\omega_B t - \phi_{2n}) + \beta_{2n}\sin(2n\omega_B t - \phi_{2n})]\right\}, \quad (7.43b)$$

where

$$I_0 = I_{00}\,a_0\,M_{11} = I_{00}a_0\,R_u = (I_{00}\,a_0/2)(|r_p|^2 + |r_s|^2). \quad (7.44)$$

In Eq. (7.43a), $\{a_0, (a_{2n}, b_{2n}); n = 1, \ldots, 16\}$ are the dc and unnormalized ac Fourier coefficients . In addition, $\{\phi_{2n}; n = 1, \ldots, 16\}$ are the phases of the individual Fourier components given in terms of the phases C_{S1} and C_{S2} of the two rotating compensators in the second column of Table 7-1. These latter phases are defined by $C'_1 = 5(\omega_B t - C_{S1})$ and $C'_2 = 3(\omega_B t - C_{S2})$, where C'_1

Table 7-1. A listing of the non-zero, normalized Fourier coefficients and calibration phase angles of Eq. (7.43a) for the dual rotating-compensator multichannel ellipsometer, where $c_j = \cos^2(\delta_j/2)$ and $s_j = \sin^2(\delta_j/2)$ $(j = 1,2)$.

Freq. term (dc, cos2*n*C, or sin2*n*C)	**Phase (ϕ_{2n})**	**Fourier coefficient of ($2nC - \phi_{2n}$) (α_{2n} or β_{2n})**
dc	—	$1 + m_{12}c_1\cos2P' + m_{13}c_1\sin2P' + m_{21}c_2\cos2A' + m_{31}c_2\sin2A' + m_{22}c_1c_2\cos2P'\cos2A' + m_{23}c_1c_2\sin2P'\cos2A' + m_{32}c_1c_2\cos2P'\sin2A' + m_{33}c_1c_2\sin2P'\sin2A'$
cos2*C*	$12C_{S2} - 10C_{S1}$	$-0.5\,m_{24}\sin\delta_1\,s_2\sin2(P' - A') + 0.5\,m_{34}\sin\delta_1\,s_2\cos2(P' - A')$
sin2*C*	$12C_{S2} - 10C_{S1}$	$-0.5\,m_{24}\sin\delta_1\,s_2\cos2(P' - A') - 0.5\,m_{34}\sin\delta_1\,s_2\sin2(P' - A')$
cos4*C*	$10C_{S1} - 6C_{S2}$	$-0.5\,m_{44}\sin\delta_1\sin\delta_2\cos2(P' - A')$
sin4*C*	$10C_{S1} - 6C_{S2}$	$-0.5\,m_{44}\sin\delta_1\sin\delta_2\sin2(P' - A')$
cos6*C*	$6C_{S2}$	$m_{41}\sin\delta_2\sin2A' + m_{42}c_1\sin\delta_2\cos2P'\sin2A' + m_{43}c_1\sin\delta_2\sin2P'\sin2A'$

(*cont.*)

Table 7-1. (*cont.*)

Freq. term (dc, cos2*n*C, or sin2*n*C)	Phase (ϕ_{2n})	Fourier coefficient of ($2nC - \phi_{2n}$) (α_{2n} or β_{2n})
sin6C	$6C_{S2}$	$-m_{41}\sin\delta_2\cos2A' - m_{42}c_1\sin\delta_2\cos2P'\cos2A' - m_{43}c_1\sin\delta_2\sin2P'\cos2A'$
cos8C	$20C_{S1}-12C_{S2}$	$0.5m_{22}s_1s_2\cos2(P'-A') - 0.5m_{23}s_1s_2\sin2(P'-A') + 0.5m_{32}s_1s_2\sin2(P'-A') + 0.5m_{33}s_1s_2\cos2(P'-A')$
sin8C	$20C_{S1}-12C_{S2}$	$0.5m_{22}s_1s_2\sin2(P'-A') + 0.5m_{23}s_1s_2\cos2(P'-A') - 0.5m_{32}s_1s_2\cos2(P'-A') + 0.5m_{33}s_1s_2\sin2(P'-A')$
cos10C	$10C_{S1}$	$-m_{14}\sin\delta_1\sin2P' - m_{24}\sin\delta_1 c_2\sin2P'\cos2A' - m_{34}\sin\delta_1 c_2\sin2P'\sin2A'$
sin10C	$10C_{S1}$	$m_{14}\sin\delta_1\cos2P' + m_{24}\sin\delta_1 c_2\cos2P'\cos2A' + m_{34}\sin\delta_1 c_2\cos2P'\sin2A'$
cos12C	$12C_{S2}$	$m_{21}s_2\cos2A' - m_{31}s_2\sin2A' + m_{22}c_1s_2\cos2P'\cos2A' + m_{23}c_1s_2\sin2P'\cos2A' - m_{32}c_1s_2\cos2P'\sin2A' - m_{33}c_1s_2\sin2P'\sin2A'$
sin12C	$12C_{S2}$	$m_{21}s_2\sin2A' + m_{31}s_2\cos2A' + m_{22}c_1s_2\cos2P'\sin2A' + m_{23}c_1s_2\sin2P'\sin2A' + m_{32}c_1s_2\cos2P'\cos2A' + m_{33}c_1s_2\sin2P'\cos2A'$
cos14C	$20C_{S1}-6C_{S2}$	$-0.5\,m_{42}\,s_1\sin\delta_2\sin2(P'-A') - 0.5\,m_{43}\,s_1\sin\delta_2\cos2(P'-A')$
sin14C	$20C_{S1}-6C_{S2}$	$0.5\,m_{42}\,s_1\sin\delta_2\cos2(P'-A') - 0.5\,m_{43}\,s_1\sin\delta_2\sin2(P'-A')$
cos16C	$10C_{S1}+6C_{S2}$	$0.5\,m_{44}\sin\delta_1\sin\delta_2\cos2(P'+A')$
sin16C	$10C_{S1}+6C_{S2}$	$0.5\,m_{44}\sin\delta_1\sin\delta_2\sin2(P'+A')$
cos20C	$20C_{S1}$	$m_{12}s_1\cos2P' - m_{13}s_1\sin2P' + m_{22}s_1c_2\cos2P'\cos2A' - m_{23}s_1c_2\sin2P'\cos2A' + m_{32}s_1c_2\cos2P'\sin2A' - m_{33}s_1c_2\sin2P'\sin2A'$
sin20C	$20C_{S1}$	$m_{12}s_1\sin2P' + m_{13}s_1\cos2P' + m_{22}s_1c_2\sin2P'\cos2A' + m_{23}s_1c_2\cos2P'\cos2A' + m_{32}s_1c_2\sin2P'\sin2A' + m_{33}s_1c_2\cos2P'\sin2A'$
cos22C	$10C_{S1}+12C_{S2}$	$-0.5\,m_{24}\sin\delta_1 s_2\sin2(P'+A') - 0.5\,m_{34}\sin\delta_1 s_2\cos2(P'+A')$

Table 7-1. (*cont.*)

Freq. term (dc, cos2*n*C, or sin2*n*C)	Phase (ϕ_{2n})	Fourier coefficient of ($2nC - \phi_{2n}$) (α_{2n} or β_{2n})
sin22*C*	$10C_{S1}+12C_{S2}$	$0.5\, m_{24} \sin\delta_1\, s_2 \cos2(P' + A')$ $- 0.5\, m_{34} \sin\delta_1\, s_2 \sin2(P' + A')$
cos26*C*	$20C_{S1}+6C_{S2}$	$0.5\, m_{42}\, s_1 \sin\delta_2 \sin2(P' + A')$ $+ 0.5\, m_{43}\, s_1 \sin\delta_2 \cos2(P' + A')$
sin26*C*	$20C_{S1}+6C_{S2}$	$-0.5\, m_{42}\, s_1 \sin\delta_2 \cos2(P' + A')$ $+ 0.5\, m_{43}\, s_1 \sin\delta_2 \sin2(P' + A')$
cos32*C*	$20C_{S1}+12C_{S2}$	$0.5m_{22}s_1s_2 \cos2(P' + A') - 0.5m_{23}s_1s_2 \sin2(P' + A')$ $- 0.5m_{32}s_1s_2 \sin2(P' + A')$ $- 0.5m_{33}s_1s_2 \cos2(P' + A')$
sin32*C*	$20C_{S1}+12C_{S2}$	$0.5m_{22}s_1s_2 \sin2(P' + A') + 0.5m_{23}s_1s_2 \cos2(P' + A')$ $+ 0.5m_{32}s_1s_2 \cos2(P' + A')$ $- 0.5m_{33}s_1s_2 \sin2(P' + A')$

and C_2' are the true angles of the fast axes of the first and second compensators, measured from the plane of incidence in a counterclockwise-positive sense when looking opposite to the direction of the light beam. Thus, $-5C_{S1}$ and $-3C_{S2}$ are the angles of the fast axes at time zero, defined (as in the other instruments) as the onset of data acquisition for the given pixel (or pixel group). In Eq. (7.43b), $\{I_0, (\alpha_{2n}, \beta_{2n}); n = 1, \ldots, 16\}$ are the dc and normalized ac coefficients, given by $I_0 = a_0 i_0$, $\alpha_{2n} = a_{2n}/a_0$, and $\beta_{2n} = b_{2n}/a_0$, and presented in the third column of Table 7-1. Among the 33 possible Fourier coefficients in Eqs. (7.43), eight ac coefficients vanish even for the most general form of the Mueller matrix. The vanishing coefficients are those corresponding to $n = 9$, 12, 14, and 15 in the sums of Eqs. (7.43).

In Eq. (7.44), the average irradiance in the theoretical waveform is expressed as the product of three factors: (i) the ellipsometer spectral response function I_{00}, (ii) the dc Fourier coefficient a_0, and (iii) the (1,1) Mueller matrix element of the sample M_{11}. In fact in the absence of errors, I_{00} is independent of the sample and the optical element settings, whereas M_{11} depends only on the sample. As can be noted in Table 7-1, a_0 is a complicated function that depends on the sample (i.e., the normalized Mueller matrix elements $m_{ij} = M_{ij}/M_{11}$), the true polarizer and analyzer angles P' and A', and the compensator retardance spectra δ_1 and δ_2. The second equality in Eq. (7.44) results from the fact that for a non-depolarizing specular sample the (1,1) Mueller matrix element is simply equal to the reflectance for

unpolarized light [see, e.g., Eq. (7.19a)]. The third equality in Eq. (7.44) is only valid for non-depolarizing and isotropic samples.

7.3.3.2 Data Collection

For an error-free system in the dual-rotating compensator configuration with $\omega_{1C} = 5\omega_B$ and $\omega_{2C} = 3\omega_B$, the irradiance at any given pixel (or pixel group) obeys the following experimental expression:[59]

$$I'(t) = I'_0\left\{1 + \sum_{n=1}^{16}[\alpha'_{2n}\cos 2n\omega_B t + \beta'_{2n}\sin 2n\omega_B t]\right\}, \tag{7.45}$$

where $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); \quad n = 1, 2, \ldots, 16\}$ are the dc and normalized ac Fourier coefficients to be determined experimentally. Equation (7.45) differs from the theoretical expression Eq. (7.43b) in that the waveform does not include the phase correction angles ϕ_{2n} defined in Table 7-1. These terms are omitted from the experimental waveform because they cannot be determined until a calibration is performed that yields C_{S1} and C_{S2}. As in Eq. (7.43b), $\{(\alpha'_{2n}, \beta'_{2n}); n = 9, 12, 14, 15\}$ all vanish, even for the most general sample.

Data collection for the dual-rotating compensator is performed in a manner similar to the simpler multichannel instruments. If the detector is read N times per fundamental optical cycle (π/ω_B), leading to N equal exposure times $t_e = \pi/N\omega_B$, then N spectra are generated having the following form:[59]

$$S_j = I'_0\int_{(j-1)\pi/N\omega_B}^{j\pi/N\omega_B}\left\{1 + \sum_{n=1}^{16}[\alpha'_{2n}\cos 2n\omega_B t + \beta'_{2n}\sin 2n\omega_B t]\right\}dt; \quad (j = 1, \ldots, N), \tag{7.46a}$$

$$= \frac{\pi I'_0}{N\omega_B} + \sum_{n=1}^{16}\frac{I'_0}{n\omega_B}\left[\sin\left(\frac{n\pi}{N}\right)\right] \times\left[\alpha'_{2n}\cos\left(\frac{(2j-1)n\pi}{N}\right) + \beta'_{2n}\sin\left(\frac{(2j-1)n\pi}{N}\right)\right]. \tag{7.46b}$$

For each pixel (or pixel group) of the photodiode array, Eq. (7.46b) represents a system of N equations in 25 unknowns $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1, 2, \ldots, 8, 10, 11, 13, 16\}$. In the first implementation of such an instrument, $N = 36$ waveform integrations were obtained over the fundamental optical cycle.[18] For this choice, one can generate a set of 36 equations in 36 unknowns by extending the series in Eq. (7.46b) to include all even Fourier coefficients up to β'_{36} (but not including α'_{36}). The corresponding 36×36 matrix of coefficients can be

inverted to deduce $\{I'_0, \alpha'_{2m}(m = 1, \ldots, 17), \beta'_{2n}(n = 1, \ldots, 18)\}$. Of these 36 quantities, 11 Fourier coefficients should vanish (corresponding to $m, n = 9, 12, 14, 15, 17$; $n = 18$). Thus, for $N = 36$, there are 11 consistency checks in the inversion of the waveform integrals to deduce the Fourier coefficients.

7.3.3.3 Data Reduction

Once the dc and normalized ac Fourier coefficients $\{I'_0, (\alpha'_{2n}, \beta'_{2n}); n = 1, \ldots, 8, 10, 11, 13, 16\}$ are determined from inversion of Eq. (7.46b), the next step is a phase correction of the coefficients. Phase correction is performed as usual using

$$I_0 = I'_0 \tag{7.47a}$$

$$(\alpha_{2n} \quad \beta_{2n})^{\mathrm{T}} = \Re(\phi_{2n})(\alpha'_{2n} \quad \beta'_{2n})^{\mathrm{T}}; \quad (n = 1, 2, \ldots, 8, 10, 11, 13, 16). \tag{7.47b}$$

Table 7-1 provides expressions for the phase angles ϕ_{2n} of Eq. (7.47b). Alternative more general expressions associated with the 5:3 frequency ratio include: $\phi_{2n} = \{\mathrm{sgn}(3L - 5K)\}\ \{(3LC_{S2} - 5KC_{S1})\}$, $(K = 0, 2, 4; L = 0, 2, 4)$, where $2n = |3L - 5K|$, valid for $n = 1, \ldots, 7, 10$; and $\phi_{2n} = (3LC_{S2} + 5KC_{S1})$ $(K = 2, 4; L = 2, 4)$, where $2n = 3L + 5K$, valid for $n = 8, 11, 13, 16$. As noted previously, C_{S1} and C_{S2} are the absolute phase angles for the two rotating compensators, defined in analogy with C_S for the $\mathrm{PSC_r}(\omega_C)\mathrm{A}$ multichannel ellipsometer such that $-5C_{S1}$ and $-3C_{S2}$ are the angular positions of the fast axes of the first and second compensators at time zero of data acquisition for a given pixel.

From the phase-corrected Fourier coefficients, the (1,1) element of the unnormalized Mueller matrix M_{11} of the sample, as well as the normalized Mueller matrix elements $\{m_{ij} = M_{ij}/M_{11}; i = 1,\ldots,4; j = 1,\ldots,4\}$ can be determined according to the following equations:[59,60]

$$M_{11} = I_0/(I_{00}a_0) = R_u = (1/2)(|r_p|^2 + |r_s|^2), \tag{7.48}$$

$$\begin{aligned} a_0 = t_1t_2\{&t_1t_2 + \alpha_8 \cos 4(P' - A') + \beta_8 \sin 4(P' - A') - t_1\alpha_{12} \cos 4A' \\ &- t_1\beta_{12} \sin 4A' - t_2\alpha_{20} \cos 4P' - t_2\beta_{20} \sin 4P' + \alpha_{32} \cos 4(P' + A') \\ &+ \beta_{32} \sin 4(P' + A')\}^{-1}, \end{aligned} \tag{7.49}$$

$$m_{12} + im_{13} = (a_0/s_1t_2)\exp(-2iP')\{-B_8 \exp(4iA') + t_2B_{20} - B_{32}\exp(-4iA')\}, \tag{7.50}$$

$$\begin{aligned} m_{14} = (a_0/t_2 \sin\delta_1)\{&2\alpha_2 \sin 2(P' - 2A') + 2\beta_2 \cos 2(P' - 2A') - t_2\alpha_{10} \sin 2P' \\ &+ t_2\beta_{10} \cos 2P'\}, \end{aligned} \tag{7.51}$$

$$m_{21} + im_{31} = (a_0/t_1 s_2)\exp(-2iA')\{-B_8^* \exp(4iP') + t_1 B_{12} - B_{32}\exp(-4iP')\}, \quad (7.52)$$

$$m_{22} + im_{23} = (a_0/s_1 s_2)\exp(-2iP')\{B_8 \exp(2iA') + B_{32}\exp(-2iA')\}, \quad (7.53)$$

$$m_{24} + im_{34} = (2ia_0/s_2 \sin\delta_1)\{B_2 \exp 2i(P' - A')\}, \quad (7.54)$$

$$m_{32} + im_{33} = (ia_0/s_1 s_2)\exp(-2iP')\{B_8 \exp(2iA') - B_{32} \exp(-2iA')\}, \quad (7.55)$$

$$m_{41} = (a_0/t_1 \sin\delta_2)\{2\alpha_{14}\sin 2(2P' - A') - 2\beta_{14}\cos 2(2P' - A') + t_1\alpha_6 \sin 2A' - t_1\beta_6 \cos 2A'\}, \quad (7.56)$$

$$m_{42} + im_{43} = (-2ia_0/s_1 \sin\delta_2)\{B_{14} \exp 2i(A' - P')\}, \quad (7.57)$$

$$m_{44} = (2a_0/\sin\delta_1 \sin\delta_2)\{-\alpha_4 \cos 2(P' - A') - \beta_4 \sin 2(P' - A')\}. \quad (7.58)$$

In the derivation of these expressions, the general Mueller matrix from Ref. 8 was applied for the two compensators. In Eqs. (7.49) through (7.58), $B_{2n} = \alpha_{2n} + i\beta_{2n}$, $B_{2n}^* = \alpha_{2n} - i\beta_{2n}$, $s_j = \sin^2(\delta_j/2)$, and $t_j = \tan^2(\delta_j/2)$ ($j = 1,2$). In addition, δ_j ($j = 1,2$) represent the slow-to-fast axis phase shifts for the first and second rotating compensators. Again, the two equalities for M_{11} at the far right in Eq. (7.48) are appropriate only for non-depolarizing, isotropic samples. For the evaluation of Eqs. (7.47) through (7.58) in data reduction, one requires I_{00}, δ_1, δ_2, C_{S1}, C_{S2}, $P' = P - P_S$, and $A' = A - A_S$, which are determined in calibration as described in Section 7.3.3.4.

As an example of the capabilities of the dual-rotating compensator multichannel ellipsometer, Fig. 7.18 shows the 15 spectra in m_{ij} measured in transmission at normal incidence for a helicoidally sculptured MgF_2 thin film[61] nominally 4.7 μm thick and consisting of 12.3 right-handed pitches on a glass substrate. This is a demanding application of the multichannel ellipsometer owing to the weak anisotropy exhibited by this film, meaning that the normalized Mueller matrix is approximately given by $m_{ij} \approx \delta_{ij}$, i.e., the identity matrix. The data of Fig. 7.18 were obtained as an average over ten 0.25 s optical periods, leading to a total acquisition time of 2.5 s. In this application, the well-defined features appearing near 2.8 eV are associated with resonances in the circular birefringence and dichroism analogous to the Cotton effect in structurally homogeneous chiral media.[62] For the helicoidal sculptured thin film at normal incidence, these resonances occur at a wavelength of $\lambda_0 = 2\Omega < n > \approx 440$ nm, where $2\Omega \approx 382$ nm is the helicoid pitch and $< n > \approx 1.15$ is the effective index of refraction of the film for $\lambda = \lambda_0$. From the normalized Mueller matrix spectra of Fig. 7.18, the polarization modification effect of the sculptured thin film on any incident polarization state at any given photon energy ($2.0 \leq E \leq 3.75$ eV) can be computed.

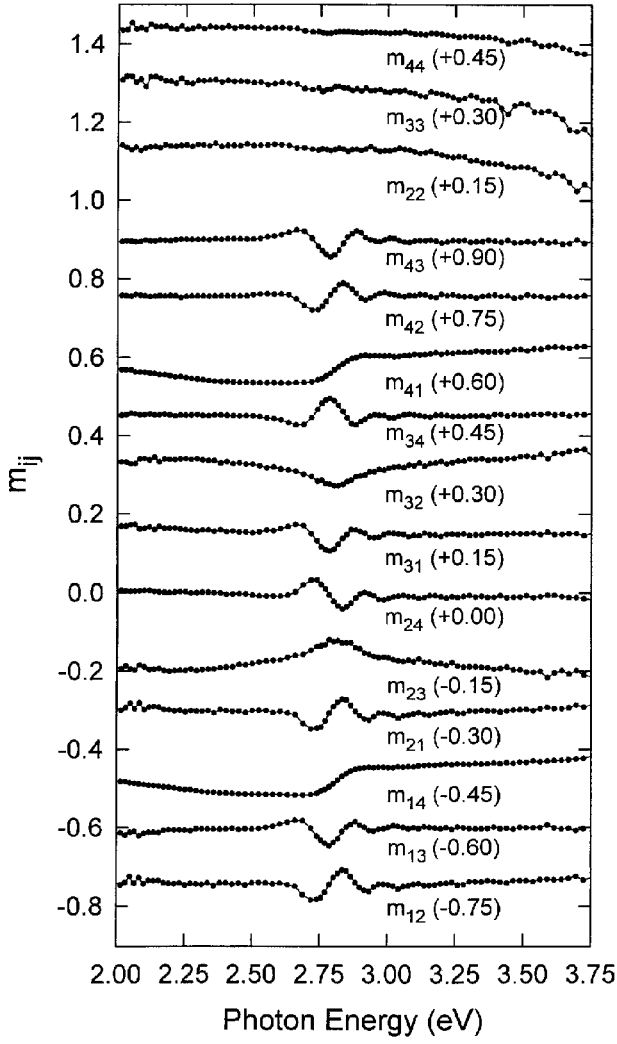


Figure 7.18 Fifteen spectra in the normalized Mueller matrix m_{ij} measured in transmission at normal incidence for a helicoidally-sculptured MgF_2 thin film nominally 4.7 μm thick and consisting of 12.3 right-handed pitches on a glass substrate. These data were obtained using a dual rotating-compensator multichannel ellipsometer as an average over ten 0.25 s optical periods, with a total acquisition time of 2.5 s. The features appearing near 2.8 eV are associated with resonances in the circular birefringence and dichroism. [Reproduced from Ref. 18.]

7.3.3.4 Calibration

For the dual rotating-compensator multichannel ellipsometer, the conventional sequence of calibration involves (i) measurement of the retardance spectra δ_1 and δ_2 in straight-through, (ii) determination of the optical element offset and phase angles (P_S, A_S, C_{S1}, C_{S2}) with the sample in place after its alignment, and (iii) measurement of the spectral response function I_{00} from initial measurements of the starting substrate.[63] The approach for these procedures is analogous to that described in Section 7.3.2 for the single rotating-compensator multichannel ellipsometer. To review, δ_1 and δ_2 are determined once in the initial development of the ellipsometer (or after any major system realignment) using the straight-through $PC_{1r}(5\omega_B)C_{2r}(3\omega_B)A$ configuration without a sample. For the most general sample, the offset and phase angles must be determined in a separate step every time a sample is mounted and aligned, i.e., just prior to the real time measurements. For an isotropic and non-depolarizing sample, complete self-calibration of the offset and phase angles is possible, i.e., calibration from the actual sample measurements themselves. Ideally, the spectral response function is determined from the initial real time measurements of the starting sample prior to thin film deposition or materials processing. In this case, however, the starting sample characteristics must be known so that its spectroscopic reflectance can be predicted; otherwise a known reference sample must be used.

An alternative calibration sequence is also possible that accounts for the effects of sample alignment on the beam path through the second compensator. As noted in Section 7.3.2.4, slight differences in beam path through the compensator, resulting when measuring samples with differing alignments, can lead to irreproducibilities in the retardance spectrum appropriate for the analysis of data for such samples. This problem can be minimized by measuring the retardance spectrum with the sample in place after the offset and phase angle calibration, but before the spectral response function calibration. In fact, with the dual rotating-compensator multichannel ellipsometer, the retardance spectra of the two compensators can be obtained in self-calibration under certain circumstances.

Retardance Calibration

Section 7.3.2.4 provides a description of the internal alignment procedure necessary to eliminate artifacts in the retardance spectra of the MgF_2 biplate compensator, the design of choice for wide spectral range in the dual rotating-compensator ellipsometer. Here, it is assumed that such a procedure has been performed for both compensators.

For measurements in the straight-through $PC_{1r}(5\omega_B)C_{2r}(3\omega_B)A$ configuration, in the absence of errors, only those Fourier coefficients with $n = 2$, 4, 6, 8, and 10 are non-zero. Under these conditions, the amplitudes of the $n = 4$, 6, and 10 uncorrected Fourier coefficients provide δ_1 and δ_2 directly, according to:

$$\delta_1 = 2 \arctan \sqrt{\frac{|B'_8|}{|B'_{12}|}}, \tag{7.59a}$$

$$\delta_2 = 2 \arctan \sqrt{\frac{|B'_8|}{|B'_{20}|}}, \tag{7.59b}$$

where $|B'_{2n}| = \sqrt{(\alpha'_{2n})^2 + (\beta'_{2n})^2}$. Equations (7.59a) and (7.59b) do not require phase-corrected coefficients. In addition, the unnormalized Fourier coefficients (a'_{2n}, b'_{2n}) can be used in place of (α'_{2n}, β'_{2n}) because the dc Fourier coefficient $I'_0 = i'_0 a_0$ cancels from the numerator and denominator in the argument of the inverse tangent function. As a result, the retardance calibration can be performed without the need to determine the optical element offsets P_S and A_S or the phase angles C_{S1} and C_{S2}. The procedures to obtain the four offset and phase angles are described shortly.

In a second approach for the determination of δ_1 and δ_2, the $PC_{1r}(5\omega_B)SC_{2r}(3\omega_B)A$ ellipsometer in the actual reflection configuration is used, and the calibration angles P_S and A_S are required. In this second approach, the sample must be non-depolarizing and isotropic so that its Mueller matrix is given by Eqs. (7.19) with $D = 1$. Thus, it is most convenient to perform this particular procedure as described in the next paragraph, after performing the offset and phase angle calibration that provides (P_S, A_S, C_{S1}, C_{S2}).

Once A_S is determined in the $PC_{1r}(5\omega_B)SC_{2r}(3\omega_B)A$ configuration [see, e.g., the self-calibration procedures of Eqs. (7.61) and (7.62), which can also be applied in a separate step prior to film growth or materials processing], the analyzer is set so that $A' = \pm\pi/4$. Then the second compensator retardance can be determined from:

$$\delta_2 = 2 \arctan \sqrt{\frac{2|B'_2|}{|B'_{10}(A' = \pm\pi/4)|}}. \tag{7.59c}$$

In this expression, it is emphasized that $|B'_{10}| = \sqrt{(\alpha'_{10})^2 + (\beta'_{10})^2}$ depends on the analyzer reading $A = A' + A_S$ which must be set at $A = (\pm\pi/4) + A_S$, in order for the equation to be valid. In contrast, $|B'_2| = \sqrt{(\alpha'_2)^2 + (\beta'_2)^2}$ depends only the sample and not on the polarizer or analyzer angular readings. Thus, Eq. (7.59c) can be used irrespective of the polarizer setting.

Once P_S is determined in the same $PC_{1r}(5\omega_B)SC_{2r}(3\omega_B)A$ configuration, then the polarizer can be set so that $P' = \pm\pi/4$ and the first compensator retardance can be determined from:

$$\delta_1 = 2 \arctan\sqrt{\frac{2|B'_{14}|}{|B'_6(P' = \pm\pi/4)|}}. \qquad (7.59d)$$

This expression can applied analogously to Eq. (7.59c), but with the polarizer (rather than the analyzer) set at $P = (\pm\pi/4) + P_S$.

Equations (7.59c) and (7.59d) exhibit their optimum sensitivity for a sample with $\psi = 45°$ and $\Delta = 90°$, i.e., for a metallic surface; they fail when $\psi = 0°$ or 90° or when $\Delta = 0°$ and ±180°. These sensitivities arise from the fact that the factors $|B'_{2n}|$ in the equations exhibit the proportionality $|B'_{2n}| \propto (\sin 2\psi \sin\Delta)/a_0$. As a result, the sample dependence cancels from the numerators and denominators in the inverse tangent arguments of Eqs. (7.59c) and (7.59d). Similarly, the unnormalized ac Fourier coefficients can also be used since the dc coefficient cancels as well. If real time measurements are performed with $A' = P' = \pm\pi/4$, and the starting sample is non-depolarizing and isotropic, then Eqs. (7.59c) and (7.59d) can be applied in a self-calibration mode using the data collected just prior to film growth or sample processing.

Offset and Phase Angle Self-Calibration

The angular offsets and phase shifts can be determined in the actual $PC_{1r}(5\omega)SC_{2r}(3\omega)A$ configuration from data collected either in a separate step after sample mounting and alignment or during the actual sample measurement, i.e., in the self-calibration mode (most often using the starting sample surface prior to film growth or sample processing). As usual, the phase functions are exploited, defined by

$$\Theta'_{2n} = (1/2)[\arctan(\beta'_{2n}/\alpha'_{2n})] = (1/2)[\arctan(\beta_{2n}/\alpha_{2n}) + \phi_{2n}], \qquad (7.60)$$

where $n = 1, 2, \ldots, 8, 10, 11, 13, 16$. In addition, the phases ϕ_{2n} in Eq. (7.60) are given in Table 7-1 or below Eqs. (7.47). As in the single rotating-compensator configuration, quadrant corrections to the phase functions will be needed to obtain self-consistent calibration and measurement results. These corrections will depend on the instrument values (P', A', C_{S1}, C_{S2}). As usual, it is best to pre-align the fast axis of each compensator relative to its motor shaft to ensure that C'_1 and C'_2 are near zero at the onset of data collection ($t = 0$) for a reference pixel. In

this way, approximate values of (P', A', C_{S1}, C_{S2}) are known in advance of the calibration.

The phase functions for the $n = 1, 2, 3, 5, 7, 8, 11$, and 13 harmonic components are simple enough to provide calibration information independent of the sample. For the most general sample in which case no conditions are placed on the normalized Mueller matrix elements, the phase functions can provide the following two linear combinations of the compensator phase shifts and polarizer and analyzer offset angles:

$$\begin{aligned} P - P_S + 5C_{S1} &= \Theta'_{10} \pm (\pi/4) = (\Theta'_4 + \Theta'_6) \pm (\pi/4) = (1/2)(\Theta'_4 + \Theta'_{16}) \\ &= (\Theta'_{16} - \Theta'_6) \pm (\pi/4) = (1/2)(\Theta'_{22} - \Theta'_2) \end{aligned} \quad (7.61a)$$

$$\begin{aligned} A - A_S + 3C_{S2} &= \Theta'_6 \pm (\pi/4) = (\Theta'_{10} - \Theta'_4) \pm (\pi/4) = (1/2)(\Theta'_{16} - \Theta'_4) \\ &= (\Theta'_{16} - \Theta'_{10}) \pm (\pi/4) = (1/2)(\Theta'_{26} - \Theta'_{14}). \end{aligned} \quad (7.61b)$$

For the most general sample, no linear combination of the phase functions has been found that can provide independent expressions for (P_S, A_S, C_{S1}, C_{S2}). For such a sample, complete self-calibration is not yet possible, and a separate measurement step as a function of the polarizer and analyzer angles prior to real time sample measurement is used to complete the calibration, as described in detail elsewhere.[59] Here, the calibration using a homogeneous (i.e., non-depolarizing), isotropic sample will be treated, in which case $\{M_{13}, M_{31}, M_{14}, M_{41}, M_{23}, M_{32}, M_{24}, M_{42}\}$ all vanish. With such a sample, complete self-calibration is possible starting with the following simple expressions for C_{S1} and C_{S2}:

$$5C_{S1} = \Theta'_2 + \Theta'_8 = \Theta'_{14} - \Theta'_4 = \Theta'_{26} - \Theta'_{16} = \Theta'_{32} - \Theta'_{22}, \quad (7.62a)$$

$$3C_{S2} = \Theta'_2 + \Theta'_4 = \Theta'_{14} - \Theta'_8 = \Theta'_{22} - \Theta'_{16} = \Theta'_{32} - \Theta'_{26}. \quad (7.62b)$$

Although many other linear combinations are possible, Eqs. (7.62) when applied with Eqs. (7.61) provide (P_S, A_S, C_{S1}, C_{S2}) with the minimum number of phase function terms (either three or four).

The self-calibration of Eqs. (7.62) is designed for real time measurement in the reflection geometry using the starting isotropic substrate surface prior to anisotropic thin film growth. The dual rotating-compensator ellipsometer also has many applications in the transmission measurement geometry for the analysis of anisotropic optical film growth. In this geometry, one can define the fixed polarizer transmission axis as the reference axis, in which case $P' = P - P_S \equiv 0$. Then for a starting isotropic substrate surface, self-calibration can be performed prior to anisotropic film growth using the following expressions:[63]

$$A - A_S = -(2\Theta'_4 - \Theta'_8), \tag{7.63a}$$

$$5C_{S1} = (1/2)(\Theta'_4 + \Theta'_{16}), \tag{7.63b}$$

$$3C_{S2} = (1/2)(3\Theta'_4 - 2\Theta'_8 + \Theta'_{16}). \tag{7.63c}$$

These equations are also designed to be valid in the transmission geometry in the absence of a sample. In other words, they can be applied either in straight-through prior to sample mounting and alignment, or in the self-calibration mode on an isotropic substrate at the start of real time measurement.

As an example of a calibration applying Eqs. (7.63), Fig. 7.19 shows (a) $A_S = A - A' = -A'$, (b) C_{S1}, and (c) C_{S2} for the dual rotating-compensator multichannel ellipsometer as measured in the transmission geometry without a sample.[63] For these results, the 1024-pixel photodiode array was operated without grouping. The value of A_S in Fig. 7.19(a) exhibits a standard deviation of 0.06° about an average value of −0.14°. Any systematic trends in A_S with photon energy lie within ±0.03° of the average value. Such behavior attests to the validity of the analyzer offset angle calibration, and demonstrates that any chromatic alignment errors are small. In Figs. 7.19(b–c), C_{S1} and C_{S2} exhibit linear relationships versus pixel number over the k = 450–850 pixel range (corresponding to the photon energy range of 2.27–4.15 eV) with slopes of −0.003609° and −0.003605°. These linear relationships arise because the absolute phases of the first and second compensators evolve according to $C'_1 = 5(\omega_B t - C_{S1})$ and $C'_2 = 3(\omega_B t - C_{S2})$, where ω_B corresponds to a base mechanical frequency of 2 Hz. In the experimental waveform of Eq. (7.45), time zero for a given pixel is defined as the time at which the S_1 integration of Eq. (7.46a) commences. At this time, the fast axes of the first and second compensators are at angles $C'_1 = -5C_{S1}$ and $C'_2 = -3C_{S2}$. Between the readout of two successive pixels, this time origin shifts by the single pixel readout time t_x of 5 μs. As a result, C_{S1} and C_{S2} should be linear functions of pixel number with slopes of $\omega_B t_x = -0.0036°$. The slopes of the linear fits in Figs. 7.19(b–c) are in close agreement with the expected value. The maximum difference from the expected value is ~0.3%, and the two independently determined slopes agree to within 0.1%. Such agreement in the measured slopes in successive calibrations over long periods attests to the high accuracy and stability of the 5:3 frequency ratio for the two compensators of this instrument.[63]

Spectral Response Calibration

For the unnormalized Mueller matrix, in the absence of depolarizing effects, M_{11} is no longer defined as unity, but rather is the specular

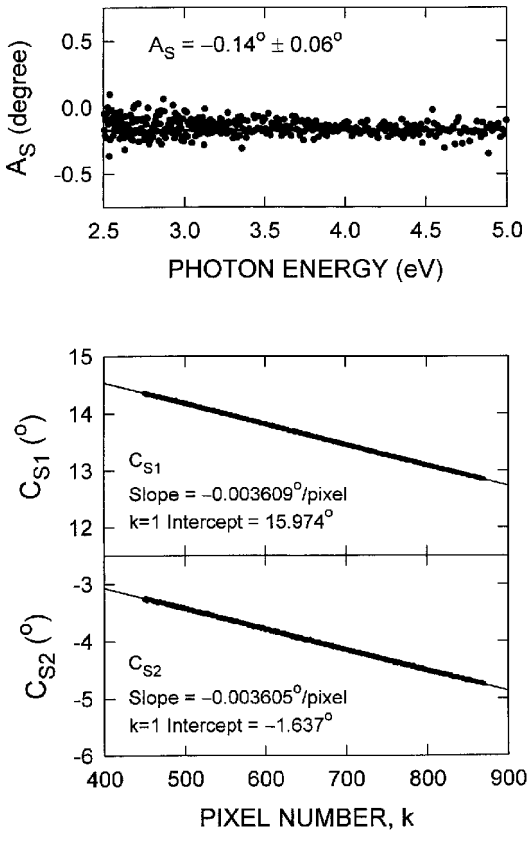


Figure 7.19 Calibration results for a dual rotating-compensator multichannel ellipsometer obtained in the straight-through $PC_{1r}(5\omega_B)C_{2r}(3\omega_B)A$ configuration without a sample: (a) the analyzer angular offset A_S plotted versus photon energy from Eq. (7.63a); (b) the first rotating-compensator phase angle C_{S1} plotted versus pixel number k from Eq. (7.63b); and (c) the second rotating-compensator phase angle C_{S2} plotted versus pixel number k from Eq. (7.63c). In (a), the spectral average of A_S is $-0.14°$ and the standard deviation from the average is 0.06°. In (b) and (c), the spectral range of the experimental data is 2.27–4.15 eV, and the lines represent the best fit linear relationships to C_{S1} and C_{S2} with the slopes and intercepts listed. [Adapted from Ref. 63.]

reflectance of the sample for incident unpolarized light, i.e., the ratio of the irradiance in the reflected light beam to that in the incident beam when the latter is completely unpolarized. In the transmission sample measurement geometry, M_{11} is the transmittance for unpolarized light. The critical step in the measurement of M_{11} is the determination of the ellipsometer spectral response function I_{00} as defined by Eq. (7.44). In the transmission

geometry, I_{00} can be obtained easily by removing the sample and setting the Mueller matrix ellipsometer in straight-through. In this case, $M_{11} = 1$ so $I_{00} = I_0/a_0$. In a real time measurement in the reflection geometry, if the nature of the initial starting surface at time zero is known, then this surface can be employed as a calibration standard for subsequent measurements during processing. Using this approach, one can write:[59]

$$M_{11}(t) = \frac{I'_0(t)a_0(0)}{I'_0(0)a_0(t)} R_u(0), \tag{7.64}$$

where $I'_0(0)$ and $I'_0(t)$ are the measured dc components of the waveform obtained from the known starting surface at time zero and from the unknown surface after a time t, respectively. The a_0 spectra are the results for the corresponding surfaces obtained after substituting the measured Fourier coefficients $(\alpha'_{2n}, \beta'_{2n})$ into the transformations of Eq. (7.47b), and in turn substituting these transformed coefficients into Eq. (7.49). Finally, $R_u(0) = M_{11}(0)$ and $M_{11}(t)$ are the assumed known and unknown (1,1) elements of the unnormalized Mueller matrices for the surface at time zero and time t, respectively. Once $M_{11}(t)$ has been determined using the known starting surface, then the other unnormalized matrix elements can be deduced by multiplying each of the expressions in Eqs. (7.50) through (7.58) by M_{11}.

For more complex starting surfaces, one can use separate measurements of a Si wafer as a reference calibration surface (denoted by 'cal'), as described in Sections 7.3.1 and 7.3.2. In this case, I_{00} is determined according to $I_{00} = I'_{0,\text{cal}}/[a_{0,\text{cal}}R_{u,\text{cal}}]$ for the calibration surface, and then M_{11} is obtained for any other surface according to $M_{11} = I'_0/I_{00}a_0$. This approach will be accurate as long as the overall instrument configuration and alignment is retained between measuring the calibration surface and the unknown surface under study.

As an example, Fig. 7.20 shows the spectrum in M_{11} plotted versus photon energy for the MgF_2 sculptured thin film of Section 7.3.3.3 and Fig. 7.18. These results were obtained from a set of waveform integrals collected ex situ in the transmission geometry in 2.5 s. In fact, this data set is the same as that from which the normalized Mueller matrix of Fig. 7.18 was extracted.[63] To obtain the spectrum in M_{11} of Fig. 7.20, I_{00} was determined prior to sample measurement from spectra in $I'_{0,\text{cal}}$ and $a_{0,\text{cal}}$, collected in the transmission geometry in the absence of a sample. In this I_{00} calibration, the equation $I_{00} = I'_{0,\text{cal}}/[a_{0,\text{cal}}T_{u,\text{cal}}]$ was applied, where $T_{u,\text{cal}} = 1$ is the transmittance for unpolarized light in the absence of a sample. Once I_{00} is established, M_{11} for the sculptured thin film is determined from measurements of I'_0 and a_0 in the same transmission geometry with the sample in place by applying $M_{11} = I'_0/I_{00}a_0$.

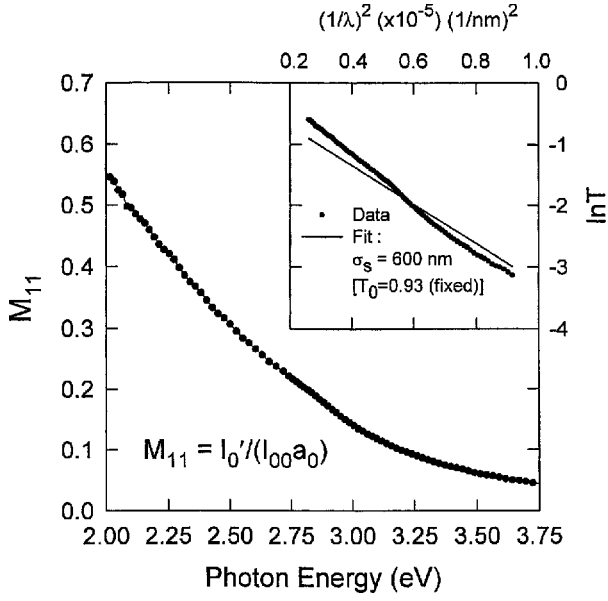


Figure 7.20 Spectrum in the unnormalized Mueller matrix element M_{11} measured in transmission through the helicoidally-sculptured MgF_2 thin film of Fig. 7.18. This spectrum was deduced from the same set of waveform integrals as was the normalized Mueller matrix of Fig. 7.18. The inset shows the data plotted as $\ln(T_u)$ versus λ^{-2}, where T_u is the transmsittance of the sample for unpolarized light. The solid line is the best fit linear relationship obtained by applying a model in which the transmission losses are due to non-specular light scattering. [Reproduced from Ref. 63.]

The spectrum in M_{11} in Fig. 7.20 exhibits a significant decrease from ~0.55 at 2 eV to ~0.05 at 3.75 eV that cannot be attributed to absorption or reflection losses, but rather to non-specular scattering of the incident light beam at the surface of the sculptured thin film. These scattering losses limit the spectral range to photon energies $E < 3.75$ eV. In spite of the significant decrease in M_{11}, a weak feature is detectable near 2.8 eV which is a manifestation of the resonance in the circular birefringence and dichroism at $\lambda = \lambda_0 = 440$ nm. If it is assumed that macroscopic roughness on the sculptured thin film surface generates the non-specular scattering, then the effect can be simulated using the simple expression: $M_{11} = T_u \approx T_{0,u} \exp\{-[2\pi(1 - \langle n \rangle)\sigma_s/\lambda]^2\}$. Here σ_s is the full-width of the assumed Gaussian-distributed surface profile and $T_{0,u}$ is the unpolarized transmittance in the absence of the macroscopic roughness. The solid line in the inset is a single-parameter fit to the data (points) for this simple surface scattering model.

7.4 Concluding Remarks

Instruments with sufficiently high spectral scanning speed for real time spectroscopic ellipsometry have been developed based on self-compensating, rotating-element, and phase modulation principles. Each type of instrument has its strengths and weaknesses with respect to instrument speed, spectral point-density, spectral resolution, measurement precision, and accuracy, and each has different trade-offs that must be addressed to achieve satisfactory performance in specific applications. Rotating-element instruments have been most successful, both as a commercial product and as a research tool. The two primary advantages of the rotating-element design include (i) its fully parallel operation when used with an integrating photodiode array detector and (ii) its versatility which has led to the development of instruments for real time Stokes vector and Mueller matrix spectroscopies. As a result, the primary focus of this review has been the rotating-element multichannel ellipsometers in the rotating-polarizer, single rotating-compensator, and dual rotating-compensator configurations.

The rotating-polarizer multichannel ellipsometer is best suited for studies of strongly-absorbing material structures, e.g., semiconductor or metal thin films on semiconductor substrates, structures that are isotropic and homogeneous over the full extent of the incident beam (and hence non-depolarizing). The advantage of the rotating-polarizer instrument in such applications is its simplicity, ease of calibration, and potentially wide spectral range (1.5 to 6.5 eV in one instrument), as well as its high speed, precision, and accuracy. This instrument provides only partial information on the Stokes vector of the reflected beam, however, and is not accurate when the ellipticity angle of the reflected beam polarization state is small. Thus, for sample structures that incorporate thick weakly absorbing layers or transparent substrates, structures that are inhomogeneous and depolarizing, or structures that include anisotropic materials or films, the rotating-polarizer configuration becomes less accurate and/or provides insufficient information. In fact, if some of these complexities are unwittingly present, e.g., inhomogeneity or anisotropy, then the rotating-polarizer ellipsometer can provide erroneous results.

For more complex sample structures, the introduction of one or more rotating MgF_2 compensators can improve the accuracy and/or provide more information at the expense of more complicated calibration and data reduction procedures. The single-rotating compensator multichannel ellipsometer provides the complete Stokes vector of the beam reflected from the sample surface. As a result, it is suitable for the analysis of sample structures that incorporate weakly absorbing layers or transparent substrates, and can also handle some relatively simple sample inhomogeneities.

As a further step toward the analysis of even more complex samples, the dual rotating-compensator multichannel ellipsometer provides the full Mueller matrix of the sample. As a result, it is suitable for the analysis of samples exhibiting both heterogeneity and anisotropy. In fact, owing to its very recent development, the full range of applications of this instrument has yet to be demonstrated. Thus, it is expected that the characterization of anisotropic thin film growth will be one of the important future applications of multichannel ellipsometry. With the recent development of dual rotating-compensator ellipsometry, it would appear that there is little room for further instrumentation advances in the future. However, numerous minor and major improvements on a number of levels are possible in the future.

For the multichannel instruments in general, advances in the design of sources and detection systems can lead to increases in spectral range and improvements in speed, accuracy, and precision. The recent application of a tandem Xe-D_2 source in rotating polarizer multichannel ellipsometry has led to a spectral range of 1.5 to 6.5 eV.[16] Other sources or source combinations can be tailored for extended ranges in the near-infrared for characterization of the free carrier characteristics of degenerate semiconductors used as transparent conductors or in the vacuum ultraviolet for characterization of dielectric materials used in microelectronics and in photolithography. Current detection systems employ unintensified Si photodiode arrays owing to their high performance at relatively high light levels. Breakthroughs in detector technology, e.g., multichannel photomultipliers, may find application in the ellipsometers of the future. Rotating-compensator multichannel ellipsometers suffer from a loss in precision and accuracy at photon energies where δ is far from the quarterwave value ($\delta = 90°$). Such instruments, in particular those in the dual rotating-compensator configuration, would benefit greatly from achromatic rotating compensators such as the three reflection rhomb. Thus, mounting and alignment techniques that avoid stress and ensure transmitted beam stability under rotation of the rhomb are necessary in implementing such an improvement.

Finally, the applications of multichannel ellipsometry that deserve more attention in the future involve high resolution probing of surfaces. Although the emphasis of the present article has been on instruments that probe surfaces with a collimated beam (1-10 mm diameter spot size), a rotating-polarizer multichannel ellipsometer with a focused beam (0.1 mm spot size) has been applied in-line within a semiconductor process environment.[64] The high speed of the multichannel ellipsometer (minimum of 16 ms per measurement) provides the capability of spectroscopic mapping of surfaces. Future directions involve reducing the spot size closer to the diffraction limit for the longest wavelengths of the spectrum and

adapting focusing optics to more advanced configurations such as the dual rotating-compensator ellipsometer. This may require new directions in the design of sources (e.g., increased brightness) and detection systems (e.g., increased sensitivity at low light levels). The dual rotating-compensator ellipsometer may provide improved capabilities in the characterization of patterned surfaces with feature sizes smaller than the spot size of the instrument, owing to its ability to extract the full Mueller matrix. Finally the use of polarization discriminating optics in near-field scanning optical microscopy is the first step in developing metrological instrumentation with resolution below the diffraction limit.[65]

Acknowledgements

The preparation of this chapter was supported in part by the National Science Foundation under Grant Nos. DMR-9217169, DMR-9820170, and DMR-0137240, and by the National Renewable Energy Laboratory under subcontract Nos. XAF-8-17619-22 and AAD-9-18668-9.

7.5 References

1. I. P. Herman, *Optical Diagnostics for Thin Film Processing*, Academic Press, New York (1995).
2. A. C. Boccara, C. Pickering, and J. Rivory, editors, *Proceedings of the First International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam (1993); also published as *Thin Solid Films*, **233–234** (1993).
3. R. W. Collins, D. E. Aspnes, and E. A. Irene, editors, *Proceedings of the Second International Conference on Spectroscopic Ellipsometry*, Elsevier, Amsterdam (1998); also published as *Thin Solid Films*, **313–314** (1998).
4. D. E. Aspnes, in *Optical Properties of Solids New Developments*, edited by B. O. Seraphin, North-Holland, Amsterdam (1976), p. 799.
5. R. H. Muller and J. C. Farmer, *J. Phys. (Paris), Colloque*, **C10**, Suppl. 12, 57 (1983).
6. J. B. Theeten and D. E. Aspnes, *Ann. Rev. Mater. Sci.*, **11**, 97 (1981).
7. F. L. McCrackin and J. P. Colson, in *Ellipsometry in the Measurement of Surfaces and Thin Films*, edited by E. Passaglia, R. R. Stromberg, and J. Kruger, Nat. Bur. Stand., Misc. Publ. 256, Washington DC (1964), p. 61.
8. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, North-Holland, Amsterdam (1977).
9. J. A. Woollam and P. G. Snyder, *Mater. Sci. Eng. B,* **5**, 279 (1990).
10. R. W. Collins, *Rev. Sci. Instrum.*, **61**, 2029 (1990).
11. W. M. Duncan and S. A. Henck, *Appl. Surf. Sci.*, **63**, 9 (1993).
12. (a) SOPRA S.A., 26 rue Pierre Joigneaux, 92270 Bois-Colombes, France; (b) J. A. Woollam Co., Suite 102, 645 M Street, Lincoln, NE 68508, USA.

13. I. An, Y. M. Li, H .V. Nguyen, and R. W. Collins, *Rev. Sci. Instrum.*, **63**, 3842 (1992).
14. I. An, H. V. Nguyen, A. R. Heyd, and R. W. Collins, *Rev. Sci. Instrum.*, **65**, 3489 (1994).
15. R. W. Collins, I. An, H. Fujiwara, J. Lee, Y. Lu, J. Koh, and P. I. Rovira, *Thin Solid Films*, **313-314**, 18 (1998).
16. J. A. Zapien, R. W. Collins, and R. Messier, *Rev. Sci. Instrum.*, **71**, 3451 (2000).
17. J. Lee, P. I. Rovira, I. An, and R. W. Collins, *Rev. Sci. Instrum.*, **69**, 1800 (1998).
18. J. Lee, J. Koh, and R. W. Collins, *Opt. Lett.*, **25**, 1573 (2000).
19. R. H. Muller and J. C. Farmer, *Rev. Sci. Instrum.*, **55**, 371 (1984).
20. J. C. Farmer and R. H. Muller, *J. Electrochem. Soc.*, **132**, 313 (1985).
21. S. T. Mayer and R. H. Muller, *J. Electrochem. Soc.*, **135**, 2133 (1988).
22. J. L. Ord, *Surf. Sci.*, **16**, 155 (1969).
23. H. J. Mathieu, D. E. McClure, and R. H. Muller, *Rev. Sci. Instrum.*, **45**, 798 (1974).
24. Y.-T. Kim, R. W. Collins, and K. Vedam, *Surf. Sci.*, **233**, 341 (1990).
25. R. W. Collins and Y.-T. Kim, *Analytical Chem.*, **62**, 887A (1990).
26. Y. Cong, I. An, R. W. Collins, K. Vedam, H. Witham, L. J. Pilione, and R. Messier, *Thin Solid Films*, **193/194**, 361 (1990).
27. G. K. T. Conn and G. K. Eaton, *J. Opt. Soc. Am.*, **44**, 546 (1954).
28. W. Budde, *Appl. Opt.*, **1**, 201 (1962).
29. B. D. Cahan and R. F. Spanier, *Surf. Sci.*, **16**, 166 (1969).
30. D. E. Aspnes, *Opt. Commun.*, **8**, 222 (1973).
31. D. E. Aspnes and A. A. Studna, *Appl. Opt.*, **14**, 220 (1975).
32. B. Johs, D. Doerr, S. Pittal, I. B. Bhat, and S. Dakshinamurthy, *Thin Solid Films*, **233**, 293 (1993).
33. J.-P. Piel, J.-L. Stehle, and O. Thomas, *Thin Solid Films*, **233**, 301 (1993).
34. N. V. Nguyen, B. S. Pudliner, I. An, and R. W. Collins, *J. Opt. Soc. Am. A*, **8**, 919 (1991).
35. S. H. Russev, *Appl. Opt.*, **28**, 1504 (1989).
36. G. Laurence, F. Hottier, and J. Hallais, *Rev. Phys. Appl. (Paris)*, **16**, 579 (1981).
37. S. A. Henck, W. M. Duncan, L. M. Lowenstein, and S. W. Butler, *J. Vac. Sci. Technol. A*, **11**, 1179 (1993).
38. S. N. Jasperson and S. E. Schnatterly, *Rev. Sci. Instrum.*, **40**, 761 (1969).
39. V. M. Bermudez and V. H. Ritz, *Appl. Opt.*, **17**, 542 (1978).
40. G. E. Jellison, Jr., and F. A. Modine, *J. Appl. Phys.*, **53**, 3745 (1982).
41. B. Drevillon, J. Perrin, R. Marbot, A. Violet, and J. L. Dalby, *Rev. Sci. Instrum.*, **53**, 969 (1982).
42. B. Drevillon, *Prog. Cryst. Growth Charact. Mater.*, **27**, 1 (1993).
43. D. E. Aspnes, *J. Opt. Soc. Am.*, **64**, 639 (1974).
44. P. S. Hauge and F. H. Dill, *Opt. Commun.*, **14**, 431 (1975).
45. D. E. Aspnes, *J. Opt. Soc. Am.*, **65**, 1274 (1975).
46. I. An and R. W. Collins, *Rev. Sci. Instrum.*, **62**, 1904 (1991).
47. I. An, Y. Cong, N. V. Nguyen, B. Pudliner, and R. W. Collins, *Thin Solid Films*, **206**, 300 (1991).

48. B. Hong, J. Lee, R. W. Collins, Y. Kuang, W. Drawl, R. Messier, T. T. Tsong, and Y. E. Strausser, *Diamond and Related Mater.*, **6**, 55 (1997).
49. Y. Talmi and R. W. Simpson, *Appl. Opt.*, **19**, 1401 (1980).
50. D. E. Aspnes, *J. Opt. Soc. Am.*, **64**, 812 (1974).
51. B. Johs, *Thin Solid Films*, **234**, 395 (1993).
52. J. M. M. de Nijs, A. H. M. Holtslag, A. Hoeksta, and A. van Silfhout, *J. Opt. Soc. Am. A*, **5**, 1466 (1988).
53. I. An, J. Lee, B. Hong, and R. W. Collins, *Thin Solid Films*, **313–314**, 79 (1998).
54. U. Rossow, *Thin Solid Films*, **313–314**, 97 (1998).
55. S.-M. F. Nee, *Appl. Opt.*, **35**, 3570 (1996).
56. J. Lee and R. W. Collins, *Appl. Opt.*, **37**, 4230 (1999).
57. J. Lee, P. I. Rovira, I. An, and R. W. Collins, *J. Opt. Soc. Am. A*, **18**, 1980 (2001).
58. J. Opsal, J. Fanton, J. Chen, J. Leng, L. Wei, C. Uhrich, M. Senko, C. Zaiser, and D. E. Aspnes, *Thin Solid Films*, **313–314**, 58 (1998).
59. R. W. Collins and J. Koh, *J. Opt. Soc. Am. A*, **16**, 1997 (1999).
60. R. W. Collins, P. I. Rovira, A. S. Ferlauto, J. Koh, I. An, J. A. Zapien, R. Messier, and C. R. Wronski, *Mater. Res. Soc. Symp. Proc.*, **569**, 43 (1999).
61. K. Robbie, M. J. Brett, and A. Lakhtakia, *J. Vac. Sci. Technol. A*, **13**, 2992 (1995).
62. P. I. Rovira, R. A. Yarussi, R. W. Collins, V. C. Venugopal, A. Lakhtakia, R. Messier, K. Robbie, and M. J. Brett, *Appl. Phys. Lett.*, **71**, 1180 (1997).
63. J. Lee, J. Koh, and R. W. Collins, *Rev. Sci. Instrum.*, **72**, 1742 (2001).
64. C. Pickering, J. Russell, V. Nayar, J. Imschweiler, H. Wille, S. Harrington, C. Wiggins, J.-L. Stehle, J.-P. Piel, and J. Bruchez, *Thin Solid Films*, **313–314**, 446 (1998).
65. E. B. McDaniel, S. C. McClain, and J. W. P. Hsu, *Appl. Opt.*, **37**, 84 (1998).

Critical Reviews of Some Applications

8 SiO_2 Films

Eugene A. Irene

Department of Chemistry, University of North Carolina

8.1 Introduction

8.1.1 Preeminence of SiO_2 in Microelectronics: the Ellipsometry Connection

SiO_2 is one of the most important and studied materials in the field of microelectronics. The underlying reasons for the preeminence of SiO_2 in microelectronics lie with the unique properties of the SiO_2 film and more importantly the Si-SiO_2 interface. Ellipsometry has played a crucial role in the development of SiO_2 and other films used in microelectronics, and ellipsometry has also had a major role as a general method for thin film characterization. The SiO_2 related issues, especially the link between SiO_2 and ellipsometry are the focii of this chapter. The specific objectives of this chapter are twofold. One objective is to demonstrate the central role of ellipsometry in microelectronics through its role in the development of knowledge about SiO_2. The second objective is to use SiO_2 as a nearly perfect material with which to apply and demonstrate many ellipsometry techniques and variances of techniques for thin film measurement.

The primary application of SiO_2 films is as an electronically passivating film for Si based electronic devices that comprise an overwhelming majority of present day electronic devices. This specific application cannot presently be accomplished to the required level with any other material. Therefore, a brief discussion of the unique electronics passivation function of SiO_2 on single crystal Si wafer surfaces is appropriate and will follow. In addition, SiO_2 films have other important functions within Si technology and microelectronics such as for electrical isolation between devices on a densely populated chip, as a masking film for lithography, for metal contact/line insulation, and for packaging. These other functions for SiO_2 films can be and oftentimes are accomplished by other materials, and thus are not unique to SiO_2. Many details about SiO_2, such as the uses of SiO_2 in Si technology, the formation of SiO_2 films via oxidation of Si and other methods, the kinetics of formation of SiO_2 films and SiO_2 film properties, have been written about extensively and reviews are available in

the literature.[1–4] Only a brief discussion is reiterated here with the intention to make connections with the field of ellipsometry, the subject of this Handbook.

8.1.2 Electronic Passivation

Electronic passivation of a semiconductor surface refers to the reduction of the surface electronic states from bare surface intrinsic levels (approximately 10^{15} states $eV^{-1}cm^{-2}$) to levels below which Si based electronic devices (metal oxide semiconductor field effect transistors, MOSFETs) can be operated (below approximately 10^{12} states per eV per cm^2). In order to ensure high quality and reliable device operation, Si surface electronic state levels of 10^{10} eV^{-1} cm^{-2} and below are generally demanded of state-of-the-art device process. However, thus far the only reliable way to achieve this level of electronic passivation is through the use of SiO_2 films prepared by the thermal oxidation of Si. No other semiconductor-film combination has been found that can achieve the required level of electronic passivation attainable with SiO_2 films on Si. Needless to say, this fact has led to the overwhelming prominence of SiO_2 in Si technology and indeed the dominance of Si technology itself. As will be presented below, many of the important details of this growth process and properties of the materials can be elucidated through the use of ellipsometry.

SiO_2 has other very important uses in Si technology as was mentioned above. As a wide band gap dielectric (9 eV), SiO_2 can support large electric fields without significant leakage current. Consequently, SiO_2 is used as a gate dielectric in MOSFETs to both passivate the Si as defined above, and to support the electric field necessary to invert the carrier type in the MOSFET channel and thereby enable the device to turn on, that is to enable current to flow from the source to the drain elements through the inverted channel of the MOSFET. A cross section of a simple MOSFET is shown in Fig. 8.1, in order to identify the device elements and where the passivation SiO_2 film is located, viz. in the gate region between the source and drain. For the N-Channel MOSFET in Fig. 8.1, a positive potential applied to the gate electrode causes the channel region in the Si beneath the gate oxide to become electron enriched. This Si region that was p-type is now n-type with the positive gate bias and this process where the carrier type is changed is termed inversion. The inversion layer near the $Si\text{-}SiO_2$ interface connects the n-type source and drain. Electronic states at this interface (interface electronic states) would deleteriously affect the electron conduction process and the device characteristics. By virtue of its excellent electrical insulating quality, SiO_2 is used to isolate one device from another on a chip. It is also used as an overall electrical isolating

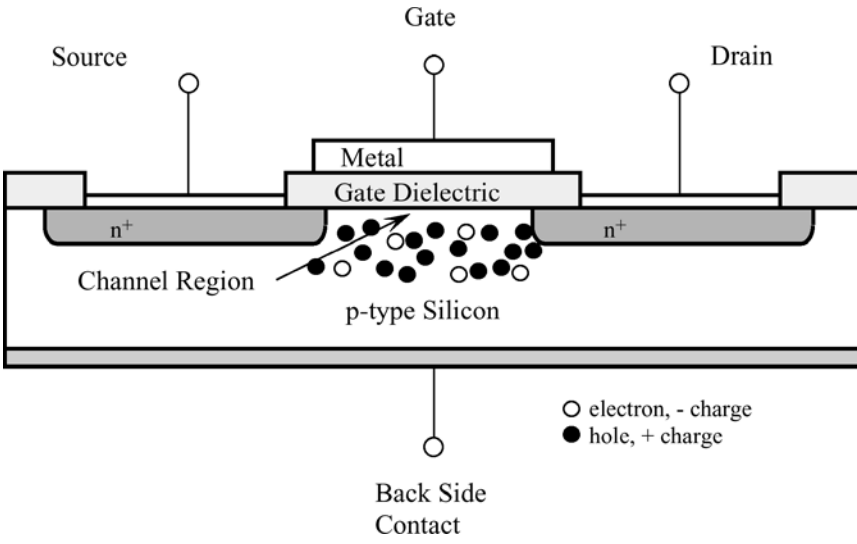


Figure 8.1 Pictorial representation of an N-Channel MOSFET. The gate dielectric SiO_2 film is located below the metal gate contact and above the channel region of the transistor.

encapsulant although modern devices mostly use other dielectrics for that purpose. SiO_2 is readily removed using HF aqueous solutions, and using plasmas containing halogens. Consequently, it is a desirable masking film for use in lithography to support photoresists.

It is true that except for electronic passivation as was discussed above, other dielectrics and materials could function equally well and sometimes better than SiO_2 for the various applications in Si technology. However, in order to minimize the number of other materials used in the technology, and because only SiO_2 will electronically passivate Si so thoroughly, SiO_2 is typically preferred in applications in which it is acceptable albeit not required. While SiO_2 is used mostly in Si technology, because it is an excellent dielectric, it is chemically stable, transparent in the visible spectrum, and the preparation methods are well known, SiO_2 films also find wide applications within other semiconductor and optical technologies.

8.1.3 Properties of SiO_2 Films

Before proceeding to the important role of ellipsometry for SiO_2 in Si and various other technologies it is useful to briefly review some of the important properties of SiO_2, in addition to its ability to electronically passivate the Si surface, that render it such a useful material.

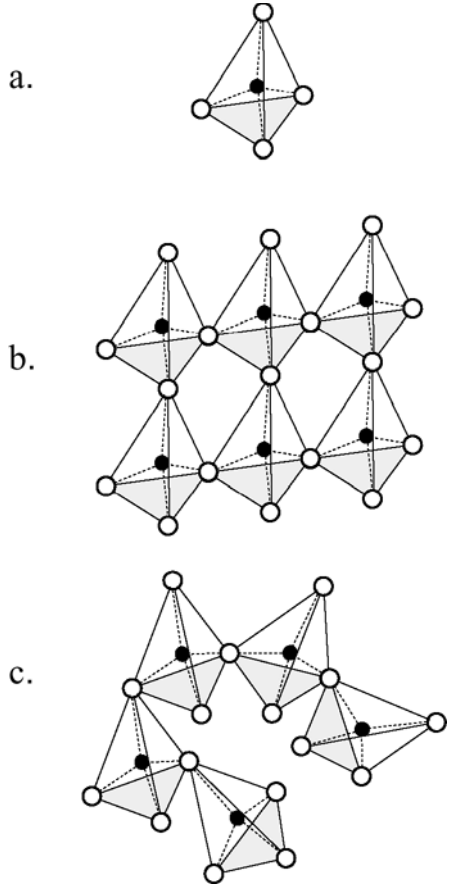


Figure 8.2 Structure of SiO_2: a) SiO_4 tetrahedra building blocks; b) SiO_4 tetrahedra arranged with long range order, c-SiO_2; c) SiO_4 tetrahedra arranged without long range order, a-SiO_2. Open circles represent oxygen and closed circles represent silicon.

Structurally, SiO_2 is composed of SiO_4 tetrahedra as illustrated in Fig. 8.2a. The tetrahedra are joined to one another by means of the oxygen atoms at the corners of the tetrahedra, the so-called bridging oxygens. If the bond angles for the Si-O-Si bonds, viz. the bond angles between adjacent tetrahedra, are tightly distributed, crystalline SiO_2 (c-SiO_2) results (Fig. 8.2b). If the bond angles are randomly distributed, then an amorphous random network results (Fig. 8.2c). Both forms of SiO_2 are important. For electronic thin film applications the amorphous form is the most important and is obtained from both the thermal oxidation of Si

and chemical vapor deposition at process temperatures less than 1100° C. It should be noticed that the fundamental building blocks of amorphous SiO_2 (a-SiO_2) are the SiO_4 tetrahedra which have considerable short-range order that narrowly fixes the stoichiometry. However, because of the randomly distributed bridging bond angles, this form of SiO_2 displays no long-range order and is therefore amorphous. Amorphous SiO_2 has an open network structure in which some foreign species can reside and others can readily diffuse through. Due to the inability to pack the building blocks efficiently when random bridge angles occur, the density of a-SiO_2 is lower than for the crystalline modifications.

As was pointed out in previous chapters, ellipsometry measures Δ and Ψ that can be recast as the dielectric function, ε, for a pure substance. According to the Clausius-Massotti relationship:

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{4\pi}{3}\rho\alpha \tag{8.1}$$

where ε is the dielectric function for a pure material and α is the bond polarizability, a microscopic quantity, α, that is connected to the macroscopic measurable dielectric response, ε. ρ is the number per volume or density of polarizable species. Also ε is a complex quantity with a real and imaginary part, ε_1 and ε_2, respectively and ε is given as:

$$\varepsilon = \varepsilon_1 + i\varepsilon_2 \quad \text{and} \quad \varepsilon_1 = n^2 - k^2, \quad \varepsilon_2 = 2nk \tag{8.2}$$

where n is the real part of the refractive index and k is the absorption constant. It should be noted that the refractive index N is in general complex:

$$N = n + ik \tag{8.3}$$

So that for a material such as SiO_2 with a 9 eV band gap and thus where $k = 0$ in the visible part of the optical spectrum:

$$\varepsilon = n^2 \tag{8.4}$$

Therefore, a decrease in ρ can cause a decrease in the real part of the refractive index, n. This analysis assumes that there is no difference in the α's for c-SiO_2 and a-SiO_2, which is a reasonable assumption, since the polarizability is mainly determined by the local chemical bonding (the short-range order) which is essentially the same for c-Si and a-Si.

Optically, both c-SiO_2 and a-SiO_2 are wide band gap (~9 eV) materials, and thus optically transparent in the visible spectrum (1.5−5 eV). Of course some band tailing is expected for a-SiO_2 as a result of the lack of

long-range order, and this would effectively narrow the gap slightly by a few tenths of an eV and add localized electronic states. However, no device problems associated with a-SiO_2 have been reported when it is made using high temperature thermal oxidation of Si in dry O_2. The optical transparency permits the assumption made above that the optical absorption is essentially zero ($k = 0$) for both c-SiO_2 and a-SiO_2 throughout the visible portion of the optical spectrum and this assumption enables a reduction in the number of variables to be determined by ellipsometry.

The best electronic quality SiO_2 is grown upon a single crystalline Si substrate via oxidation of the Si surface in pure oxygen ambient at elevated temperatures. Also, SiO_2 is often deposited using chemical vapor deposition (CVD) where deposition occurs largely from a gas phase reaction that produces the desired product or a precursor that in turn rapidly changes to the desired product upon condensation. The Si substrate is typically pure intrinsic single crystal Si that is cut into wafers and highly polished and sometimes doped with electrically active impurities such as B and P. The single crystal Si wafers are virtually free of crystallographic defects. As was discussed in previous chapters on the fundamentals of ellipsometry, particularly ellipsometry performed on thin film systems, it is a decided advantage to know ahead of time the optical properties (ε_1 and ε_2 or n and k) of the substrate. Since ellipsometry directly evaluates two parameters, Δ and Ψ, a single ellipsometric measurement can yield only two unknown parameters. For Si a number of high quality measurements of ε_1 and ε_2 or n and k are available from the literature.[5] While the available databases are similar over much of the spectrum covered, there are differences. In principle it is always possible to obtain a database on the bare substrate prior to film deposition. In practice one must consider that even in controlled environments a bare surface, even a relatively inert surface like Au or Pt, will accumulate nanometers of airborne and even vacuum borne contamination. For a chemically reactive surface like Si in addition to contamination, a native oxide will form in seconds. Thus high quality substrate measurements must be made with the utmost care. Typically the newest measurements available are the best, because the workers are aware of the older measurements and any problems associated with the older measurements. For Si this author prefers the newest measurements contributed by Jellison[6] in which the utmost care has been taken with both the measurement and data analysis.

It is worthwhile to quantify the sensitivity of ellipsometry to the presence of overlayers on a substrate whether the overlayer is an intentional film or contamination. Table 8-1 shows calculated results for a Si surface coated with a film with $n = 1.5$ and $k = 0$ for 632.8 nm light and the angle of incidence $\phi = 70°$. Under these conditions of the calculation for an imaginary film that is similar to SiO_2, it is seen that Δ changes by about 0.3°

Table 8-1. Calculated Δ, Ψ for various film thickness with conditions: n = 1.5 on Si, λ = 632.8 nm, ϕ = 70°.

Δ	Ψ	Thickness (nm)
179.257	10.448	0.0
178.957	10.448	0.1
178.657	10.449	0.2
178.356	10.450	0.3
178.056	10.451	0.4
177.756	10.453	0.5
176.257	10.462	1.0

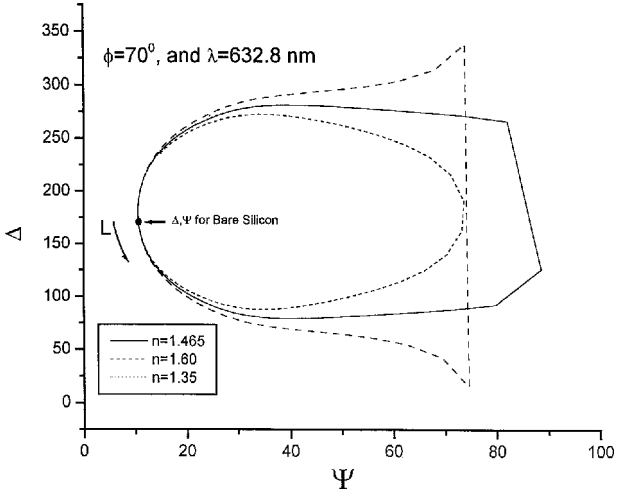


Figure 8.3 A simulated map of, Δ, Ψ space for three transparent films (k = 0) with different real parts for the refractive index on a single crystal Si substrate calculated with ϕ = 70° and λ = 632.8 nm. The bare Si point (L = 0) is indicated by an arrow and the direction for increasing film thickness, L, is also indicated by an arrow.

per 0.1 nm of film. Considering that a properly aligned ellipsometer with high quality optics is capable of precision of about an order of magnitude better in Δ and Ψ or 0.01 to 0.02°, submonolayer sensitivity is readily achievable with a properly aligned and calibrated ellipsometer.

Figure 8.3 shows a simulated map of Δ, Ψ space for three different films on a Si substrate. The film with n = 1.465 closely resembles SiO_2.

The three films are transparent ($k = 0$) but they have different n's. Near the origin, namely where the film thickness is zero, which represents Δ and Ψ for bare Si, Ψ is nearly unchanged as the film thickness (L) increases while Δ decreases sharply. Thus for thin transparent films on Si (near the origin), Δ mainly determines the film thickness.

It is seen that beyond about 10 nm, it is Ψ that changes while Δ is nearly constant. Also, it is observed that the Δ, Ψ space is repetitive or cyclic. Thus one cannot distinguish thickness in the first repetition from the second and so on. These repetitions are called ellipsometric periods (P) and for a single film on a substrate P is given by:

$$P = \frac{1}{2} \frac{\lambda}{\sqrt{N_1^2 - N_0^2 \sin^2\phi}} \tag{8.5}$$

P is determined by λ, ϕ and the complex refractive indexes for the film, N_1, and ambient, N_o. For non-absorbing films and ambient, $k = 0$, $N = n$ and P is real and a repetitive curve in Δ, Ψ space results. Thus in order to obtain film thickness, one must know *a priori* in which ellipsometric period is the film thickness. Another point gleaned from Fig. 8.3 is that near half of an ellipsometric period (right side of the figure) the Δ, Ψ thickness trajectories are separated maximally while near $L = 0$ and for full periods there is minimum separation of the trajectories. This means that in order to obtain the most reliable n values for films, one should have sample film near 1/2 an ellipsometric period. Conversely it is difficult to obtain reliable n values for thin films near $L = 0$. In addition there are many studies that indicate that the refractive index for SiO_2 films on Si is not constant with film thickness,[7–9] and because of the technological importance of ultra-thin SiO_2 films (<5 nm) this subject will be treated in some detail later.

As was discussed in previous chapters, the simplest model used for the analysis of ellipsometry measurements on films grown or deposited upon substrates is a model that includes a single homogenous film with a sharp interface on an infinite substrate, all in an ambient with $n = 1$ and $k = 0$. This single film model is shown as Fig. 8.4a. If substrate n and k values are supplied and if $k = 0$ for the film, then it is possible to obtain L_f and n_f (the subscript f is to indicate the film) for the film from a single measurement of Δ, Ψ, since λ and ϕ are experimentally defined. This ideal model can be applied with reasonable confidence to SiO_2 films on Si wafers when the SiO_2 film thickness is greater than about 20 nm. This simple model has been used extensively to characterize the SiO_2 film growth dynamics resulting from the oxidation of Si and more details will be presented in this chapter. However, now the microelectronics requirements are for thinner SiO_2 films. For the thin film regime the interface

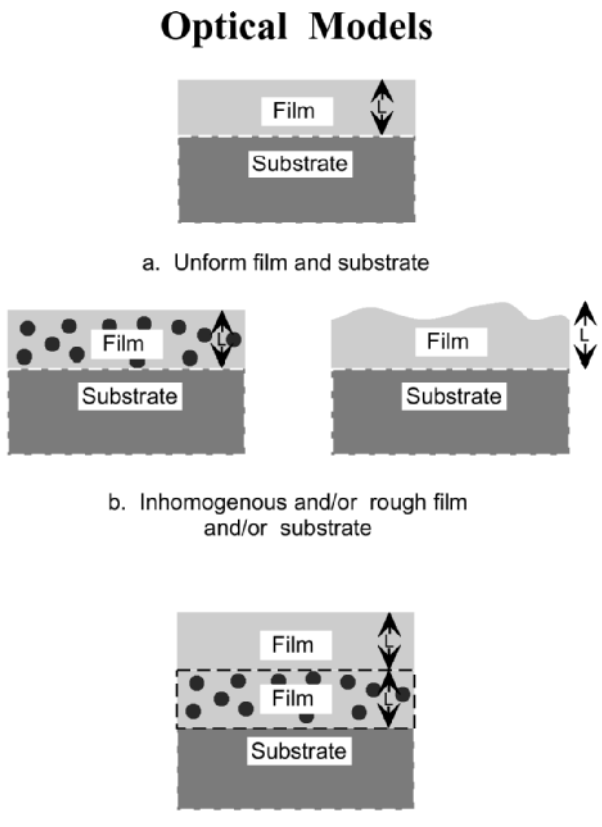


Figure 8.4 Several common optical models for ellipsometric analysis of films on substrates. The number of unknown parameters associated with each model is also included (F for film and S for substrate).

between the SiO_2 film and the Si surface is considered to be composed of both a chemical and a physical interface that add parameters that need to be determined, and this will also be discussed in detail later. It can be said now that these complications present a formidable challenge to obtaining reliable indexes and film thickness for thin films and represent a frontier area for ellipsometry and thin film metrology in general.

The oxidation of Si is arguably the single most important method of preparing SiO_2 films. This chemical process appears to be simple, but it is not. To form a molecule of SiO_2 from the reaction of Si with oxygen, a molar volume change of 120% or a factor of 2.2 takes place. This means that the volume of SiO_2 formed from Si is 2.2 times the starting volume of Si from which it is derived. Such a large volume change upon oxidation (called the Pillings-Bedworth ratio in metallurgy) typically causes large stresses resulting in the oxide peeling away from the substrate. Peeling of the film occurs when the stresses exceed the adhesion forces or the mechanical strength of the film material. Peeling does not happen for SiO_2 films grown on Si. Significant stresses have been measured during the oxidation of Si, but the stresses are smaller than anticipated based on the volume change. To explain these observations, a viscous flow model for Si oxidation has been formulated[10] and the details will follow. Ellipsometry has been found useful in determining the stress effects, since the refractive index and the film density are altered by stress[11,12] and these measurements will be discussed below.

The oxidation of Si in dry O_2 is the preferred method to prepare the highest quality SiO_2 films on Si for microelectronics applications. Even trace amounts (25 ppm) of water in the O_2 have been found to measurably influence the kinetics of film formation and water related charge trapping in the oxide.[13,14] The advantage of having water in the oxidation ambient is that water is a more virulent oxidant and is therefore useful when growing thick films of SiO_2 in that the rate of film formation is increased. The kinetics of SiO_2 film formation both using dry and water containing O_2 have been extensively studied[15–19] using *in situ* real time ellipsometry and some of the details will be shown below.

8.2 Historical Perspective – Prior to 1970

In this section the simultaneous and cooperative evolution of the fields of ellipsometry and microelectronics after 1960 is considered. It is clear that SiO_2 played a significant role in the development of both fields. There is no attempt in this chapter to present an exhaustive historical perspective of either ellipsometry or SiO_2. Rather, the intention is to point out significant papers and the key ideas in those papers that in the authors' opinion simultaneously shaped the evolution of ellipsometry and SiO_2.

The first set of papers that are significant are by Archer[20–22] with the 1957 paper[20] as the most significant, since it paved the way for research that led to the many innovations and improvements that occurred in the following ten years. This early study dealt with the oxidation of both Si and Ge in room air and temperature. The film thickness values were

obtained from ellipsometry measurements made using 546.1 nm light at two angles of incidence (61.26° and 70.00°). Archer characterized the film thickness at various times using the above mentioned three phase optical model (Fig. 8.4a) and from the results made judgments about the mechanism of oxidation. In a 1962 paper[21] Archer extended the analysis of oxidation data using exact relationships between ellipsometric measureables (Δ, Ψ) and optical properties including film thickness and the real part of the refractive index. For this study a digital computer was used for the data analysis. In addition Δ, Ψ maps were calculated and experimental data were compared to the calculations. In a parallel effort, McCrackin and colleagues[23,24] at the then National Bureau of Standards (now NIST) published similar papers on the exact solutions and on the use of computer algorithms. This effort led to a popular Fortran program[25] that was available at that time to the public. The early papers of Archer and McCrackin used single wavelength polarizer-compensator-sample-analyzer (PCSA) ellipsometers and assumed ideal compensators. In 1967 Archer and Shank[26] pointed out the importance of considering nonideal compensators and performed compensator calibration on nonideal compensators. These ideas were incorporated in the 1969 McCrackin Fortran program. So by 1969 and largely attributed to Archer and McCrackin and their colleagues, procedures and understanding were in place so that ellipsometry could be performed accurately and precisely and the results could be analyzed using exact optical relationships to obtain film and substrate properties and film thickness. The fact that Archer, who was at Bell Laboratories, performed measurements on Si and Ge with potential applications to the emerging electronics industry and thereby provided seminal studies linking ellipsometry and SiO_2 was probably no accident. At that time there was considerable effort expended at Bell Laboratories with the development of transistors using Si and Ge and the staff was no doubt eager to make contributions to the field.

Emerging in the early 1960s and parallel with the application of ellipsometry for studying SiO_2 films were a number of pioneering studies about the formation and properties of SiO_2 films prepared via oxidation of Si.[27–31] None of these studies used ellipsometry. Rather most of the studies used optical interference methods. However, these studies were important because several of the key features about the oxidation of Si to produce high quality films of SiO_2 were revealed. Also, the understanding gained in these early studies that treated mostly thicker films provided the basis for subsequent and more accurate studies using ellipsometry and extended the understanding about Si oxidation to thin SiO_2 films. These early studies agree in the broad picture of Si oxidation. Two studies provide the best summaries of the understanding developed in that era with one study[28] by Deal and Grove (the same A. Grove who later was a

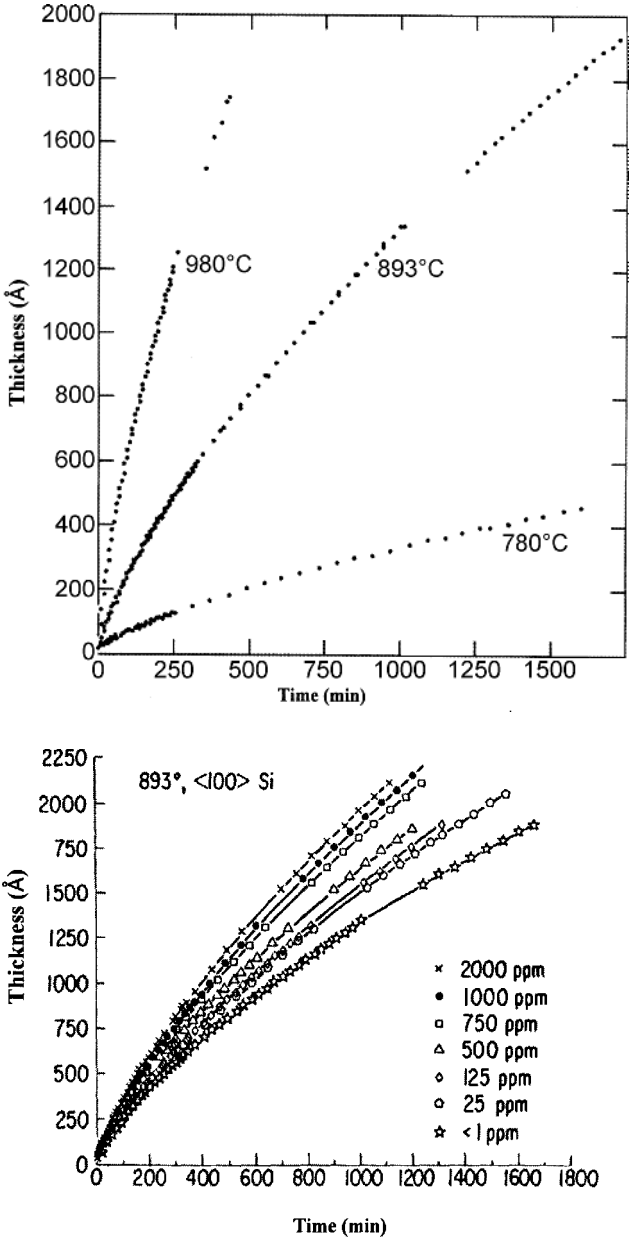


Figure 8.5 Data for the thermal oxidation of Si (100): a) in dry 1 atm O_2 at three temperatures in terms of the SiO_2 film thickness (L) versus the time (t) for oxidation; b) at one temperature but with various amounts of H_2O added to the O_2. The data was obtained using *in situ* real time ellipsometry at a wavelength of 632.8 nm and 70.00° angle of incidence, and the data was analyzed using a single film optical model.

founder of INTEL) often quoted relative to the model for Si oxidation in which the thermal oxidation of single crystal <100> and <111> Si was treated using both dry O_2 and steam. The resulting Si oxidation model is often called the linear-parabolic (LP) model. The other study[27] by Ligenza is concerned with the effects of the single crystal Si substrate crystallographic orientation on the Si oxidation kinetics. We discuss these topics separately and in some depth below, because when ellipsometry was introduced as an effective tool for thin film measurements these early studies were redone using ellipsometry and new insights were revealed that are best understood in the context of these early studies.

Before proposing an analytical model that describes the time evolution of the SiO_2 film growth on Si, it is useful to consider the overall behavior of the growth process as the early workers had done. An overall picture can be obtained from Fig. 8.5a that displays SiO_2 film thickness versus oxidation time data for several Si oxidation temperatures performed in dry oxygen. Figure 8.5b displays data from a systematic study of H_2O additions to oxygen and the effects on the Si oxidation kinetics. This data was generated using *in situ* real-time ellipsometry from the author's laboratory[15,16] and more details about this data will be discussed later. For now it is clear from the shape of the data that the relationship between the SiO_2 film thickness and the oxidation time is neither purely linear nor purely parabolic. If the rate of formation of SiO_2 was invariant with film growth, then linear growth kinetics is expected. On the other hand if the growth were limited by diffusional transport of reactant(s) through the growing oxide film, then purely parabolic growth would be expected. Furthermore, there have been several studies that indicate that during the oxidation of Si, the oxidant species is most likely molecular oxygen (O_2) that migrates inward through the growing oxide film to the Si surface where the oxidation proceeds.[32,33] The most conclusive study[32] first used naturally occurring isotope mixtures of O_2 which is mostly O^{16} for the oxidation of Si. Then after growing the SiO_2 film for some time the ambient was switched to O^{18} enriched O_2. Upon further oxidation in the O^{18} enriched O_2, the workers found that new oxide (SiO_2^{18}) was found at the Si-SiO_2 interface rather than at the SiO_2 surface. This clearly indicated that the oxidant species is migrating inward to the interface where it reacts with Si. Based on the shape of the data which indicates that transport alone cannot explain the oxidation data and that the oxidant species, primarily uncharged O_2, are actually migrating, a very successful oxidation model was derived in the 1960s. Figure 8.6 depicts the essential features of this so-called linear parabolic model and shows that essentially two fluxes are considered. This representation is a departure from the original derivation wherein three fluxes were considered. The flux of oxidant from the gas to the SiO_2 surface is ignored, because this gaseous flux is

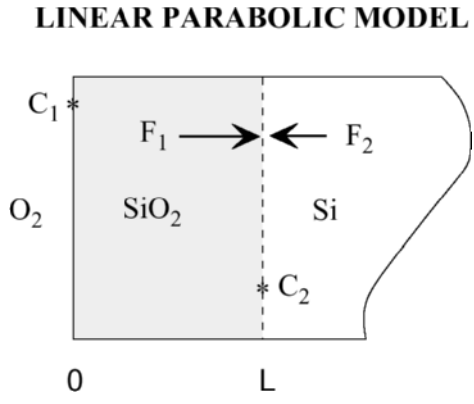


Figure 8.6 Schematic representation of the Si oxidation model showing the transport flux as F_1 and the reaction flux as F_2. C_1 and C_2 are the concentrations of oxidant at the outer and inner interfaces, respectively and L is the SiO_2 film thickness.

generally fast relative to the other fluxes in the solid phases and therefore not kinetically significant in a series process. The flux of oxidant across the growing oxide film, F_1, is given by Fick's first law as:

$$F_1 = D\frac{dC}{dx} \tag{8.6}$$

where D is the oxidant species diffusivity, and dC/dx is the spatial concentration gradient of oxidant in one dimension. For a steady state situation the gradient can be approximated as:

$$\frac{dC}{dx} \approx \frac{C_1 - C_2}{L} \tag{8.7}$$

where C_1 is the solubility of oxidant in SiO_2 and C_2 is the smaller oxidant concentration at the Si-SiO_2 interface where oxidant is removed by reaction, and L is the SiO_2 film thickness. The other flux, F_2, is the number of oxide molecules forming per cm^2 s and is expressed kinetically by a first-order chemical dependence on the oxidant concentration at the Si-SiO_2 interface, C_2:

$$F_2 = k_2 C_2 \tag{8.8}$$

where k_2 is the surface reaction rate constant. Of course the oxidation reaction involves both the oxygen and Si concentrations. However, for a

given Si orientation the Si atom concentration per unit area is constant and thus included in k_2. This Si orientation issue will be discussed in the following paragraphs and then in the context of more modern studies where the careful application of ellipsometry has enabled further resolution of what are ostensibly a confusing set of results. By imposing the condition of steady state an equivalence of the fluxes follows as:

$$F = F_1 = F_2 \tag{8.9}$$

which means that the series fluxes are self-regulatory and permitted the use above of Fick's first law. A rate equation is then formulated in terms of the rate of formation of the oxide film:

$$\frac{dL}{dt} = \Omega F \tag{8.10}$$

where Ω is the conversion constant for from O_2 flux to SiO_2 solid film, i.e., the number of oxidant molecules incorporated into a unit volume of SiO_2 solid (2.3×10^{22} cm^{-3}). C_2 can be eliminated and the rate equation rewritten as:

$$\frac{1}{\Omega}\frac{dL}{dt} = \frac{k_2 D C_1}{k_2 L + D} \tag{8.11}$$

This equation is readily integrated to yield:

$$AL^2 + BL = t + \text{constant} \tag{8.12}$$

where A and B are the reciprocals of the parabolic, k_p, and linear, k_l, rate constants, respectively. The integration constant is evaluated at some initial time and thickness $t = t_o$, $L = L_o$, so as to be able to shift the coordinate axes to any position in L,t space. With this boundary condition, we obtain the following:

$$t - t_0 = \frac{L^2 - L_0^2}{k_p} + \frac{L - L_0}{k_l} \tag{8.13}$$

The use of the region L_o, t_o in L,t space is twofold: at $t_o = 0$, L_o represents an initial oxide thickness either as a native oxide or from some previous processing; and L_o, t_o can be used to define a region of oxidation that does not conform to linear-parabolic kinetics, i.e., an offset to the model. For oxidation in dry O_2 such a region has been identified extending to several tens of nm. Later this important initial oxidation regime will be discussed

further as it has been the subject of a number of recent oxidation models and while in the 1960s through the 1990s the initial growth regime has been more of an artifact than technologically important, now it is of utmost importance in microelectronics. Of particular relevance here is that the initial regime of oxidation comprises SiO_2 from 0 to 20 nm thick. This regime of film thickness is difficult to determine using ellipsometry because n is difficult to measure, as was discussed above and n has been found to vary with thickness. Special procedures involving several techniques including ellipsometry have been used to determine n for thin films and will be presented later in this chapter.

Experimentally obtained oxidation data are often interpreted using the L-P oxidation model. The linear, k_l, and parabolic, k_p, rate constants are obtained from curve fitting routines and the variation of the parameters with process variables are explained using the ideas contained within the L-P model.

Also of significance in the 1960s era was a report by Ligenza[27] that the kinetics of the oxidation of Si was significantly dependent on the Si substrate orientation. Specifically, Ligenza had shown that for high pressure steam oxidation the Si<111> orientation displayed the fastest oxidation with the Si<110> next and followed by the Si<100> surface. According to the LP model this might be interpreted in terms of the number of available Si atoms at the various surfaces. However, as is seen in Table 8-2 the Si<110> orientation has the greatest area density of Si atoms for the three major Si orientations, so a more convoluted model was constructed that depended on the number of available bonds at the Si surface. It has not been made clear why the author chose bonds over atoms as being kinetically significant. As was mentioned above, the original LP model contains the orientation dependence implicitly in the linear rate constant, k_l (or k_2), since the substrate surface ought to dominate

Table 8-2. Si Surface Atom Densities.

Si Orientation	Density Assuming Planar Surfaces 10^{14} (atoms/cm^2)	Density Based on Vicinal Planes 10^{14} (atoms/cm^2)
<110>	9.59	
<111>	7.83	
<311>	10.86	7.21
<511>	11.48	7.05
<100>	6.78	

the surface reaction. Below when more modern ellipsometry results are presented for the oxidation of various Si orientations, Table 8-2 will be discussed in more detail and the need for a convoluted model to explain the orientation behavior will be obviated.

8.3 Modern Studies – Since 1970

From the 1970s to the present, both industrial and academic microelectronics science and technology have been dominated by Si technology. Si technology involves numerous films on single crystal Si substrates (see Fig. 8.1) that in combination yield metal oxide semiconductor field effect transistors (MOSFETs) and the associated circuitry that are included in modern integrated circuits (ICs). In order to control the electronic device properties for the myriad of electronic devices on an IC, process control for the manufacturing steps was recognized as the strategy to follow. Processes to produce devices and arrays of devices called IC chips are essentially a series of individual process steps many of which involve thin films. The steps involve surface preparation and characterization, film formation, film characterization, film removal, and device feature development that is done by lithography. In this chapter the focus is on films and particularly SiO_2 films that are of utmost relevance to microelectronics. Section 8.3 commences with a discussion of the applications and impact of ellipsometry to thick films in 8.3.1. By thick films, it is herein meant films of 20 nm and thicker, and thin films that will be treated in 8.3.2. will encompass SiO_2 films thinner than 20 nm. While there will be considerable overlap in the Thick Films and Thin Films sections, the demarcation made here albeit artificial will enable a focus on different issues that are of interest to the overlap of the fields of microelectronics and ellipsometry.

8.3.1 Thick SiO_2 Films

Silicon technology through the 1970s and 1980s was dominated with films thicker than 20 nm. In the earliest studies, single wavelength ellipsometry (SWE) was used extensively for film thickness measurement. In the late 1970s spectroscopic ellipsometry (SE) was shown to be readily implemented[34] and superior for most applications. Nonetheless, the question of whether to use SE or SWE for a specific microelectronics problem was and is today a valid one. SE will always yield more information, but the measurement can be relatively time consuming, more complicated, and expensive. The operational answer to the question of which ellipsometry

technique to use lies in equipment limitations, principally speed and accuracy of the hardware, and most importantly, a consideration of what is the desired result from the measurement. Speed is important when one desires to follow film formation processes *in situ* and during the processing. For example, a spectral scan using a monochromator drive for SE in the visible range (200–800 nm) could take 10 min or longer. However, with recent developments from Collins' laboratory (see Collins chapter) using optical multichannel analyzers (OMAs), accurate SE scans can be made in the order of seconds and even less, and if some precision can be sacrificed acquisition time of milliseconds is attainable. However, our experience to date is that once a particular process with the materials system is reasonably characterized in terms of the constituents, optical properties and optical models, typically using SE and a variety of other *in* and *ex situ* techniques, SWE can often be employed to yield sufficient process monitoring information during the process. This may not be the case with more complex future processes and complex materials, but many of the examples to follow from our laboratory on SiO_2 film issues are of the type where SWE can be used effectively. Presently, SE is a key technique in our laboratory that is used to characterize the system under study, and SE is performed *in situ* and during processes, i.e., in real-time. SE is used to determine process variables, specifically the best optical models from which to interpret ellipsometric data. In the examples that follow, we use SE and SWE in concert as the main techniques. However, a wide variety of surface analytical and microscopy techniques is required to establish appropriate optical models. It should be mentioned that the use of multiple angles of incidence also yields another method to enhance the accuracy of ellipsometric results and when combined with multiple wavelengths the acronym VASE for variable angle spectroscopic ellipsometry is often used. However, the VASE technique has limited application for *in situ* studies due to the fixed geometry restrictions of typical process chambers.

8.3.1.1 Optical Models

In this section the reasoning that underlies the selection of appropriate SiO_2 optical models is discussed. The first issue to arise for an ellipsometric measurement of a film on a surface, SiO_2 on Si for example, is to determine which optical model best represents the situation, and thus which model to use for the analysis of the Δ, Ψ data. For the case of thick SiO_2 films on Si, the matter of the correct optical model is a relatively easy one to resolve. SiO_2 is known to be a stable and stoichiometric compound. Thus for thick films an appropriate and adequate model is a single film or three phase model as shown in Fig. 8.4a. This is the case because

there is extensive literature on the formation of SiO_2 that demonstrates that the Si single crystal surface is planar and smooth on the 0.1 nm scale and the oxide grows conformally and thus is as smooth as the substrate. There is also the assumption that the interface between the Si and SiO_2 is sharp. For films greater than about 20 nm this is a good practical assumption. For thin films this assumption cannot be made without serious error, and this interface and its characteristics are the subject of considerable ongoing research. However, it is rather clear that there is an interface region of arguable dimension and composition that needs to be accounted for in thin film optical models and this will be discussed in Section 8.3.2. on the subject of thin films, because for thin films the incorporation of an interface in the optical model is important. Figure 8.5 shows the results of the use of *in situ* SWE during the thermal oxidation of Si for both dry O_2 (8.5a) and wet O_2 oxidation (8.5b) with a simple three phase model. It is clear from these results that the kinetics of oxidation as was discussed above can be determined.

Typically what is done to evaluate this kind of film growth or deposition experiment is to first determine the optical model to be tried, and then to collect from the literature and/or determine the optical properties of each component in the model in the cases where acceptable properties are not readily available. For the simple case of Si oxidation which is called film growth, the reaction constituents are the O_2 gas and the Si substrate, and a film of these constituents grows on the Si surface. As was briefly discussed above, this film growth process is contrasted with film deposition where Si and O containing gases are flowed into a reactor, mixed and reacted to form the desired SiO_2 product that condenses upon a Si or any other surface. For analysis in terms of a three phase model, there are five *a priori* unknowns excluding the optical properties of the ambient: the complex refractive index for Si substrate (n_s, k_s), the same for the SiO_2 film (n_f, k_f) and the film thickness (L_f).

For SWE there are two quantities measured: Δ and Ψ. With knowledge of the experimental conditions such as λ and ϕ it remains impossible to obtain the five material unknowns with high certainty. However, all is not lost for this particular situation. The Si substrate has been thoroughly characterized and ample databases[5,6] are available, even at elevated temperatures[35,36] which would be needed for *in situ* real-time SWE Si oxidation process measurements. Similarly one can find databases for SiO_2 films as well.[5] It has been determined that the optical properties for SiO_2 at high temperatures are very weak functions of temperature.[35] Thus the room temperature SiO_2 database need not be corrected unless extreme accuracy is required. Also SiO_2 with about a 9 eV band gap is transparent in the visible and so $k_f = 0$. This leaves only L_f to be determined in this problem, and of course it is readily obtained from an inversion of the

ellipsometric equations. It should be remembered that *in situ* real-time SWE experiments are typically performed to follow the time evolution of the materials in a process. Thus the differences measured during a process are usually more important in elucidating film formation kinetics than are the absolute values. If accurate absolute values are required, then the system can be brought to room temperature for careful measurement and/or measurements can be made using films on previously characterized substrates to determine the temperature dependence of the film optical properties.

Suppose that a different and unknown film is being grown or deposited onto Si or another known substrate. Then there would be three unknowns for the film: n_f, k_f, and L_f. In this case, SWE cannot determine the three unknowns accurately from the two measurables. If the film is transparent, then there are two unknowns that can in principle be determined from the two measurables. At this point one needs to reconsider a topic discussed above in Section 8.1.3, namely that the sensitivity of ellipsometry is not uniform throughout Δ, Ψ space. With reference to Fig. 8.3, it is now possible to analyze what is meant by "in principle" both n_f and L_f can be determined from a single SWE measurement for a transparent film on a known substrate. Figure 8.3 shows that near the origin, i.e., where the film thickness is small, the Δ, Ψ trajectories for the transparent films with different n's are nearly the same while the trajectories diverge for larger film thickness and are maximally different at half an ellipsometric period ($P/2$). Therefore, ellipsometry has maximum sensitivity to determine n_f near $P/2$. For SiO_2 films using He-Ne red light (633 nm) and for $\phi = 70.00°$, $P/2 \cong 144$ nm. This means that it would be best to try to grow or deposit films near P/2; for an unknown material that may appear to be a difficult task. However, film growers can usually control their process sufficiently and obtain approximate film thickness from other techniques such as transmission electron microscopy, crystal rate monitors, profilometers, etc. Therefore, it is usually possible to fabricate a sampling of film thickness on known substrates for ellipsometric evaluation. Once preliminary ellipsometric evaluation is performed then better guesses about near half period thickness can be made. Of course this implies that n_f does not significantly vary with thickness. This is approximately true for many thick films and it has been verified for thick SiO_2 films prepared by thermal oxidation of Si. However, as we shall see later this is definitely not true for thin films and it has been verified not to be true for thin SiO_2 films. The fact that microelectronics is moving toward ultra thin films of SiO_2 less than 5 nm thick for the gate dielectric means that special problems arise for accurate ellipsometric measurement and these problems and some solutions will be discussed below in the Thin Films section. One suggestion here that we routinely employ is to prepare simulations of Δ, Ψ

versus film thickness space for the near half period thick film n_f. Then as more n_f data from various thickness are accumulated these data are added to the simulated figure. From the relationship of these data to the simulated trajectories, one can begin to envisage the variation of n_f with film thickness.

SE measurements on thick films afford considerably more latitude in obtaining the unknowns from the optical model. Even without VASE, SE enables more sophisticated regression analysis. This kind of analysis where experimental data and model calculated data are compared and the model parameters are obtained that minimize the differences has been treated in some detail in earlier chapters. Therefore, only a brief treatment is given here in relation to SiO_2 films. For a single pure homogenous material surface, SE measures the dispersion in $\Delta(\lambda)$, $\Psi(\lambda)$ from which $\varepsilon(\lambda)$ is obtained. From the equations above for ε, it is seen that ε is related to k, and the absorption coefficient α is given as a function of k:

$$\alpha = \frac{4\pi k}{\lambda_0} \tag{8.14}$$

where λ_o is the wavelength in vacuum. Essentially SE accesses the absorption spectrum, $\alpha(\lambda)$, in the wavelength range scanned. Ellipsometry not only yields the dispersion in n, $n(\lambda)$, but also in k, $k(\lambda)$, without recourse to the optical sum rules, the so-called Kramers-Kronig relationships which enable conversion from ε_1 to ε_2 and vice versa, but are approximate because they are integrals over infinite wavelengths and an infinite λ range is not experimentally covered. Therefore, SE is very useful in determining optical property databases on new materials or on materials changes with processing. Having the spectrum of optical properties allows an investigator to choose spectral regions or even single wavelengths that optimize a particular measurement. For example it has been found[37] that certain wavelength regions of the spectrum for semiconductors are more sensitive to thermal changes. Figure 8.7 shows the temperature (T) dependence of the dielectric functions for some semiconductors used in our research. Thus, the more sensitive regions can be selected to determine temperature while less sensitive regions can be used when small changes in temperature cannot be avoided or readily accounted for. For example in plasma processing the temperature of the sample will rise due to exposure to the plasma.[38,39] This rise in temperature is both uncontrollable and unavoidable. Therefore, if the temperature rise is unknown or impossible to account for, it would be judicious to find a spectral region for SE that is relatively insensitive to temperature. For example for Si at 3.65 eV there is very low T sensitivity while at 3 eV there is high T sensitivity. Attention to these aspects of the SE spectra has been made in

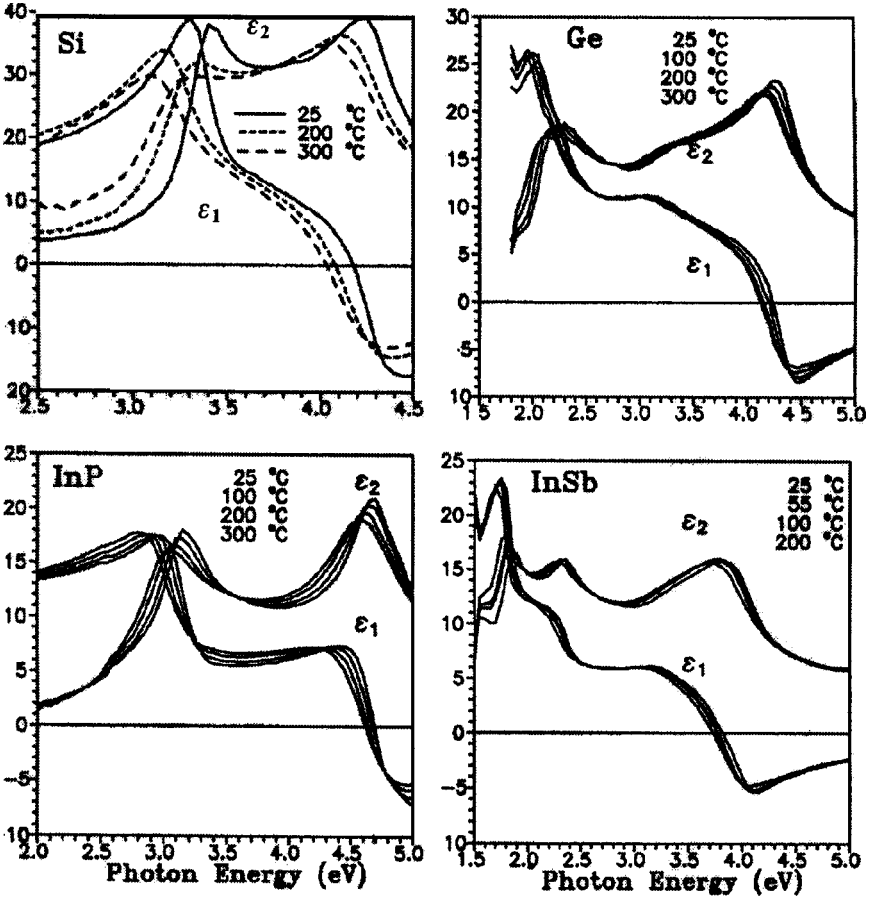


Figure 8.7 Dielectric function (ε_1 and ε_2) spectra as a function of temperature for several semiconductor surfaces.

previous reviews.[37,40] A last point about spectral regions should be mentioned. While many investigators intentionally or inadvertently put forth the seductive argument that the inclusion of more λ's and more ϕ's improves accuracy in a regression analysis, this line of argument can be misleading. If one knows *a priori* that a spectral region is sensitive to uncontrolled process changes and/or that some ϕ's are less sensitive to polarization changes, then to include these knowingly inaccurate data can only make the results less accurate irrespective of the quantity of data. When commencing the measurement of a materials system, it is best to perform computer simulations using known and guessed optical properties, in order to learn about the optimum experimental conditions. Then the

actual experimental measurements should be performed at as close to optimum conditions as is possible.

For other cases such as homogenous films on a pure substrate, several films, inhomogeneous materials, etc., ellipsometry obtains Δ, Ψ at each λ as for the simple pure material discussed above. However, from this measurement one does not directly obtain ε. Rather, from the direct application of the ellipsometry formulas one obtains the pseudo dielectric function, $<\varepsilon>$. In essence $<\varepsilon>$ is simply the value obtained when one analyzes the ellipsometric measurables for a complex system. In order to obtain ε or L for a specific pure homogenous film in the complex system, the optical model for the system must be known. Figure 8.4 shows some optical models for various systems. The complexities and procedures for extracting unknowns from the optical models have been treated in depth in previous chapters, but some specific issues related to SiO_2 films will be treated below.

Returning to the specific subject of SiO_2 films, it is useful to first look at the SE spectrum of pure Si itself shown in Fig. 8.8 in the visible spectral region for both crystalline and amorphous Si. As discussed above, the ε_2 spectrum is similar to the absorption spectrum for a pure homogenous material. The characteristic peaks in the c-Si ε_2 spectrum at 3.4 and 4.3 eV correspond to the interband transitions for Si, viz. the direct gap energies,

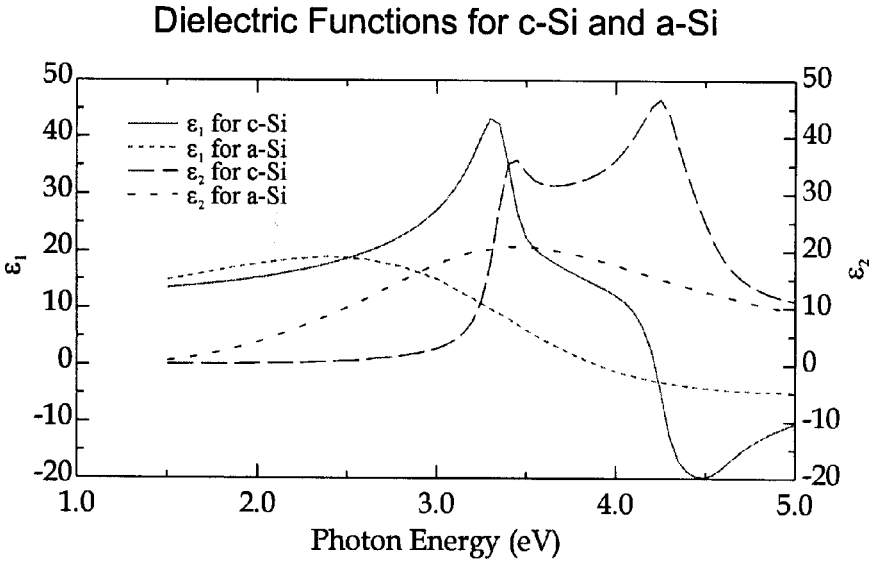


Figure 8.8 Dielectric function ε in terms of ε_1 the real part and ε_2 the imaginary part for both amorphous and crystalline Si, a-Si and c-Si.

in the spectral range covered and the peak position, magnitude, and shape are all material specific and can even reflect process differences which give rise to differences in structure or morphology which affect ε_2. This is seen dramatically in the contrasting relatively broad featureless spectra of ε_2 for a-Si. These optical differences for a material with the same atoms and with nearly identical bonding and short-range order are therefore attributable to the loss of long-range order and its effect on the electron energy band structure and optical absorption. Even for these chemically similar materials there are distinct differences that are easily detected and distinguished using SE. As was previously reported,[37,41] damage to the Si surface resulting from ion beam and/or plasma exposure causes amorphization and hence a mixture of c-Si and a-Si that can be modeled using the SE spectra for a mixture of c-Si and a-Si using effective medium approximations (EMA). Such a mixed film can be modeled as is shown in Fig. 8.4b. Later in the Thin Films section the c-Si and a-Si spectra will be revisited. Now however, the relevant point is that when, for example, a film of optically transparent SiO_2 is deposited or grown on Si, the Si spectrum changes, since the resultant spectrum is a composite of film and substrate as can be indicated by $<\varepsilon(\lambda)>$. Figure 8.9 shows a c-Si

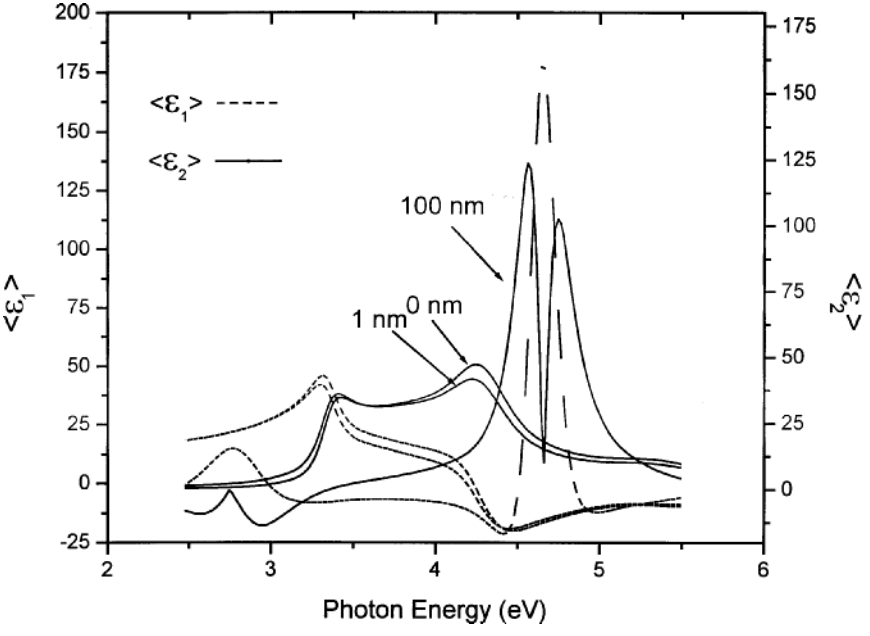


Figure 8.9 SE spectrum in terms of ε_1 and ε_2 for a single crystal Si (c-Si) substrate (0 nm) and c-Si with a thin (1 nm) and a thick (100 nm) SiO_2 film.

spectrum with a thin (1 nm) and thick (100 nm) SiO_2 film and bare c-Si for comparison. With a thin film the c-Si spectral features are clearly visible albeit attenuated as compared with the pure c-Si spectrum. However, c-Si with a thicker SiO_2 film does not show the pronounced spectral features of c-Si. Rather, the composite SE spectrum shows little resemblance to c-Si, but rather it is dominated by the optical response of the film and the optical interference obtained there from.

8.3.1.2 SiO_2 Film Growth Mechanisms

The thermal oxidation of Si provides an excellent example of the power of ellipsometry in addressing microelectronics issues and represents one of the first examples of real-time monitoring of a microelectronics industrial process, viz. Si oxidation in 1 atm O_2.[15,42] In the field of microelectronics it is usually desirable to measure various dielectric film thickness, and the SiO_2 film thickness, L, and index of refraction, n, resulting from Si oxidation. Furthermore, it is necessary to have sufficient process control to obtain specified film thickness to within several percent. As discussed above SWE measurement of Δ, Ψ enables the extraction of both desired parameters L and n for a single transparent film on a known substrate. Spectroscopic ellipsometry over specifies this problem. Using a well aligned ellipsometer and near $P/2$, n and L are obtained well within a few percent accuracy.

Figure 8.5a shows the results from a 1 atm dry thermal O_2 oxidation study of Si[15] using SWE, which was performed in flowing O_2 . From the shape of the thickness versus time data it was possible to deduce and confirm oxidation models for the crucially important SiO_2 film growth process. For example the data were compared to the linear-parabolic model and values for k_p and k_l and L_o were obtained. From k_p and k_l values versus temperature, activation energies were also obtained. Thus the oxidation process could be quantitatively described and reproduced using the model parameters from the optical model and the Si oxidation model. Furthermore, better accuracy than required could be obtained from the combination of models using ellipsometry for films thicker than 20 nm. Figure 8.10 displays data from another *in situ* real-time SWE study,[16] and which includes both dry and wet (H_2O added) O_2 oxidation of Si (from Fig. 8.5), and further illustrates the great power of in-situ SWE for detailed and intricate studies of the oxidation process. The upper and lower solid lines act as bounds and show results from H_2O in O_2 (2000 ppm H_2O in O_2 and labeled wet) and pure O_2 (dry) oxidation processes, respectively. These data were obtained in separate experimental runs. For the experiment shown in this figure an initially dry O_2 oxidation ambient was

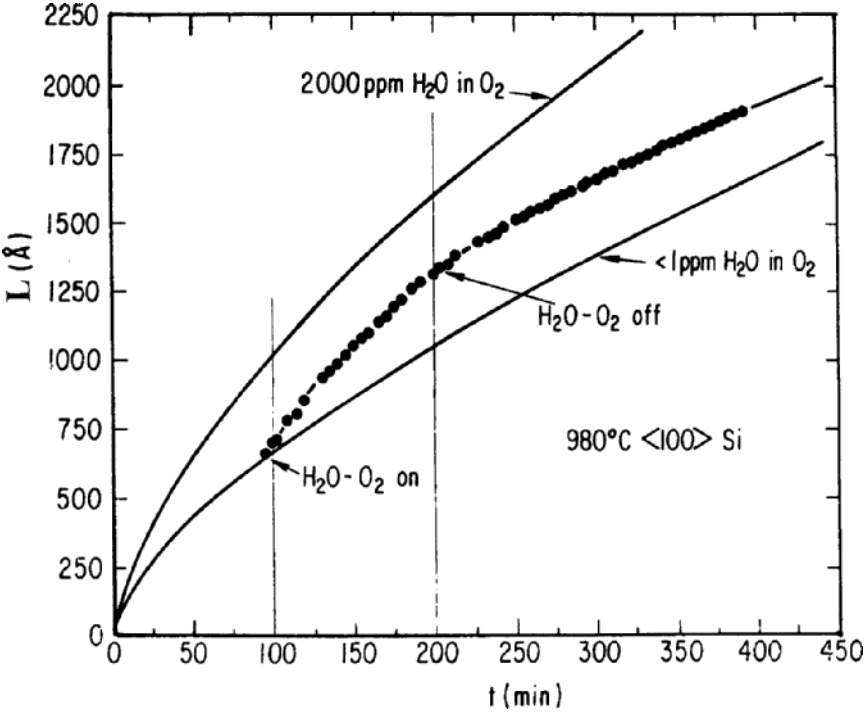


Figure 8.10 Results from in-situ SWE employed during the Si oxidation process including dry O_2 results (<1 ppm H_2O in O_2) and wet (2000 ppm H_2O in O_2) as boundaries. The filled circles represent an experiment that started with dry O_2 and switched to the wet ambient at 100 min. The data was obtained using 632.8 nm light and 70.00° angle of incidence, and was analyzed using a single film optical model.

switched at some arbitrary time (100 min) to a mixture of O_2 with H_2O (2000 ppm), whereupon the Si oxidation kinetics also abruptly changes to that characteristic of wet grown SiO_2. The analysis of these SWE experiments, using a single film or three phase model, enabled the role of trace amounts of H_2O on the SiO_2 film formation to be quantitatively assessed. It was found that both the linear and parabolic rate constants, k_l and k_p, increased with the H_2O content in the O_2. However k_p increased a greater percentage than did k_l with H_2O additions. This was explained considering that H_2O reacts with the SiO_2 network yielding OH groups and thus the reaction effectively breaks up the network of Si-O-Si bonds. When this occurs the diffusivity of O_2 and many other atoms and molecules is increased thereby increasing the rate of transport and increasing the overall oxidation rate through k_p. Also H_2O is a more virulent oxidant for Si and that explains that with H_2O addition even the interface reaction

increases as evidenced by an increase in k_1. The specific real-time experiment in Fig. 8.10 shows that almost immediately upon H_2O addition the reaction rate changes to that for H_2O in O_2. This indicates both a rapid reaction of H_2O with the SiO_2 network and a rapid change in kinetics with small amounts of H_2O in O_2. These kinds of experiments would be ambiguous if performed *ex situ* with exposure of the samples to atmospheric conditions.

8.3.1.3 Plasma Oxidation of Si

We have constructed ellipsometric hardware integrated with process hardware in order to follow electron cyclotron resonance (ECR) plasma oxidation and chemical vapor deposition of a variety of films (oxides, nitrides and Si) on a variety of semiconductor substrates, viz. Si, Ge, and InP using SE and SWE *in situ* and in real time.[38,39,43–45] The ECR process is important, because it enables the preparation of desired thin film materials at low process temperatures so as to limit solid-state diffusion, and with low ion energies to minimize damage to the semiconductor. Before focusing on the specific studies using ellipsometry a few words should be given to the ellipsometry hardware issues. Since the subject of *in situ* real time is contained in a separate chapter, here only the design philosophy is mentioned in that it relates to obtaining results required by the microelectronics industry and specifically for SiO_2 films. As has been discussed in depth in previous chapters, ellipsometry is a precision optical technique capable of sub 0.1 nm film thickness measurements when the hardware is carefully and precisely aligned and calibrated. Procedures for hardware alignment and calibration are well known. In order to perform precise alignment, ϕ must be altered to the straight-through (180° position). Most commercial ellipsometry hardware for *in situ* measurements is designed to "bolt on" to the process chamber. Process chambers are typically rigid stainless steel chambers. Therefore, using commercial equipment it is usually impossible to alter ϕ and therefore it is impossible to perform the most precise optical alignment. In order to circumvent this alignment problem, other schemes have been developed. For example previously characterized samples can be used to back calculate ϕ which can then be used for other measurements. From these kinds of procedures reasonably good results are obtainable. However, most of our ellipsometry hardware for *in situ* real-time studies was "home-built" and specifically designed including a precision optical bench that was independent of the process chamber and consequently fully alignable.[40] For accurate *in situ* real-time studies, careful attention to alignment and calibration must be given and results with known materials should at the very least be performed to determine the level of alignment and calibration for

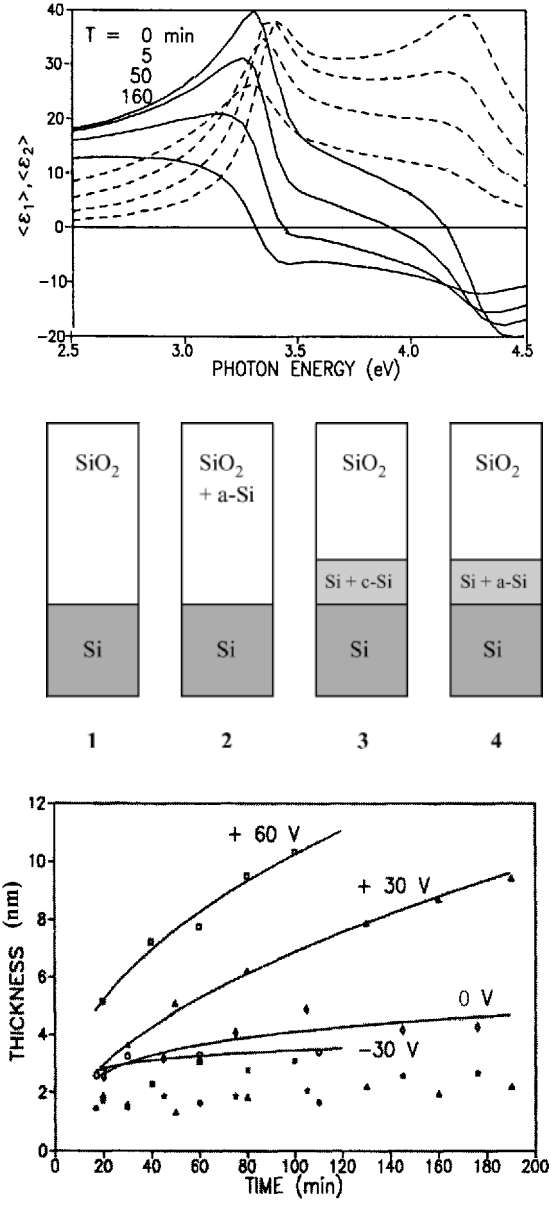


Figure 8.11 Electron cyclotron resonance plasma oxidation of Si results as obtained by *in situ* real time ellipsometry. a) *in situ* SE results as a function of oxidation time with ε_1 the solid line and ε_2 the dashed line; b) four possible optical models; c) results of data analysis according to model 4 for ECR plasma oxidation of Si at four different applied D.C. biases. Open symbols are for the SiO_2 overlayer and shaded symbols are for the interface layer.

the optical system. Rarely are convincing performance studies reported by hardware manufacturers or in ellipsometry papers.

Before the ECR plasma Si oxidation mechanism can be explored using *in situ* real-time ellipsometry, an appropriate optical model is required. Figure 8.11a shows typical SE Si plasma oxidation data[38,39,46] in terms of ε from which the time evolution of the process is observed. The SE spectra were taken upon stopping the process at various times. Figure 8.11b shows the four models considered for analysis of the data. These possible models were arrived at from various analyses of both the oxides and the Si surfaces after plasma oxidation. The quality of the fit of the data to the model was made using an unbiased estimator, δ. In this study, δ for model 4 was about 1/2 of the smallest value for the other sensible models. None of the fitting improved with the addition of voids and for model 4 about 50% a-Si and 50% SiO_2 was obtained for the volume fractions of constituents. Figure 8.11c shows the results calculated from model 4 for several D.C. substrate biases in terms of the thicknesses for the oxide overlayer (open symbols) and interface layers (black symbols). The instantaneous formation and thereafter small change in the interface layer as well as the bias effects on the oxide layer enabled elucidation of the oxidation mechanism. Essentially the Si thermal oxidation model was altered to include a flux of negatively charged oxidant species that were produced during ECR plasma oxidation of Si.[45] The rate enhancement seen with the positive D.C. bias indicates the efficacy of the negatively charged oxidant species.

8.3.1.4 SiO_2 Film Stress

Early studies of SiO_2 film stress[47,48] were performed at room temperature on films grown on Si using high oxidation temperatures greater than 1000° C which was appropriate to the technology at that time. These studies concordantly reported that the measured residual room temperature stress could be completely explained based on the thermal expansion stress, σ_{th}, which develops upon cooling from the oxidation temperature, and as a result of the difference in thermal expansion coefficients, $\Delta\alpha$, between SiO_2 and Si. σ_{th} is proportional to both $\Delta\alpha$ and ΔT as:

$$\sigma_{th} \propto \Delta\alpha \cdot \Delta T \tag{8.15}$$

Since at the oxidation temperature $\Delta T = 0$, the thermal component of the stress, σ_{th}, cannot be implicated in any oxidation models. In the late 1970's a study[49] indicated that for oxidation temperatures below 1000° C an intrinsic stress, σ_i, is observed which increased with decreasing oxidation temperatures. The existence and temperature variation of this intrinsic

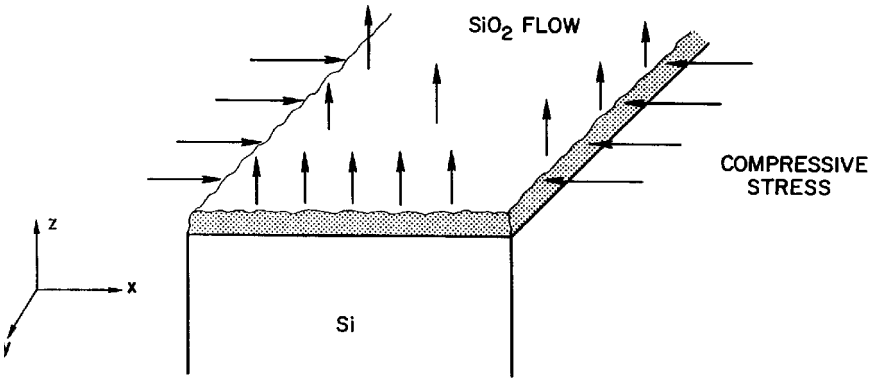


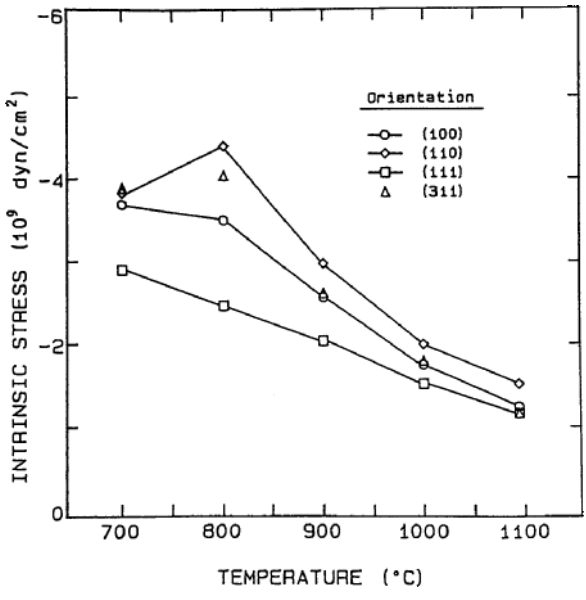
Figure 8.12 Viscous flow model where the growing SiO_2 film viscously relaxes into the free direction (z) driven by a compressive in plane (x,y) stress.

stress, σ_i, was confirmed[50,51] and a model was proposed[50] which not only explained the appearance of the stress, but also the simultaneous appearance of an increased SiO_2 film density with the same increase for lower oxidation temperatures as for σ_i.[50,52] The model called the "viscous flow model"[50] was based on earlier proposed ideas[49,53] and is explained with the use of Fig. 8.12. First as was mentioned above it is easy to confirm that there is a 120% increase in molar volume in converting an atom of Si into a molecule of SiO_2, and this fact establishes the volume requirement, in order for the reaction to proceed. This volume requirement can be met for the SiO_2 formation on the Si surface by the expansion of the as-formed SiO_2 into the free direction, viz. the direction normal to the Si surface. If all of the oxide can find this direction, then the anisotropic expansion of the newly formed SiO_2 film will occur unimpeded and without stress. The viscous flow model assumes that this free direction is "found" by the mechanism of viscous flow. At the high oxidation temperatures, above 1000° C where the oxide viscosity is sufficiently low, the oxide being constrained by adhesion in the plane of the Si surface readily flows into the normal direction. The constraint in the lateral direction and flow in the normal direction can be analogized as the flow of toothpaste from a tube as the tube is compressed normal to the direction of flow. However, at lower oxidation temperatures, the higher oxide viscosity precludes easy flow within the time frame for oxidation, and an intrinsic stress develops. Since the oxide viscosity increases as the temperature decreases, it then follows that the intrinsic stress which develops should also increase with decreasing oxidation temperature, as is observed. Along with the observation of the intrinsic stress and its temperature dependence, is the parallel observation

of an increase of the SiO_2 film density[50,52,54,55] with decreasing oxidation temperature. Using the viscous flow model, the densification of SiO_2 can be understood as the accommodation of the SiO_2 film growth system to the accumulation of stress, viz. the system attains as small a volume as possible so as to minimize the stress. Although the SiO_2 network is quite open, only a small density increase is permitted before large repulsive forces are encountered. Between the oxidation temperatures of 1100° C and 700° C about 3% density increase is observed. The experimental measurement of the density, ρ, for these thin films is worthy of further comment in the context of ellipsometry. The first report of a density increase with lower oxidation temperatures was by Taft,[52] and was based entirely on the precise measurement of the refractive index, n, values as a function of oxidation temperature and the application of the Lorentz-Lorenz relation to convert n to density. The above stated Clausius-Massotti formula relating ε to density can be rewritten in terms of n rather than ε and is called the Lorentz-Lorenz formula. Later Irene et al.[55] found nearly identical refractive indices, but went further and obtained the density directly from measurements of the film volume and mass. While this latter measurement of density is not as precise as the measurement of n, the direct measurement yielded the same temperature dependence as the n derived values and approximately the same absolute values, thereby increasing the confidence with ellipsometrically obtained ρ values. Figure 8.13 shows the measured film stress (8.13a), σ_i, and density (8.13b) as obtained from n measurements and the Lorentz-Lorenz formula, as a function of oxidation temperature. Furthermore, the same density change as a function of oxidation temperature was found using infrared spectroscopy (IR) techniques.[54] The IR spectra for SiO_2 prepared by thermal oxidation at three oxidation temperatures showed a shift towards lower frequency, ν, for the 1075 cm^{-1} band. This band is associated with the Si-O-Si bond angle, Θ, which is the angle between adjacent SiO_4 tetrahedra and is a measure of the Si-Si distance which relates directly to the SiO_2 density. The lower ν the smaller is Θ, and hence the smaller is the Si-Si distance and the higher is the film density. From the IR, a 2 to 3% increase in density is obtained in substantial agreement with the other earlier measurements.

8.3.2 Thin SiO_2 Films

The present microelectronics technology employs SiO_2 gate oxides in MOSFETs that are less than 10 nm thick and seriously contemplated future advanced technologies will require the SiO_2 gate oxide films to be less than 2 nm thick. This film thickness regime is difficult to measure accurately using ellipsometry or any other technique. Furthermore, the assumptions



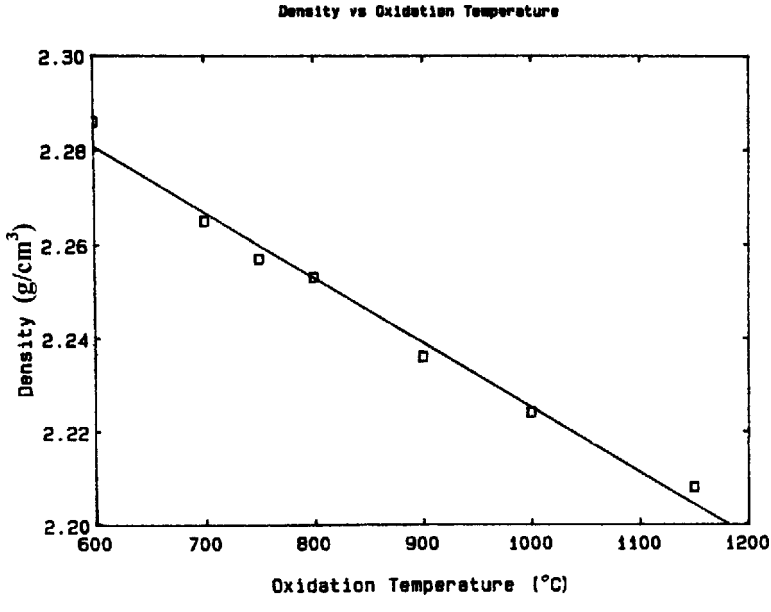


Figure 8.13 a) Intrinsic SiO_2 film stress resulting from the thermal oxidation of various Si substrate orientations as a function of oxidation temperature; b) SiO_2 film density as a function of oxidation temperature as obtained from ellipsometric measurement of n and the Lorentz-Lorenz formula.

made about uniform properties for thick films are known not to be true in this film thickness regime. The oxidation kinetics for this regime has been in dispute since the 1960s (the thickness regime $<L_o$). Also the nature of the Si-SiO_2 interface, arguably the most important interface in the universe, remains in dispute and of crucial importance in modeling thin films. These issues related to thin films of SiO_2 are now addressed as they relate to ellipsometry.

8.3.2.1 The Initial Oxidation Regime

In this section the initial Si oxidation kinetics results are discussed as these results existed up to the mid 1990s. Since that time a new technique has been developed that impacts the measurement of the oxide thickness and when combined with ellipsometry has led to a reevaluation of the initial oxidation kinetics. We will present the technique and the reevaluation studies below in Section 8.3.3.1.

The initial oxidation regime comprises the region of SiO_2 film thickness that does not conform to L-P kinetics, i.e., from zero oxide thickness up to L_o which is about 20 nm. For the purposes of discussion herein, the initial regime is divided into two parts based on film thickness. The first part includes the very initial regime that is from a bare Si surface up to about l nm, and this is followed by the regime of about l nm to 20 nm. The former regime is very important from a fundamental point of view, since it emphasizes the reaction at the Si surface without substantial SiO_2 being present. The latter regime is of particular practical importance because at this time advanced MOSFETs require gate SiO_2 to be less than 5 nm and even below 2 nm for the most advanced technologies. It is interesting to note that most Si oxidation experiments begin with Si samples that already have a native oxide of about 1 nm. However, it is possible to commence an oxidation experiment with a bare Si surface. This can be done using "brute force" by placing a Si sample in an ultra high vacuum (UHV) chamber and heating above 700° C so as to volatilize the oxide layer. Actually, the oxide itself doesn't evaporate appreciably at this temperature, but rather with SiO_2 in contact with Si, the oxide disproportionates to a volatile component, SiO as:

$$SiO_2(s) + Si(s) \rightarrow 2SiO(g)$$

Another less aggressive way to achieve a bare Si surface is to first dip the oxide coated Si surface in HF just before performing the desired reaction at the Si surface. HF will remove the surface oxide and render the Si surface hydrogen terminated and strongly hydrophobic. This H terminated

surface can be stable for hours depending on the ambient conditions. The H terminated Si surface can be gently heated to several hundred °C to remove the H and leave a bare Si surface. The HF treated H terminated Si surfaces were studied using a combination of *ex situ* and *in situ* spectroscopic ellipsometry.[56] These workers were able to easily follow the changes at the Si surfaces when the monolayer of H was removed. Of course the bare Si surface from either procedure will react rapidly to reform oxide even at UHV. In discussions of oxidation kinetics presented above, the symbol L_o was used to denote the upper limit of the initial regime of several tens of nm, and now L_n will be used to denote the upper limit of the very initial oxidation regime that is essentially the native oxide thickness.

The study of the ultra thin film regime from 0 to L_n is fraught with experimental difficulties such as the cleanliness of the Si surface, and the vacuum conditions during the experiments. One early study comprising surface energy measurements made in room ambient[57] using contact angle techniques showed a steep change in surface energy as the oxide thickness changed from 0 to 3 nm. These experiments strongly suggested that only a change in composition would explain the large observed changes in surface energy. Later Rutherford backscattering measurements on thin oxides appear in agreement,[58] however an ambiguity exists as to whether the etch solution (HF in H_2O) used to etch SiO_2 for the contact angle experiments is altering the surface energy. An alternative interpretation[59] based on image charges assumed that a fixed oxide charge, Q_f, exists near or at the outer (ambient-oxide) interface. This charge could alter the interfacial energy as the atmosphere ($\varepsilon = 1$) is replaced with H_2O ($\varepsilon = 80$). A good agreement to the contact angle data is obtained using charges located 0.3 to 0.6 nm from the surface and charge densities of the order of 10^{13} cm^{-2} were calculated. While interesting this model has not gained wide acceptance.

The usual studies of this regime, 0 to L_n, are made using UHV techniques in which particular attention is paid to the cleanliness of the Si surface. One such study[60] utilizing *in situ*, in the UHV chamber, Auger (AES) and electron energy loss (EELS) spectroscopies, and surface cleaning using high temperatures, reports the disappearance of surface states upon exposure to oxygen as well as the formation of oxide. Initially the oxide appears to be a suboxide in agreement with previous studies.[58,61–63] Another similarly careful UHV study utilizing EELS[64] reports three different stages of Si oxidation. The earliest stage after oxygen exposure at low temperatures (l00 K) involves molecular oxygen that at stage two at higher temperatures converts to atomic species. The final stage for higher exposures indicates the formation of SiO_2. Optical absorption of a UHV cleaved Si crystal[65] also shows that surface electronic states within the Si band gap disappear upon oxygen exposure; these states are associated with

dangling bonds at the Si surface, and as was discussed in Section 8.1.2 are implicated in the electronic passivation of the Si surface. The valence band states are observed to be unaffected. Therefore, these studies affirm the intuitive notion that the Si-SiO_2 interface has some transition region of graded composition and that the formation of a stoichiometric SiO_2 film occurs beyond at least a few atomic layers and at temperatures above room temperature.

The next growth regime from L_n to L_o that extends from about 1 nm to several tens of nm is the regime that is the focus of present microelectronics technology. The reason that technology is now demanding thinner high quality SiO_2 films is a direct result of the economically motivated demand for greater device densities for manufactured chips. In order to achieve greater device densities on a fixed chip area, it is obvious that smaller device areas are required, and along with this but less obvious is that thinner dielectric films are also required. This latter and presently most germane requirement is best understood by considering MOS devices that are designed to operate at certain applied electric fields. The operating fields and the electric field distribution in the gate region are determined in part by the capacitance of the gate oxide in the MOS structure which is given as:

$$C = KA/L \tag{8.16}$$

where K is the static dielectric constant for SiO_2, A is the gate contact area, and L is the dielectric thickness (capacitor plate separation) or in our case SiO_2 film thickness. Thus it is easily seen from this relationship that in order to down scale a device to smaller A but at the same time maintain the designed C, the SiO_2 film thickness, L, must also decrease so as to maintain the C and therefore maintain the device operational characteristics. Present advanced technologies find the required L below L_o or below 20 nm. For the highest quality SiO_2 grown in dry O_2, this thickness is within the offset region of the L-P model, i.e., below L_o, t_o and hence without analytic description. It will be shown below that new results for n for interfacial oxides grown on Si have enabled a further elucidation of the Si oxidation kinetics for the thin oxide regime below L_o, t_o.

In terms of the oxidation data representing this regime, the best data is obtained from in-situ ellipsometric experiments of which there are several published studies.[15,42,66–68] However, before proceeding with the results from these studies it should be noted once again that the measured ellipsometry data were analyzed using a single film model where it is assumed that the interfaces are sharp and that the oxide is homogenous. It is now known that this is not the case, and in a later section corrections will be made. However, it is interesting and instructive to know what kind of results the earlier studies yielded on the initial regime, and then to compare

these results to the corrected ellipsometry results presented below. The data of Hopper et al.[42] shows that the shape of the initial regime up to L_o is basically linear-parabolic, as is the thicker film regime but with different L-P rate constants than those used to describe the thicker film regime. Following the work of Hopper et al., similarly precise work[17,69] revealed that the shape of the L versus t data is more parabolic when the ambient contains more H_2O. This was established based on the fit of the initial regime data to a simple linear equation of the form:

$$t = k_1 L + k_2 \tag{8.17}$$

where k_1 and k_2 are simply the slope and intercept, respectively, and have no other physical significance. The quality of the fit indicates that from 780° to 980° C the dry O_2 data fits best. The parabolic term in the L-P model is derived from the consideration that the formed SiO_2 film actually protects the Si surface from further oxidation, i.e., the grown film provides a barrier to further oxidant penetration. Thus with this idea, the parabolic shape of the oxidation data is interpreted as the films ability to protect the underlying substrate from further oxidation, and one deduces that the wet grown oxides are more protective. The extensive relatively new data of Massoud et al.[17–19] generally agrees with the other data on the appearance of the initially fast oxidation regime. Massoud et al. used exponentials with appropriate damping lengths to precisely fit the data in the initial regime without implied physics.

8.3.2.2 Si Crystal Orientation Effects on Oxidation

One of the most interesting features of the above mentioned Massoud et al. data is the observed Si substrate orientation dependence for the L, t data. The earliest reports of this orientation dependence[70] showed that for high pressure oxidation in a thickness regime well beyond L_o, the Si<111> surface was the fastest oxidizing with the <110> next, followed by the <100> Si surface for the major Si orientations. According to the L-P model the Si orientation dependence derives from the reaction of oxidant with Si at the Si surface. This would mean that the oxidation kinetics should scale with the area density of Si atoms on the surface. However, as is seen in Table 8-2 the Si<110> has the greatest area density of Si atoms among these three major orientations, so in response to the experimental facts above, a more convoluted model was constructed that depended not simply on the area density of Si atoms but rather on the number of available bonds at the Si surface.[70] From the discussion above about the LP model,[28] the orientation dependence of the oxidation rate should

be incorporated in the linear rate constant, k_l, since the substrate surface ought to dominate the surface reaction. The Massoud et al. data clearly shows that for the thinnest films ($<L_o$) on the three major low index planes of Si, the order for the oxidation rate, R, is:

$$R_{l10} > R_{lll} > R_{l00}$$

This order parallels the area density of Si atoms without resorting to convoluted arguments. Sample *in situ* real-time ellipsometric data is shown in Fig. 8.14a which was obtained in dry flowing O_2 using SWE at 632.8 nm light from a He-Ne laser and incident at $\phi = 70°$. As mentioned above, this data was analyzed using a single film model. It is both interesting and

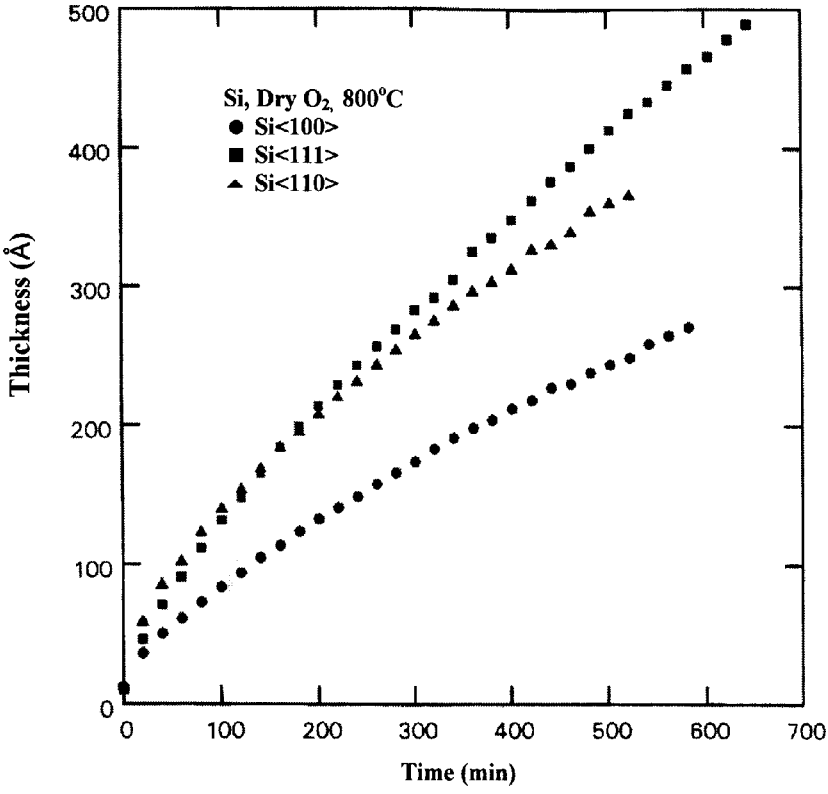


Figure 8.14 a) *In-situ* real-time Si oxidation data for various orientations of Si obtained using SWE at $\phi = 70°$ and $\lambda = 632.8$ nm and in dry O_2 at 800° C. The data was analyzed using a single film optical model; b) Similarly obtained data but at 700° C and includes the Si<311> and Si<511> orientations.

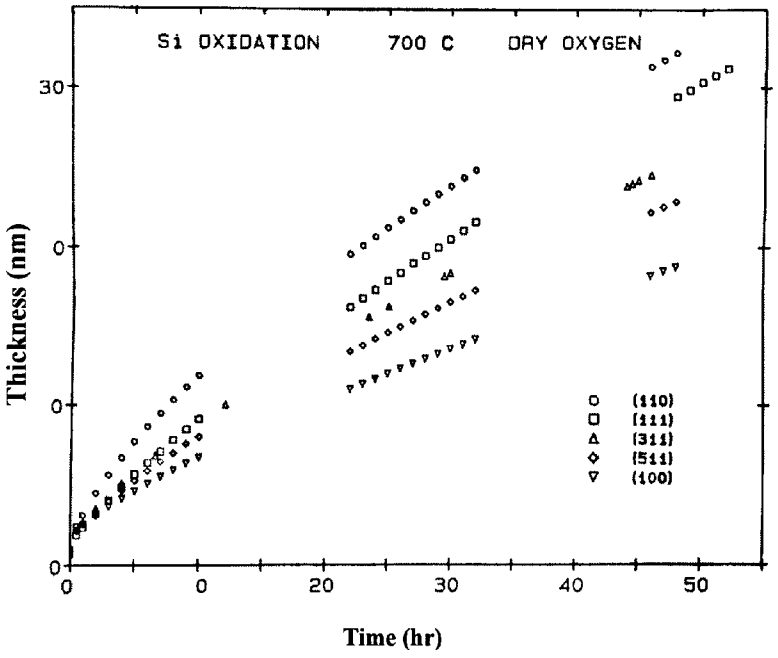


Figure 8.14 (*cont.*)

comforting to note that initially the Si<110> oxidizes fastest but there is a crossover in the order of the Si<110> and <111> orientations for thicker films (near 15 nm), yielding agreement with the early studies[70] that were performed on thicker films. An explanation for this crossover in the oxidation rates was based on the interface reaction and the effects of stress on Si oxidation.[10,49,71] The initial oxidation rate should and does scale with the area density of Si atoms. Beyond the crossover, however, another physical mechanism is required to dominate. The model invoked[10,71] is based on the observation that Young's modulus, E, for Si varies with orientation as:

$$E_{111} > E_{110} > E_{100}$$

and also on the observation that an intrinsic stress develops during thermal oxidation. The observed tensile stresses in the Si surface stretch the Si-Si bonds in the surface and thereby increase the Si surface reactivity resulting in an accelerated surface reaction rate. The relative magnitude for the observed stress should have the same order as for E above considering a constant strain resulting from the molar volume change for Si oxidation.

The actual force is obtained from the product of the stress and film thickness, hence the force grows with the oxide film growth. Thus, the dominance of this stress-modified kinetics should occur when the forces rise as the oxide grows. In the early stage of oxidation where the SiO_2 film is thin, the kinetics are dominated by the area density of Si atoms, but as both the oxide and the surface force grows, the dominance switches to that of stress dominance and hence the crossover in the oxidation rates. In short, this stress model seems to predict the qualitative aspects of the crossover observations and is even relatively quantitative with the Si<111> and <110> orientation.[71]

In an effort to confirm this model the Massoud et al. *in situ* ellipsometry data has been extended down to 600° C oxidation temperature where the stresses are larger and to include the Si<311> and <511> orientations.[68] Previously, the crossover has been seen in the oxidation data from 1100° C down to 750° C. However, for lower oxidation temperatures, the films were too thin to exhibit the crossover. The <311> and <511> orientations were chosen because ostensibly they would have higher Si atom densities if the planes are flat. However, the assumption of flatness for these planes is erroneous, because these planes are actually vicinal planes of the low index major planes, the Si<111> and <100> planes.[72] These vicinal planes are stepped with <111> risers and <100> terraces and therefore the actual area density of Si atoms is between <111> and <100> as is shown in the right side of Table 8-2. As shown in Fig. 8.14b, the experimentally determined oxidation rate order is:

$$R_{110} > R_{111} > R_{311} > R_{511} > R_{100}$$

which confirms that the area densities for these vicinal planes scales with the initial oxidation rate.

8.3.2.3 A New Initial Regime Oxidation Model

Early experiments aimed at discovering possible photon effects on Si oxidation utilized photons emanating from both a mercury arc lamp and iodine vapor lamp sources and sometimes used in combination[73] that were focused onto the cleaned and heated Si surface. An enhancement of the oxidation rate was seen for all temperatures from 955° C to 1215° C with the greatest enhancement at the lower temperatures. Using an LP analysis it was concluded that the initial regime is most affected. The conclusion was that the UV light elevates electrons over the Si-SiO_2 barrier. No further mechanism was offered. Several later investigations have been reported in which laser light has been used as the source of photons.[74–76]

In one study[74] the UV light from an Ar laser was shown to enhance the oxidation rate even after the calculated temperature rise was taken into account. Thus, both a thermal and a photonic effect was reported. Si bond breaking was proposed to account for the effect. Also, wavelength dependence was reported that indicated for light with energies greater than about 3 eV an additional enhancement was observed. In another study[75] a scanning Ar laser was used to heat a large area of the Si surface and thereby increase the oxidation rate. Again both thermal and photonic enhancements were reported. In another study in which the laser beam power density was carefully considered, an even greater photonic effect was reported.[76] This work has shown a much greater enhancement of the oxidation rate with light above 3 eV and greater effects with p-type Si and thus strongly suggests that electrons and electron related effects are important for Si oxidation. Electron effects were postulated as causative possibly through the destabilization of the O_2 molecule, and this point will be returned to below.

Reported photon enhanced oxidation effects indicate the importance of the electron barriers to Si oxidation kinetics and a new model for oxidation.[77] Thus, a consideration of the appropriate energies may also be important for normal thermal oxidation, especially in the very initial regime where oxidant transport, by whatever mechanism, is not affecting the oxidation rate. We have tested whether any of the barrier energies, shown in the energy band diagram for the Si-SiO_2 system in Fig. 8.15, could control the thermal oxidation rate, and also explored whether any of the barrier energies could alter the thermal oxidation rate. From Fig. 8.15, we see that to promote electrons from Si to the SiO_2 conduction band and thereby provide free electrons for the oxidation reaction, several routes having different energies are available. From the Si valence band photon energy above 4 eV is required, while for the Si conduction band, above about 3 eV is necessary. Intermediate in energy are the defect levels (for n and p Si) and the intrinsic Fermi level at which there are no electron states, only a probability. To determine which, if any, of these barriers may be oxidation rate limiting, we calculated the electron flux, J_{et}, from the Richardson-Dushman thermionic emission equation:

$$J_{et} = AT^2 \exp(-\chi_o/kT) \tag{8.18}$$

where χ_o is the electron barrier height, T is absolute temperature and A is the Richardson constant. Then this flux is compared with the flux of O_2 [$J(O_2)$] which is derived from the experimental oxidation rates. This comparison assumes that one electron per O_2 molecule is required for oxidation, which is justified within the specific proposed mechanism. Table 8-3 shows in the last column the calculated barrier heights such that the J_{et}

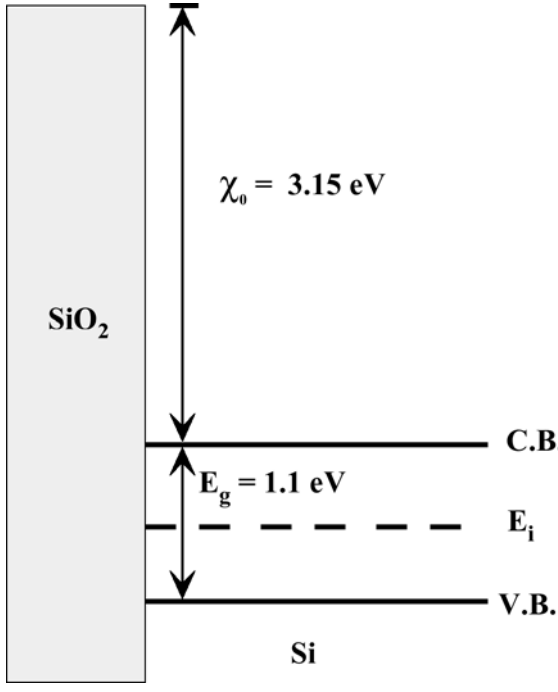


Figure 8.15 Electron energy band diagram for Si-SiO_2 system showing the barrier energies electrons from Si to SiO_2 (3.15 eV from the Si conduction band and 4.3 eV from the Si valence band).

equals $J(O_2)$ at various oxidation temperatures. It is seen that barriers of the order of 3 eV are appropriate. This is the energy barrier value for the Si conduction band electrons (see Fig. 8.15). A simple calculation confirms that there are sufficient conduction band electrons for oxidation, by a factor of ten or more, at any temperature above room temperature at which the numbers are marginal.

From these results we proposed a mechanism in which there exists a rapid relative (to the consumption of O_2) flux of O_2 to the Si surface on the SiO_2 side of the Si-SiO_2 interface, and also a rapid flux of electrons on the Si side, with the flux of electrons over the Si-SiO_2 barrier to be rate limiting. Once an electron (e^-) goes over the barrier it attaches to O_2 by a favored reaction[78] forming O_2^- which decomposes to O atoms more readily than O_2 (by 25% or more). Oxidation then proceeds readily by reaction of Si with O atoms. In a parallel way, oxidation can also occur, but much more slowly by reaction with O_2. Such a parallel reaction scheme was already suggested for the initial regime[79] and the curvature found for

Table 8-3. Barrier heights that yield equivalent thermionic electron and experimental oxygen fluxes for different oxidation temperatures, Si orientations and thickness ranges.

Oxidation Temperature (°C)	Si Orientation	Oxidation Rate (nm/min)	SiO_2 Thickness Range (nm)	Barrier Height (eV)
600	<100>	0.0004	2–3	2.87
	<110>	0.0012	2.5–6	2.79
650	<100>	0.014	2.5–7	2.95
	<110>	0.041	2.5–10	2.86
700	<100>	0.014	2.5	2.92
		0.0057	10	3.00
		0.0043	20	3.02
	<110>	0.033	2.5	2.85
		0.014	10	2.92
		0.0094	20	2.96
750	<100>	0.028	5	3.02
		0.024	10	3.03
		0.019	20	3.06
	<110>	0.057	5	2.96
		0.035	10	3.00
		0.024	20	3.04
1000	<100>	1.79	7	3.35
		0.9	20	3.43
		0.27	100	3.56
	<110>	3.45	9	3.28
		1.74	20	3.35
		0.4	100	3.52

Arrhenius plots for linear rate constants could be explained based on a parallel path reaction scheme.[80]

Finally, this e^- limited mechanism yields insight into the formation of the 1 nm native oxide that forms virtually instantly on a fresh Si surface even at room temperature, yet virtually ceases to grow after about 2 nm unless the temperature is raised. If we consider the approximately 10^{15} Si surface electronic states, most of which have eventually captured an e^- from the bulk Si, then there are some 10^{15} e^-'s available for Si oxidation. These e^-'s are existing in closely spaced levels and require only little energy promotion. Thus these electrons are available in parallel to the thermionically produced electrons. The 10^{15} electrons at one e^- per O_2

molecule would yield about 1 nm SiO_2 which is the experimentally measured native oxide thickness. Once the states are removed via oxidation, however, this native oxide can no longer form and the thermionic and/or photonic excited electrons are required for further oxidation.

8.3.2.4 Low Temperature Oxidation Data

Several thermal oxidation studies have appeared[15–19, 42] since the first thermal oxidation studies of the late 1960s that have employed *in situ* and real-time ellipsometry. The advantages of this technique are that dense data sets are obtainable which are amenable to significant data analysis and that the oxidized samples are not subject to damage or impurity contamination through exposure to the laboratory ambient during an experiment, i.e., data in a given oxidation experiment are obtained from one sample. The dense data sets of the Massoud et al. studies[17–19] confirm that a break occurs in the Arrhenius plots of the linear and parabolic rate constants with the temperature indicating a change in behavior around 950° C. This temperature coincided with the viscous flow temperature that was reported by a number of workers.

Viscous flow ideas[10] were cast in terms of possible effects on oxidation kinetics in two ways.[81] One way is to permit the stress to affect the interface reaction and the other is to alter transport. For the initial regime, the rate of oxidation is considered to be proportional to the rate at which newly formed SiO_2 can flow into the free direction (normal to the oxidizing Si wafer surface). As was written above, the specific formulation was based on the linear rate constant as:

$$R = k_l C_2 \tag{8.19}$$

where the rate, R, is proportional to the oxidant concentration at the interface, C_2, and k_l contains information about the Si surface. To include stress, we first make explicit the Si surface atom area concentration, C_{Si}, but as an effective Si concentration, C^*_{Si}. C^*_{Si} is given as the product of C_{Si} multiplied by the rate at which the oxide flows away from the Si surface, $\dot{\gamma}$:

$$R = k' C_2 C^*_{Si} \tag{8.20}$$

and

$$R = k' C_2 C_{Si} \dot{\gamma} \tag{8.21}$$

But $\dot{\gamma}$, the strain rate is given as the stress in the Si surface, σ, divided by the SiO_2 viscosity, η, according to the Maxwell viscoelastic model for a solid hence the final result for the initial rate is:

$$R = k' C_2 C_{Si} \sigma / \eta \tag{8.22}$$

As discussed above the initial rate should then scale with the product $C_{Si}\sigma$ for a given C_2, with different Si orientations, and qualitative agreement was found in the experimental data.

For the models of transport, researchers[82–84] agree that the oxide stress, being compressive, decreases the diffusivity, D. Attempts were made to describe both the initially fast oxidation regime and the thick film growth kinetics based on the stress (or strain) dependence of D and while the authors get good fits of their models to the experimental data, direct verification of any one model over another remains lacking. On the issue of data fitting to one model or another, it is useful to reconsider this question. With the advent of dense, *L*,*t* oxidation data, via *in situ* ellipsometry experiments, it is tempting to try to distinguish among models based on the quality of the fit to the data. However, caution must be exercised in that different forms of similar equations will give identical fits to the data though the forms are based on different physics. Also, errors in the data or the formulation of the correct physics may yield worse fits than an incorrect but mathematically more flexible model. The reader is cautioned that many reported incorrect oxidation models fit the *L*,*t* data rather well.

8.3.2.5 Si-SiO_2 Interfacial Layer Models

Most researchers of the Si-SiO_2 interface provide results that either suggest or directly show that the Si-SiO_2 interface is different than either Si or SiO_2. An extensive review of the Si-SiO_2 interfacial regime[85] strongly supports the notion that the interfacial region is chemically and physically distinct. Based on the findings about the Si-SiO_2 interface attempts have been made to both further determine the nature of the Si-SiO_2 interface region and to incorporate the understanding into a refined Si oxidation kinetics model.

From several published studies,[86–88] it was suggested that there is likely an epitaxial relationship between the first several layers of oxide grown on the Si surface. This is derived from structural compatibility arguments and the minimization of the molar volume difference between the two phases. In terms of an oxidation model a two step process was envisioned in which the first step produced the epitaxial layer of oxide, but with a concentration of interstitial Si atoms. The atoms resulted from

the imperfect match across the phase boundary. The second step was the oxidation of the interstitial Si atoms with the concomitant amorphization of the oxide due to the lattice expansion. Studies on the transport of oxidant through a growing oxide[89] yielded some evidence that the very initial oxide forms as a result of the motion of O atoms as opposed to the findings for thicker oxides. This new mode of transport for the very thin films suggested the possibility for an ionic transport mechanism.

More recently, a novel ellipsometric method has been developed that improves the sensitivity of ellipsometry to the interface underneath a transparent film.[90] There are several different ways to study the interface region between film and substrate as depicted in Fig. 8.16. One way is by using spectroscopic ellipsometry in air ambient (Fig. 8.16a). The disadvantage of this method is that an accurate characterization of the ultra-thin interface transition layer is complicated by the inability to discriminate the

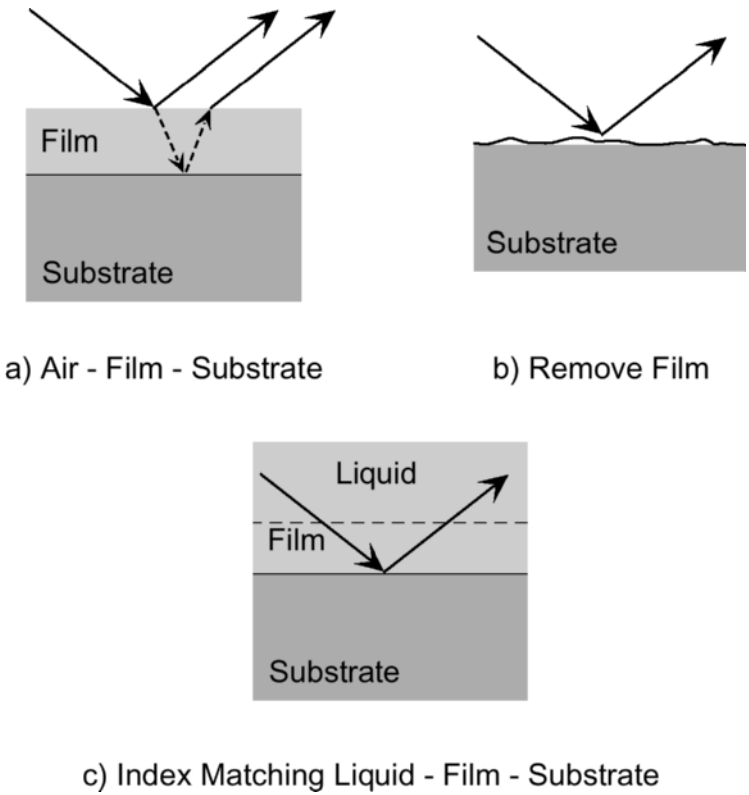


Figure 8.16 Interface characterization methods: a) observe interface with overlayer intact; b) remove overlayer: c) refractive index match to overlayer.

optical contributions of the relatively thick overlayer and the thin transition layer by the measured ellipsometric parameters. Another way is to remove the overlayer physically or chemically and then to probe the interface (Fig. 8.16b). However, this method could alter the interface region. In order to overcome these problems we have developed the technique of spectroscopic immersion ellipsometry[90,91](SIE), which uses a transparent liquid ambient that has optical properties very close to the optical properties of the dielectric overlayer thereby eliminating the optical response of the overlayer (Fig. 8.16c). Hence, this technique "optically" removes the overlayer and thus enhances the sensitivity to the interface properties. The interface sensitivity of Δ is substantially increased (more than 10x) using the liquid ambient as is shown in Fig. 8.17 (open circles) which compares the relative interface sensitivity $\delta\Delta(E) = \Delta_0(E) - \Delta_\infty(E)$ for air and CCl_4 ambient that refractive index matches to SiO_2. $\Delta_0(E)$ and $\Delta_\infty(E)$ are calculated without and with an assumed interface layer, respectively. Ψ is also obtained with improved sensitivity but much less than for Δ.

In our analysis, the working model for the interface between crystalline Si substrate and amorphous SiO_2 film is shown in Fig. 8.18. The transition region has a structure with two major components: the "physical" interface and the "chemical" interface. The "physical" interface can be represented by microroughness or protrusions of Si into the oxide. The

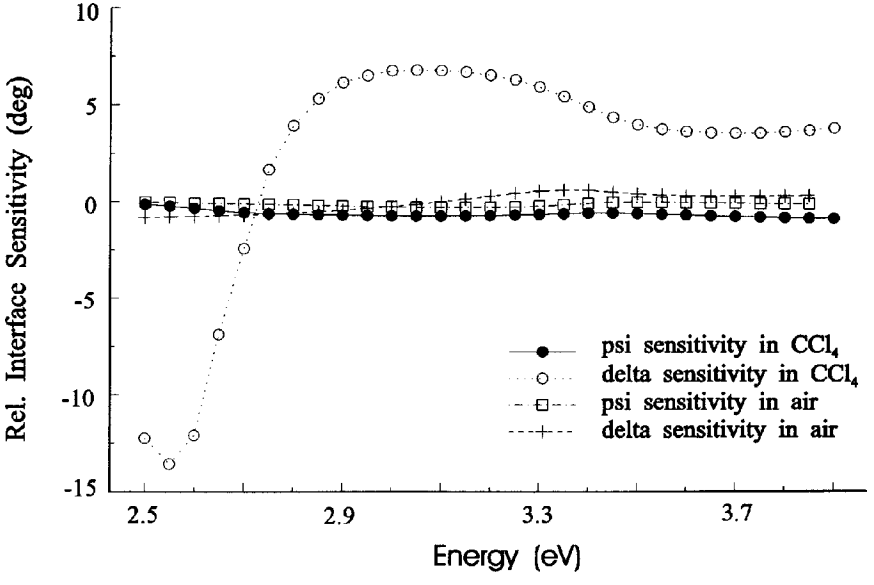


Figure 8.17 Relative interface sensitivity of Ψ and Δ versus incident photon energy with and without index matching CCl_4.

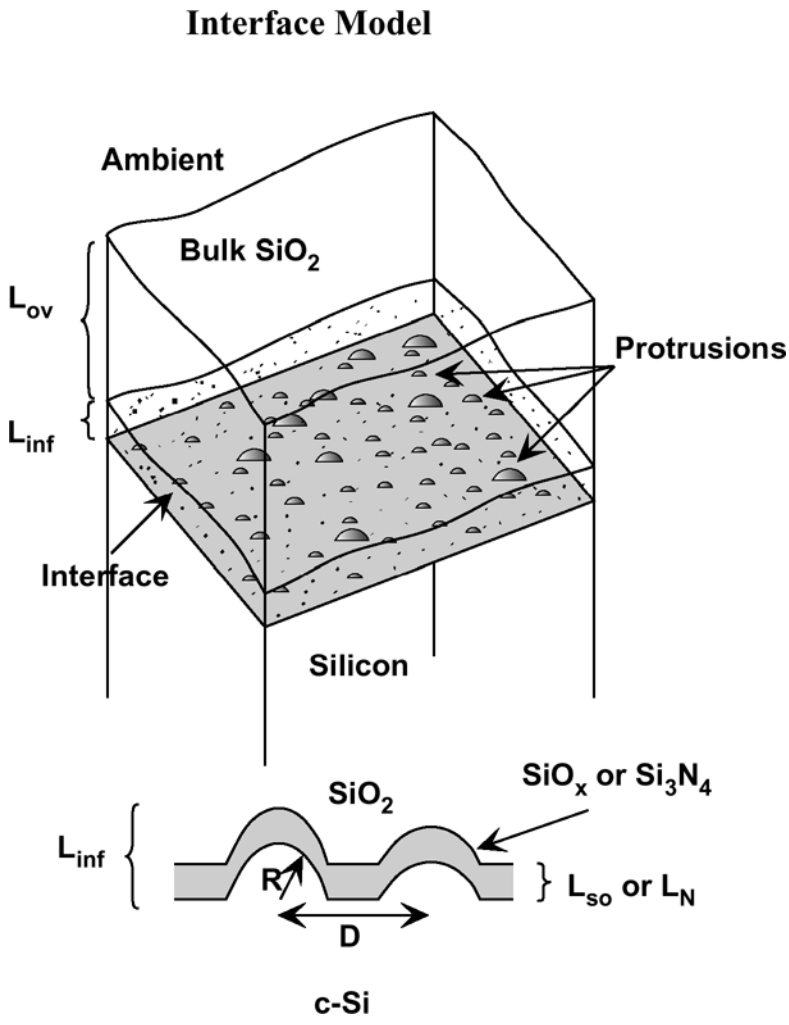


Figure 8.18 Interface model for the Si-SiO_2 interface. The interface layer (L_{inf}) made up of Si roughness (R) and suboxide (L_{so}) on a Si substrate. The bulk SiO_2 film has a thickness $L_{ov.}$

"chemical" interface consists of a suboxide, SiO_x with $0 < x < 2$. For the case of nitridation of Si the interface layer could be a nitride or even an oxynitride. The crystalline silicon protrusions are described as hemispheres with an average radius R, which form a hexagonal network with an average distance D between centers. The protrusions and the region between them are covered by a layer of suboxide assumed to be

SiO (i.e., $x = 1$) with an average thickness of L_{SiO}. An effective interface thickness is given as:

$$L_{inf} = R + L_{SiO} \tag{8.23}$$

The Bruggeman effective medium approximation (BEMA) was used to calculate the effective dielectric function of the interface.

The evolution of the Si-SiO_2 interface as a function of high temperature annealing (750–1000° C) was investigated by SIE. Figure 8.19 shows raw Δ data with the significant changes and Fig. 8.20 shows the data in terms of an effective relative interface parameter defined as:

$$\delta\Delta_{inf}(T_{an}, t_{an}) = \delta\Delta^{exp}(T_{an}, t_{an}) - \Delta_o^{exp} - \delta\Delta_{ov}^{cal}(T_{an}, t_{an}) \tag{8.24}$$

where $\Delta^{exp}(T_{an}, t_{an})$ is the experimental ellipsometric angle Δ at an annealing temperature and time, Δ_0 is the ellipsometric angle for an unannealed sample, and the term $\delta\Delta_{ov}^{cal}(T_{an}, t_{an})$ is the overlayer relaxation correction that corrects for the very small change in SiO_2 with annealing. This term

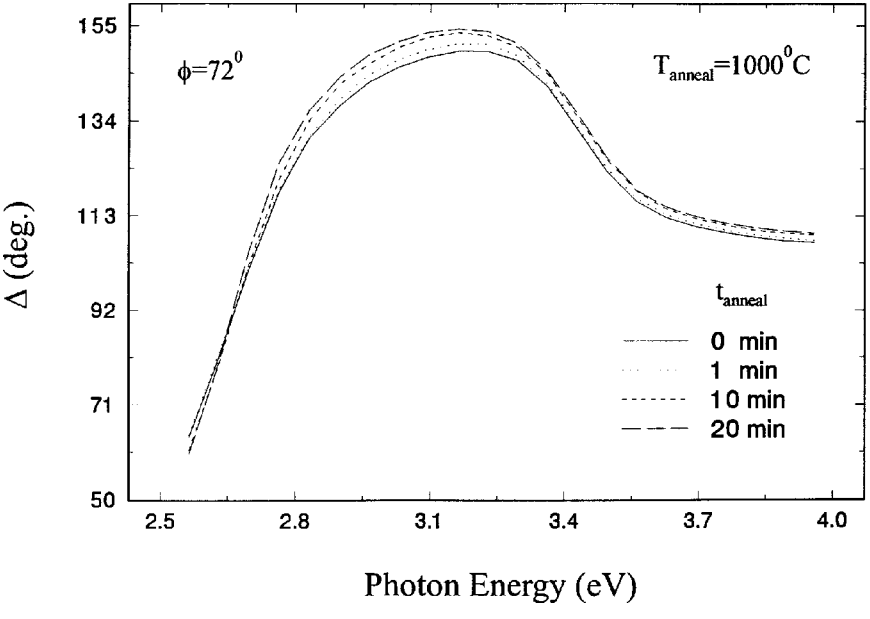


Figure 8.19 Dependence of Δ on incident photon energy for various interface annealing times. The measurements were performed on a 3 nm SiO_2 film on Si immersed in CCl_4 after anneal at the specified time and cooling from the 1000° C anneal temperature.

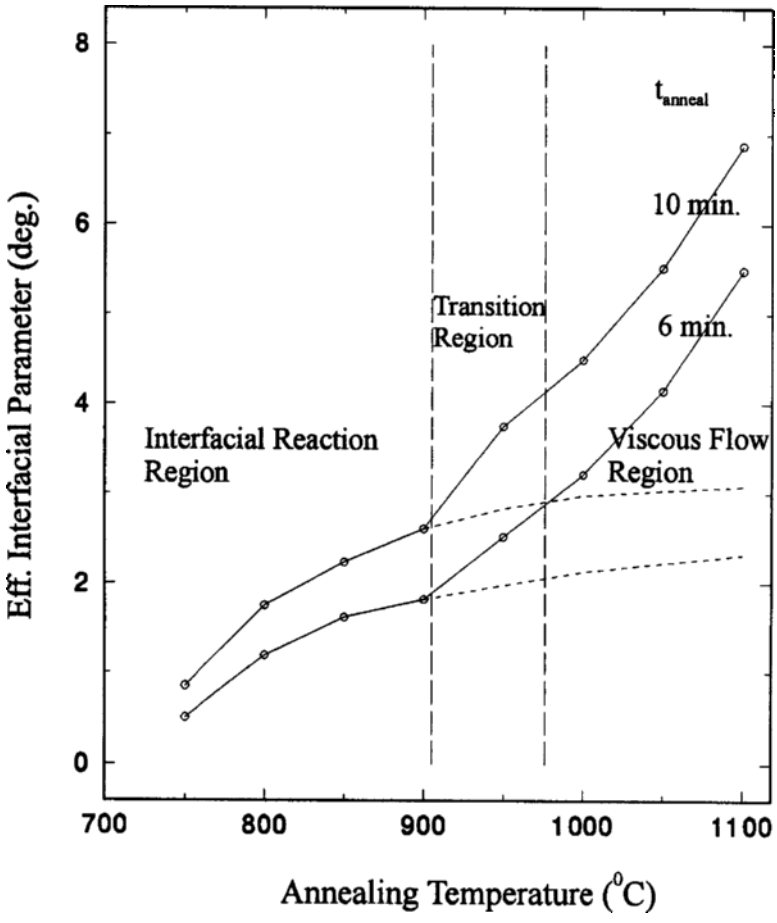


Figure 8.20 The dependence of the effective interface parameter (defined in the text) on annealing temperature for two anneal times (t_{anneal}). The measurements were performed on a 3 nm SiO_2 film on Si immersed in CCl_4 after anneal at the specified time and cooling from the anneal temperature.

is always less than a few percent of the other changes and thus could be omitted. Distinct changes are observed for anneal temperatures above 900° C that correspond to strain relaxation temperatures as was discussed above. Figure 8.21 shows modeled data in terms of the interface thickness defined above as L_{inf}, which displays the temperature-time dependent shrinkage of the interface with annealing. For short annealing times, a rapid change in the interface is observed that correlates with the disappearance of protrusions, followed by a slower change that correlates

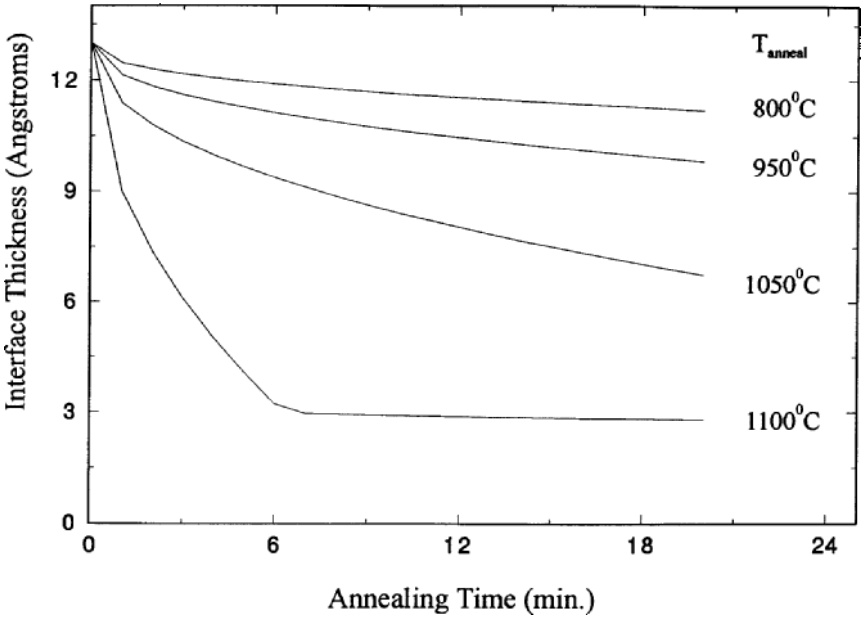


Figure 8.21 The changes in the interface layer thickness with annealing time for several anneal temperatures. The measurements were performed on a 3 nm SiO_2 film on Si immersed in CCl_4 after anneal at the specified time and cooling from the anneal temperature.

with the disappearance of the suboxide. At high annealing temperatures, we believe that viscous relaxation dominates, while at low annealing temperatures the suboxide reduction is apparent. With the use of the above optical model, we found that the thickness of the SiO layer at the interface, L_{SiO}, for all <100>, <110>, and <111> Si substrate orientations increased slightly, and the average radius of the crystalline silicon protrusions, R, decreased with the thickening of the SiO_2 overlayer. This yields an overall decrease in the interface layer (L_{inf}) as is seen in Fig. 8.21. These results are consistent with the well accepted LP Si oxidation model which yields an accurate representation of the growth of SiO_2 on Si over a wide range of thickness, temperature, and oxidant partial pressures. Also, as will be discussed later, these results are concordant with Si surface smoothing from oxidation.

As shown above the LP formula reduces to: $t \cong L^2/k_p$ for large film thickness. This is termed the parabolic growth law and implies that oxide growth is diffusion controlled. In other words, as the oxide layer gets thicker, the oxidizing species must diffuse through a larger distance to

arrive at the Si/SiO_2 interface. The reaction thus becomes limited by the rate at which the oxidizing species diffuse through the oxide. It was shown above in Section 8.3.2.1. that at elevated temperatures and with an oxygen deficiency, SiO_2 in contact with Si disproportionates to SiO. The disproportionation reaction is initiated at active defect sites already present at the Si/SiO_2 interface. In our model the Si protrusions may be considered as defects that could cause the above decomposition, since these sites are thermodynamically active due to the smaller radius of curvature. This is consistent with our results that show that with the thickening of the SiO_2, the thickness of SiO layer, L_{SiO}, at the interface increases and the average radius of the crystalline protrusions, R, decreases.

8.3.3 Recent Results on Ultra Thin SiO_2 Films and the Si-SiO_2 Interface

As was mentioned earlier in this chapter, gate dielectric films less than 5 nm thick are becoming common in MOS technology. Film thickness in this range is difficult to measure accurately. These ultra-thin films on semiconductor substrates are nearly all interface. Consequently, the nature of the interface, including interface roughness, emerges as a crucial parameter for ultra-thin films.

8.3.3.1 Film Thickness Related Results

Ellipsometry is often used to measure film thickness. However, in the ultra-thin film regime ellipsometry cannot accurately determine film thickness. This is due to both the general unavailability of refractive indices for thin films and the inability to measure refractive indices for ultra-thin films as was discussed above. Since ellipsometry measures the product of refractive index and film thickness (this product is the optical path length or optical thickness), if the index is not known or difficult to obtain accurately, then the errors in index will result in thickness errors. It is now useful to return to Fig. 8.3 that shows a simulated Δ,Ψ space trajectory for several transparent films on a Si substrate that have disparate n values. It is seen that even though the n are different, the trajectories converge as the film thickness decreases. For films that are closer in n the convergence occurs at thicker values. When the degree of convergence is within the ability of ellipsometry to measure angles accurately (about 0.01° at best) then the film n cannot be unambiguously determined using ellipsometry. While the precise conditions under which ellipsometry becomes ineffective depends on the system and instrumentation, in general refractive indexes for films

less than 10 nm are typically not obtained accurately from ellipsometry. Furthermore, it is dangerous to use the bulk film refractive index for a film that is less than 10 nm, since it is known[9,92,93] that the optical properties including the refractive index for ultra-thin SiO_2 films are different than for thick films. For SiO_2 the ultra-thin films have a higher index than the bulk film due to compressive interfacial stresses and suboxide content. Thus, the use of the low bulk refractive index for thin SiO_2 films would yield too large of a film thickness.

Other techniques such as x-ray photoelectron spectroscopy (XPS) and cross-sectional transmission electron microscopy (X-TEM) have commonly been used to measure thin film thickness. However these techniques also have significant errors associated with them. For XPS for example,[94] besides problems with adventitious C on the surface and photoelectron diffraction affecting measured intensities, the XPS measurement of film thickness requires the measurement of the attenuation length (A.L.) for the photoelectron. A.L. is typically measured in a separate experiment, and it requires knowledge of the film density which is usually not known accurately, since it can be different from the bulk density. The film thickness is obtained from a formula of the form:

$$L_{ox} = (\text{A.L.}) \sin\theta \ln\left(1 + \frac{1}{Q}\right)$$

In this formula Q is a product of two intensity ratios that are measured in separate experiments. Thus, XPS is subject to many potential sources of errors for film thickness measurements. X-TEM accesses the projection of the lattice planes throughout the cross-section sample thickness as depicted in Fig. 8.22. If the sample is rough as shown or is tilted the accuracy of the film thickness is affected. In particular roughness makes it difficult to find the interfaces, and sample tilt would yield the minimum film thickness due to the projection of all the lattice planes in the cross section. Thus, the combination of ellipsometry using thick film n's plus X-TEM can be used to bracket the real film thickness.

Without an accurate knowledge of film thickness it is difficult to fully characterize device properties, since film thickness determines the distribution of electric fields. Also accurate oxidation processes cannot be developed with inaccurate film thickness and the LP model that applies to SiO_2 thicker than 10 nm cannot be used to model or predict the ultra-thin SiO_2 film growth processes via thermal oxidation. In addition, device operation at high electric fields can be hampered in the cases where the interface between the SiO_2 film and Si substrate is rough. Both high field carrier mobility and film dielectric reliability can be reduced by interfacial roughness.

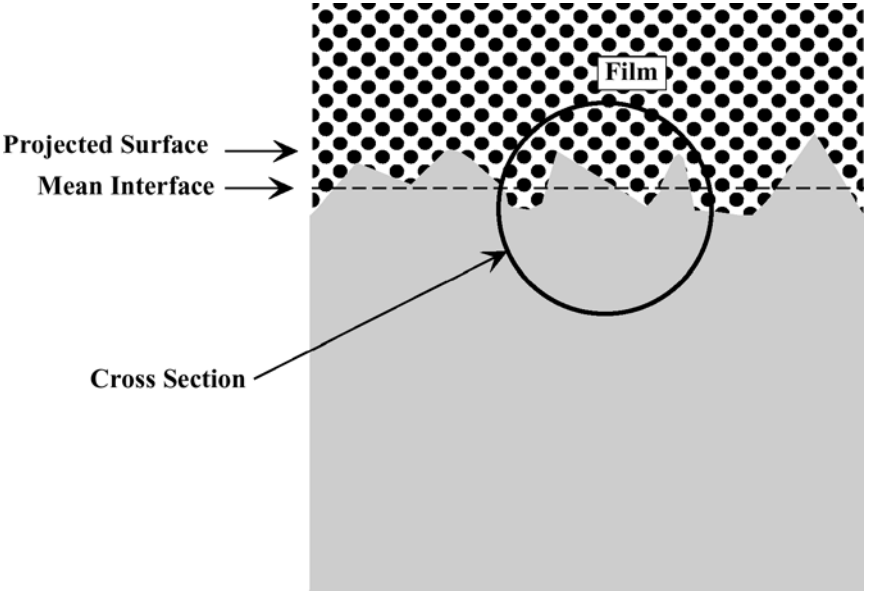


Figure 8.22 Pictorial cross-sectional transmission microscopy of a rough interface.

In order to address the measurement of these ultra-thin films, the use of special techniques is needed that have the ability to reliably measure the film and interfacial properties. It would be desirable to have techniques that are not only sensitive to the special properties, but that are also nondestructive. To this end several techniques have recently been developed that can access desired ultra-thin film properties such as refractive index, thickness, and interfacial roughness. In the following paragraphs some results using these techniques are discussed.

One technique that has been developed to obtain accurate film thicknesses uses Fowler-Nordheim (FN) electron tunneling current oscillations (FNCOs), a subject that was discussed theoretically in the 1960s,[95,96] and in the 1970s and 1980s was experimentally developed by Maserjian and colleagues.[97] As is shown in Fig. 8.23, FN tunneling current oscillations result from the interference of incident and reflected electron waves in the SiO_2 film conduction band in a MOS structure. The interference of propagating electron waves occurs when the applied oxide potential V_{ox} is greater than the electron barrier Φ_M. The solution for standing waves in this system for a trapezoidal barrier has the form of Aries functions. Thus, the oxide acts analogous to an optical etalon (a spatial filter) where the path through the barrier, the film thickness, is a parameter in the analytic solution. From Aries functions solutions and the experimentally determined

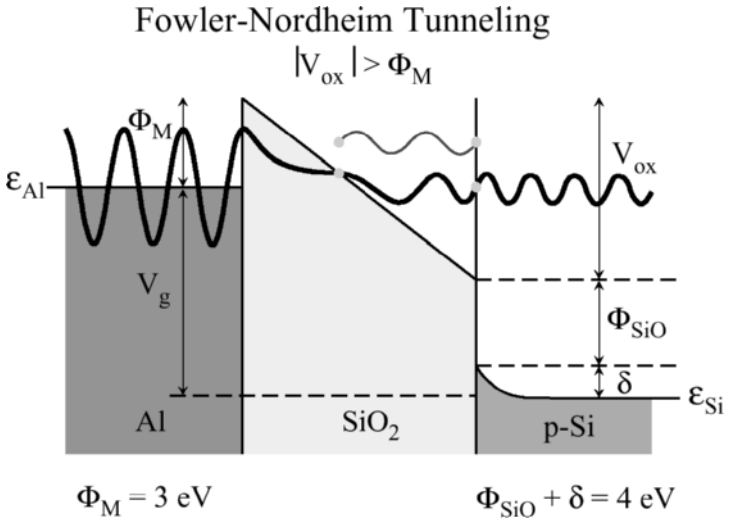


Figure 8.23 Fowler-Nordheim (FN) energy barriers. When the applied voltage $V_{ox} >$ the barrier (Φ_M) FN tunneling occurs and FN tunneling current oscillations (FNCOs) are possible from the interference of the electron waves.

maxima and minima in the oscillations, the film thickness can be obtained to an accuracy of better than 0.1 nm using a straightforward procedure recently developed in our laboratory.[98] Most of this error is due to uncertainties in the barriers and effective masses. Briefly, the new procedure starts with the FN I-V data that contains the oscillations riding a large D.C. component, and then transforms the data in terms of a dummy variable using the usual FN formula derived without considering electron wave interference. This variable is electric field independent and contains both the oscillatory and D.C. components of the tunneling current. Then using derivatives, the zeroes of the total tunneling current (A.C. and D.C.) are determined from which the electric field could be obtained as well as accurately separating the oscillatory component from the total tunneling current. The positions of the extrema of the FNCOs that are readily read from the extracted oscillations contain the film thickness information to an accuracy conservatively estimated to be better than ±0.1 nm. It should be noted that this new method to extract the small FNCOs from the large D.C. FN tunneling background current is greatly superior to an older method[97] that merely subtracts the theoretical DC FN current characteristic from the experimental data, i.e., subtracts two large numbers to obtain a small difference. The previous method works well for nearly ideal samples but leads to large errors for samples that are less than ideal and present few oscillations.

In an early study[99] ellipsometry was performed on the same samples for which the SiO_2 film thickness was determined using FNCOs,. Specifically, highly accurate single wavelength ellipsometry was used at 632.8 nm to measure the ellipsometric variables, Δ and Ψ. Using a single film model, and with the film thickness from FNCOs as input, the refractive index was calculated, thereby obverting the issue of obtaining accurate indexes from ellipsometry alone. For film thickness of about 4.4 nm, n for thermal SiO_2 films was found to be about 1.894 as compared with a value of 1.465 for bulk SiO_2 films thicker than 20 nm. Using this new value for n for ultra-thin SiO_2 films as well as known bulk values, we have generated an interpolation formula for SiO_2 refractive indexes as a function of film thickness.[100] Using these new n values for SiO_2, we have re-evaluated our previously published SiO_2 film thickness versus Si oxidation time data. Figure 8.24 shows that the corrected SiO_2 film thickness values (labeled "this work") for the ultra-thin regime are smaller than previously calculated using the lower bulk SiO_2 indexes. From this study[100] we can confidently conclude that the controversial initial thermal oxidation regime is purely linear, indicative of an interface reaction. Note that in the early Si oxidation because the bulk index was used to evaluate the *in situ* real time ellipsometry data a curvature was obtained for the initial

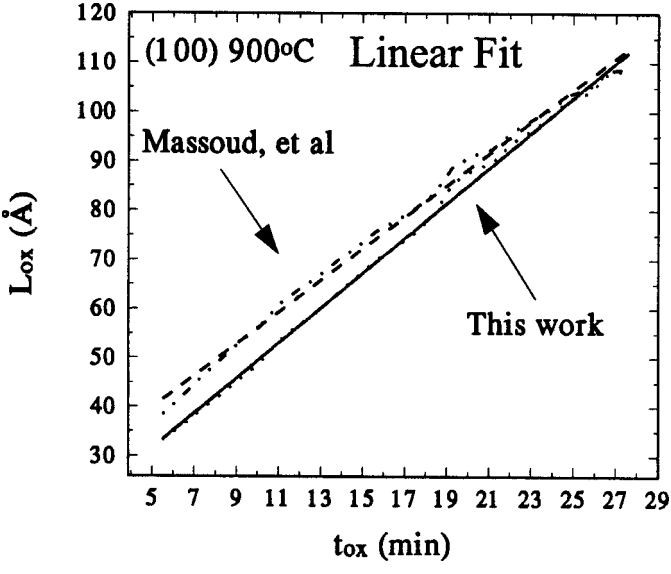


Figure 8.24 Using refractive indexes from the combined FNCO and ellipsometry method, Si oxidation data has been corrected and the older Massoud et al. data (from ref. 67) is compared with corrected data (from ref. 100).

regime. That this was the case was only very recently clarified with experimental results.

Also recently,[101] we have further refined the ultra-thin film SiO_2 refractive data. We collected variable angle SE data on SiO_2 films ranging in thickness from 2 to 8 nm and processed the data using an iterative algorithm shown in Fig. 8.25, in order to obtain consistent values for the index parameter. The result is a new recursion formula shown in Fig. 8.26 along with our previous formula[99] for comparison. These new results have extended the range of thickness for which we obtain refractive indices down to about 1.5 nm, and enable the accurate ellipsometric determination of SiO_2 film thickness to about 1 nm.

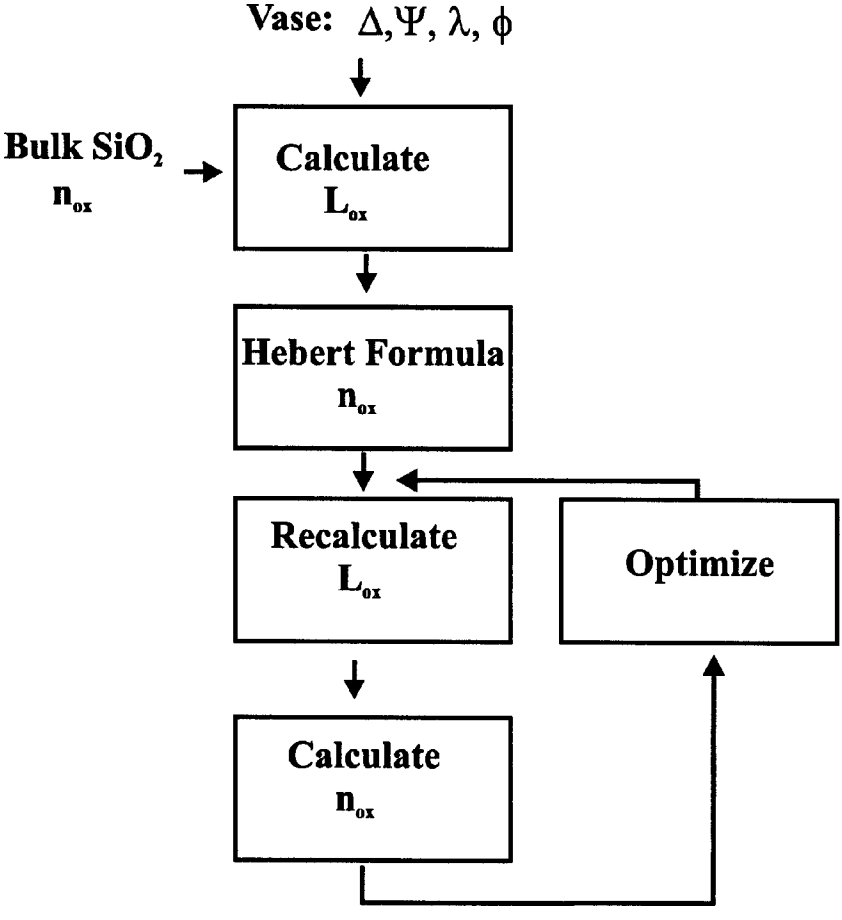


Figure 8.25 The algorithm used to correct the *n* results using VASE results.

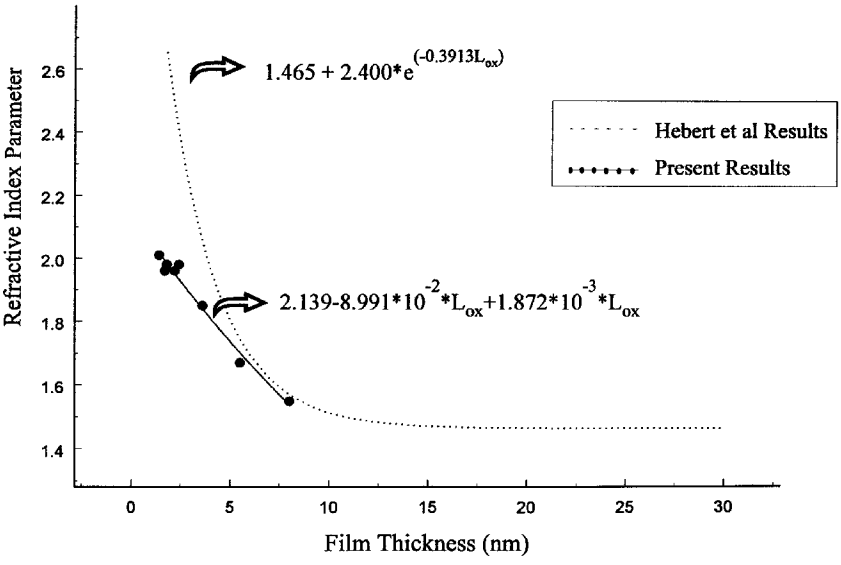


Figure 8.26 The *n* results for thin film SiO_2 using the algorithm in Fig 8.25 (present results) compared with earlier results (from ref. 100) from SWE and FNCOs.

8.3.3.2 Interfacial Roughness

For direct roughness measurements, atomic force microscopy (AFM) is commonly used to obtain an image of the surface. From the lateral *x*,*y* positions the height of a surface feature, *z*, is obtained with subnanometer accuracy. This data is typically analyzed to obtain the extent or width of the interface that is reported as the root mean square (RMS) roughness. The RMS roughness is a measure of the average magnitude of the roughness features. However, the RMS is insufficient to describe a rough surface or interface, and insufficient to distinguish one rough surface from another. We have developed a fractal analysis, in order to determine the spatial complexity of the roughness.[102,103] In fractal geometry a fractal dimension, D_F, is a scaleless non-integer parameter that is a unique descriptor of an object, analogous to the integer dimensions in Euclidean geometry. For rough surfaces D_F would be a non-integer between the integers 2 and 3, which are the Euclidean dimensions: $2 < D_F < 3$. In previous studies we have described a modified variation method algorithm for accurately extracting D_F from the AFM data,[102] and we have discussed the applicability of using D_F for Si roughness studies.[103]

We have used both AFM and SIE to follow the evolution of Si roughness resulting from both the thermal and electron cyclotron resonance

(ECR) plasma oxidation of Si.[104–106] Purposely roughened Si was used so that the level of the roughness was well above the lower detection limits of the techniques.

Figure 8.27 summarizes some of our results for thermal oxidation. Figure 8.27a displays AFM results that show that the RMS values for

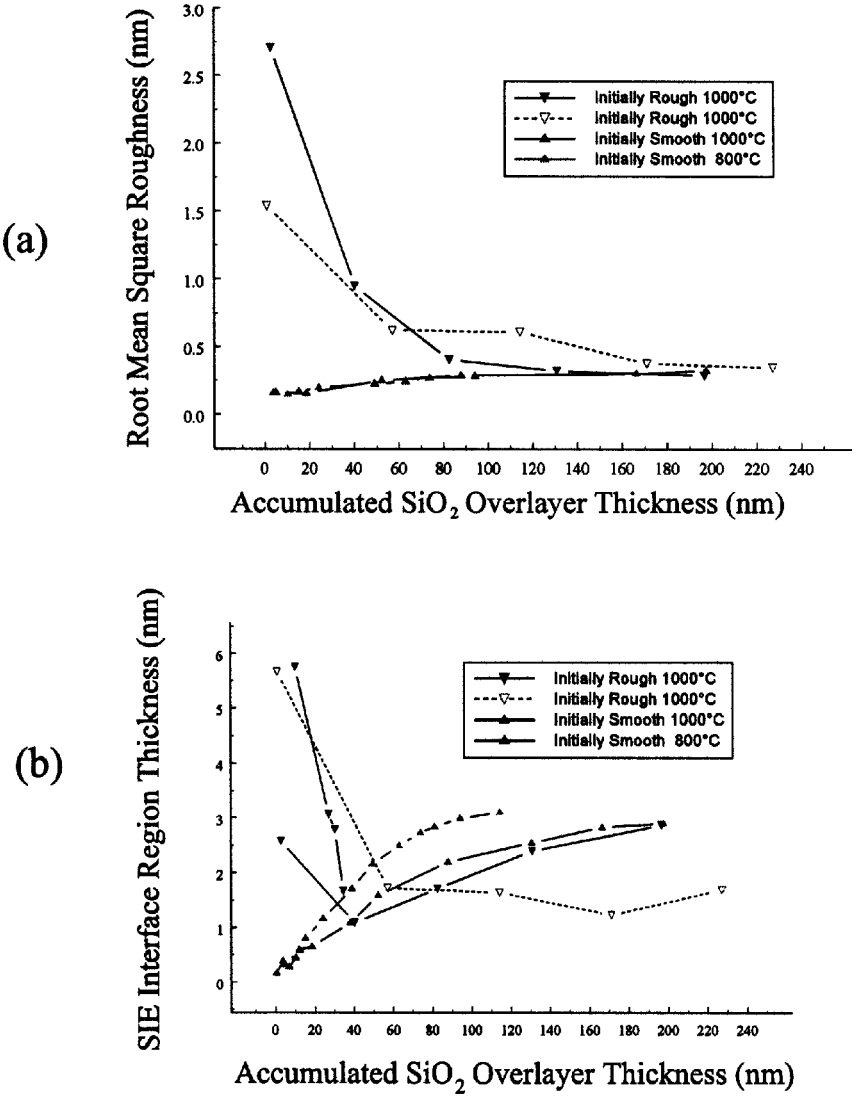


Figure 8.27 Si oxidation induced roughness from a) AFM and b) SE (SIE technique).

rough surfaces decrease with increasing extent of oxidation while for initially smooth surfaces the RMS values increase with the extent of oxidation. From this data there appears to be convergence to a limiting roughness of 0.2 to 0.3 nm RMS. It was found that D_F values for both initially rough and smooth Si surfaces monotonically decrease with increasing extent of oxidation indicating that the surface always becomes simpler as oxidation proceeds. Figure 8.27b displays SIE results that show that as oxidation time increases rough surfaces result in a smaller L_{inf} while initially smooth surfaces show an increasing L_{inf} with oxidation extent. Thus, both AFM and SIE are concordant in that initially rough surfaces become smoother, and initially smooth surfaces roughen in terms of the magnitude of the roughness features. However, the smallest features are always removed at a greater rate than large features for both initially rough and smooth surfaces (D_F always decreases with oxidation). Convergence is not seen for the SIE data but rather a crossover is observed in the values of L_{inf}. This apparently non-physical result is explained by considering the interface model (Fig. 8.18) that is used to analyze the SIE data. In this model the parameter extracted L_{inf} is the sum of the height of the Si protrusions plus the suboxide layer. The suboxide layer forms during oxidation, and grows as the oxidation rate slows and for lower temperature oxidation. Also, the effect of the suboxide would be greatest for the smooth samples where the Si protrusions are small. The extent of the suboxide then increases L_{inf}, but the suboxide does not affect the AFM derived RMS topographic measurements. Thus, we believe that the crossover seen in the SIE results is an artifact of the model used and is attributable to the suboxide formation which affects only the optical results.

Similar results have also been obtained for ECR plasma oxidation for purposely roughened and initially smooth Si surfaces.[105] For ECR plasma oxidation an additional acceleration of smoothing and roughening could be obtained by increasing the oxidation rate using a positive D.C. sample bias.

Both the smoothing and roughening effects can be understood using the Kelvin equation:

$$\Delta G = \frac{2\gamma V}{R}$$

where the change in local free energy, ΔG, is inversely proportional to the radius of curvature (R) for a small feature. γ and V are the surface energy and molar volume, respectively. The Kelvin equation teaches that small sharp features (with small R) present a greater reactivity, and hence oxidize more extensively relative to larger features. With the use of Fig. 8.28 both the smoothing of initially rough surfaces and the roughening of initially smooth Si surfaces can be explained.[106] If elliptical protrusions are assumed, then

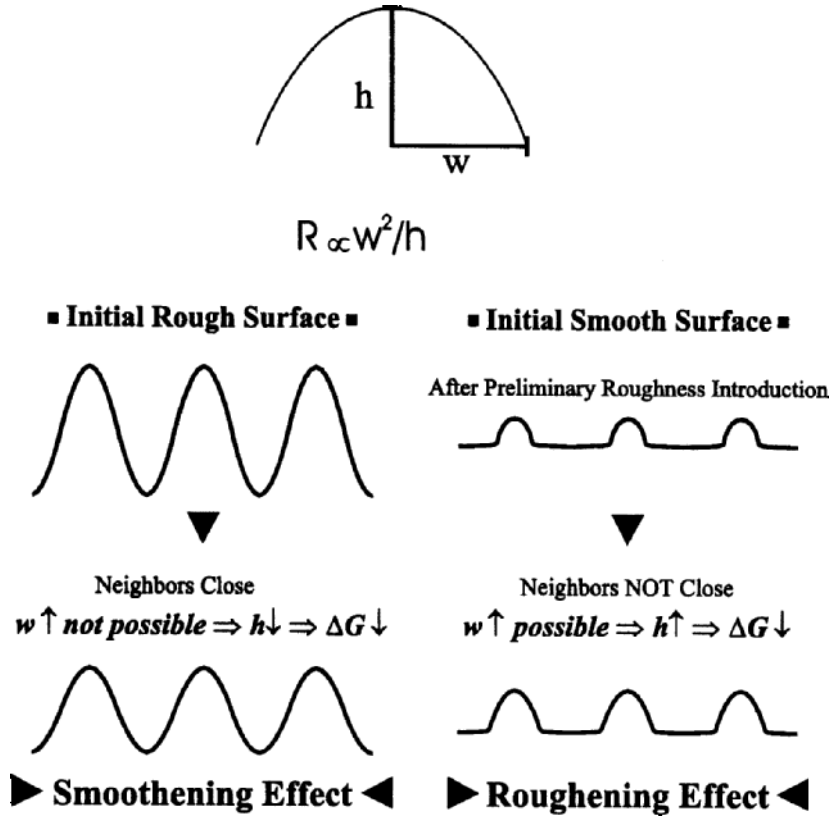


Figure 8.28 Model showing how rough and smooth surfaces change with oxidation.

the top of Fig. 8.28 shows that these protrusions can reduce their radius of curvature *R* and hence the local free energy both by an increase in the width *w* (squared) and/or a decrease in the height *h*. For the purposely roughened samples (Fig. 8.28 left side), the presence of closely spaced features requires that a decrease of *h* is the only way for the free energy to decrease as a result of oxidation. This is the case, because if *w* also increases for one feature, for an adjacent feature *w* would be forced to decrease. However, for the initially smooth surfaces (Fig. 8.28 right side) without closely adjacent features, both w and h can change so as to reduce the local free energy. In fact if *w* can increase to lower the free energy, then an increase in *h* is permitted since *R* is proportional to w^2. This leads to roughening as measured by an increasing RMS. It should also be noticed that for both the initially rough and smooth Si surfaces the D_F decreases as a result of oxidation. Also, the fact that both roughening and smoothing can occur as independent mechanisms leads to

the prediction of a limiting interfacial roughness that has now been experimentally verified in Fig. 8.27a to be about 0.3 nm.

We have recently investigated[107,108] the possible effects of roughness on interface charges in terms of fixed oxide charge, Q_f, and interface trapped charge density, D_{it}. Since the interface charge is an extensive property in that the charge level depends on the surface or interface area, an accurate measurement of surface area for the rough surfaces was required. We have found that the usual algorithms that are available with commercial AFMs are not useful for extracting areas when the roughness is in the micro-roughness regime, viz. with roughness features less than 5 nm high. This is because the commercial algorithms use a triangle method where three adjacent AFM (x,y,z) data points form a triangle, and the triangles are summed to obtain surface area. The mathematics involves squares of the lateral (x,y) and height (z) values to form the triangles. If z is much smaller than x or y which is typical for micro-roughness, the z height values will be lost in the round-off error. Thus, the area of micro-rough surfaces obtained from these algorithms is about equal to the projected area, which is far too small. To overcome this problem, we developed a new algorithm for extracting the surface area for micro-rough surfaces from AFM images.[107] This new algorithm is based on RMS and D_F values, and the way these values scale with roughness. When an accurate area for a rough surface has been obtained through the use of our new algorithm applied to AFM data, we have found that both interface electronic states and interface fixed charges scale simply with surface area.[108] These results are shown in Fig. 8.29 where it is seen that the charge values normalized using the projected contact areas show high values, which have a relatively strong correlation with the extent of oxidation. The charges decrease with oxidation indicative of smoothing and decreasing area. With the area correction most of the roughness issues are removed. Lastly, a small orientation correction is made to account for the fact that roughness introduces other orientations, which can have higher charge values. This is both a small and approximate correction compared to the area correction. We can conclude that interface charges are extensive, and thus the specific amount of the interface charge can be accounted for by the increased area for the micro-rough interfaces.

Interfacial roughness can also affect FNCOs. Specifically, it has been predicted[109] that if the thermal oxidation of a rough Si surface leads to SiO_2 thickness variations, the amplitude of the FNCOs would decrease. This is due to the fact that electron waves travelling different path lengths through the oxide would result in destructive interference of the propagating wave packets. However, the mechanism for Si thermal oxidation indicates that the reaction between oxygen and Si occurs via inward transport of oxygen to the Si surface where the oxidation takes place.[1] With this

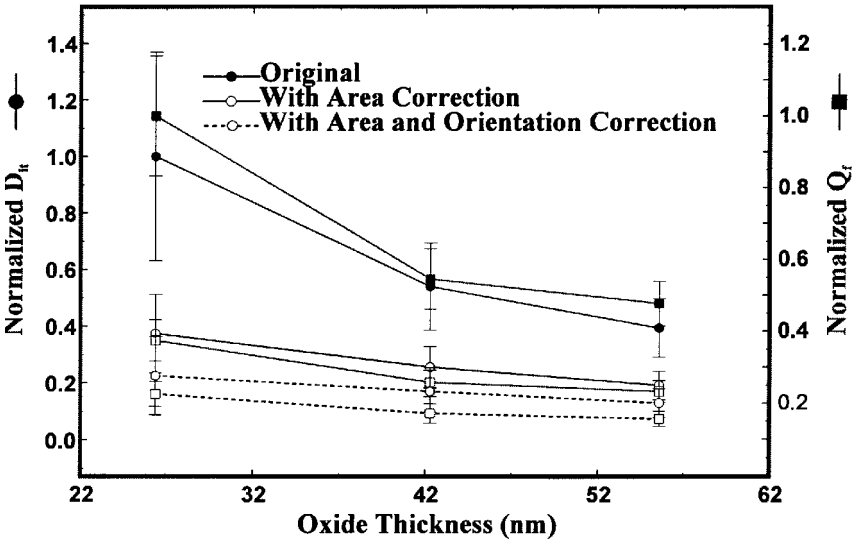


Figure 8.29 Interface trapped charge (D_{it}) and fixed oxide charge (Q_f) changes with the extent of Si thermal oxidation.

mechanism and assuming all the Si atoms at the Si surface have equivalent reactivity, the resulting oxide should be conformal. Hence, no thickness fluctuations would be anticipated as a result of the oxidation of rough Si surfaces. Indeed, our first attempt[98] to measure FNCOs resulting from the oxidation of purposely roughened Si surfaces showed no effect on the FNCO's amplitude. Rough and smooth control samples yielded the same results. However, recently, in the light of our newest results[106] discussed above we have redone this experiment. Specifically, we have attempted to use a very high spatial roughness surfaces, viz. roughened Si surfaces with a high D_F. The idea is that when there are small sharp roughness features that give rise to high D_F values, the local free energy will be increased as will the local chemical reactivity. Thus, the assumption made above about uniform Si atom reactivity will no longer hold. In this case the oxidation will be nonuniform leading to SiO_2 film thickness variations. If the thickness fluctuations are in close proximity, as would be the case for high D_F surfaces, then a diminution of the amplitude of FNCOs would result via the wave interference mechanism. The newest results[110] are summarized in Fig. 8.30 which shows FNCOs from two sample sets (A&B) each set having a smooth control. For sample set A where sample (a) is a smooth control and (b) has $D_F = 2.65$, the extracted FNCOs show no change in FNCO amplitude from the smooth to the rough sample. For sample set B where

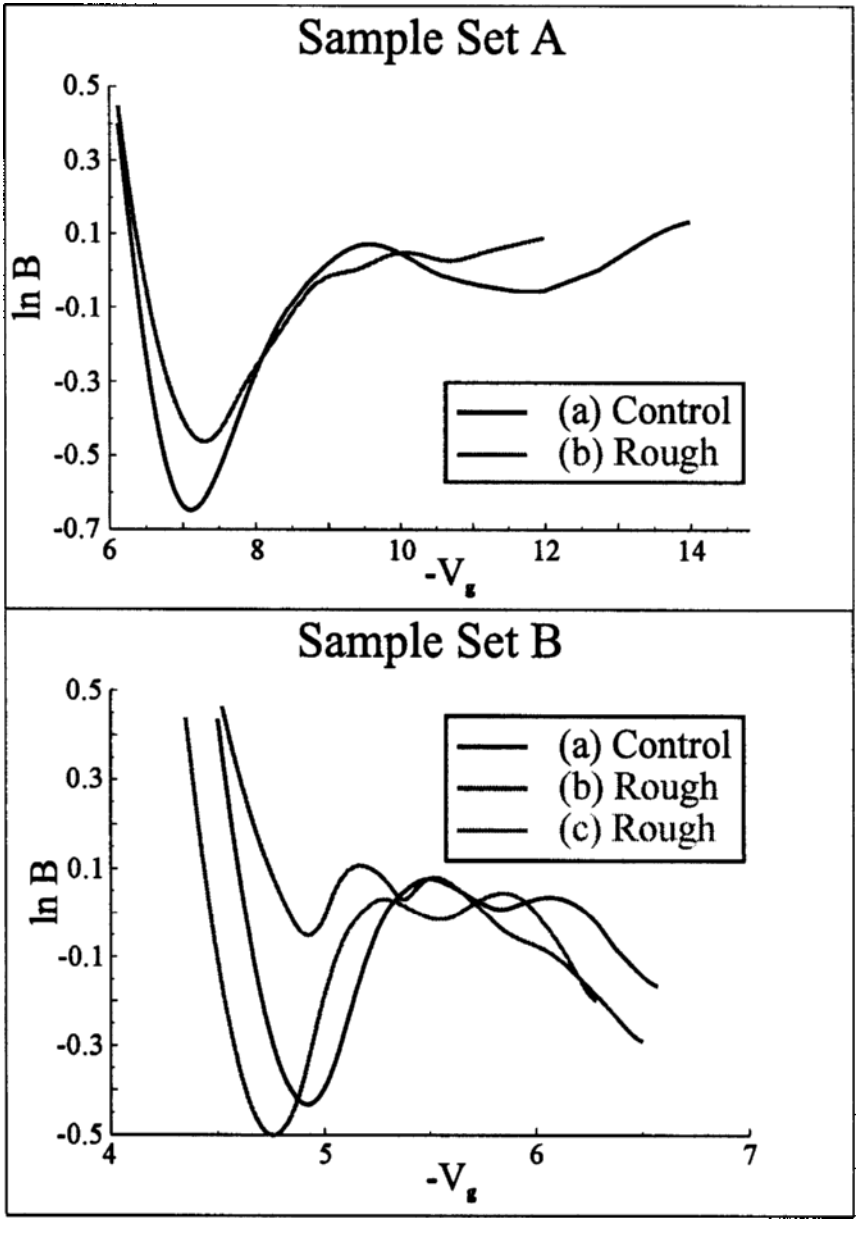


Figure 8.30 The effect of roughness on FNCOs: a) lower frequency roughness and b) high frequency roughness for sample B(c).

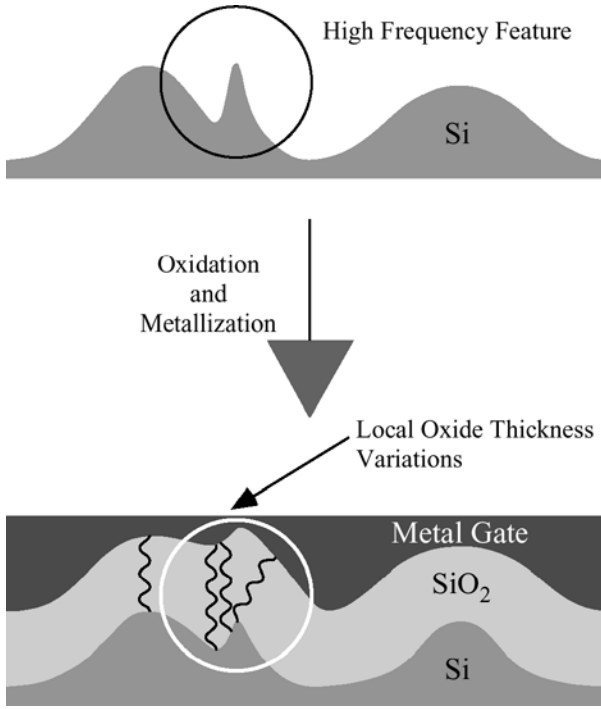


Figure 8.31 Model to explain how high frequency roughness can lead to SiO_2 film thickness fluctuations and decrease resulting FNCOs.

sample (a) is the smooth control, sample (b) has $D_F = 2.71$, and sample (c) has $D_F = 2.52$, the extracted FNCOs show a pronounced decrease in amplitude for sample (b) which has the smallest sharpest features (highest D_F). Figure 8.31 depicts our model for the electron path differences occurring in SiO_2 as a result of thickness fluctuations. Essentially the top of the figure shows a sharp high frequency feature adjacent to smoother features on a rough Si surface. Upon oxidation the sharper feature smoothens, but also results in more oxide formed due to its higher reactivity. Hence a thickness fluctuation occurs which can introduce a path difference and destructive interference for the adjacent wave packets.

8.4 Conclusions

Ellipsometry either alone or in combination with other techniques has significantly impacted every area of SiO_2 film development since the 1960s. This chapter follows the evolution of the understanding of

SiO_2 films in microelectronics emphasizing the role of ellipsometry. Ellipsometry continues to be a crucial characterization technique both in microelectronics and thin film research and in manufacturing. Indeed ellipsometers can be found in virtually every IC chip manufacturing facility around the world. The application of ellipsometry to issues involving SiO_2 films is an ideal venue for the demonstration of all the aspects of this powerful optical technique.

Acknowledgements

The author gratefully acknowledges that the authors research results in this Chapter were performed at the IBM Thomas J. Watson Research Center prior to 1982, and since that time the research was supported by the Office of Naval Research (ONR) and the National Science Foundation (NSF) Materials Research Division.

8.5 References

1. E.A. Irene, "Models for the Oxidation of Silicon,", CRC Critical Reviews in Solid State and Materials Science, Ed. J.E. Greene, Vol 14(2), pp 175–223 (1988).
2. E.A. Irene and R. Ghez, *Appl. Surface Science,* **30**, 1 (1987).
3. E.A. Irene, "Silicon Oxidation: A Process Step for the Manufacture of Integrated Circuits," Ed. P. Stroeve, in "Integrated Circuits," American Chem. Soc., p. 31 (1985).
4. E.A. Irene, *Phil. Mag. B*, **55**, 131 (1987).
5. Handbook of Optical Properties of Solids Volumes I, II and III, Ed. E.D. Palik, Academic Press, Inc., Orlando Florida.
6. G.E. Jellison, *Optical Materials*, **1**, 41 (1992).
7. E.A. Taft, and L. Cordes, *J. Electrochem. Soc.*, **126**, 131 (1979).
8. D.E. Aspnes and J.B. Theeten, *J. Electrochem. Soc.*, **127**, 1359 (1980).
9. S. Chongsawangvirod, E.A. Irene, A. Kalnitsky, S.P. Tay and J.P. Ellul, *J. Electrochem. Soc.*, **137**, 3536 (1990).
10. E.A. Irene, E. Tierney, and J. Angilello, *J. Electrochem. Soc.*, **129**, 2594 (1982).
11. E.A. Irene, D.W. Dong and R.J. Zeto, *J. Electrochem. Soc.*, **127**, 396 (1980).
12. J.K. Srivastava and E.A. Irene, *J. Electrochem. Soc.*, **132**, 2815 (1985).
13. E.A. Irene, *J. Electrochem., Soc.*, **121**, 1613 (1974).
14. E.H. Nicollian, A. Goetzberger, and C.N. Berglund, *Appl. Phys. Lett.*, **15**, 174 (1969); E.H. Nicollian, C.N. Berglund, P.F. Schmidt and J.M. Andrews, *J. Appl. Phys.*, 42, 5654 (1971).
15. E.A. Irene and Y.J. Van der Meulen, *J. Electrochem. Soc.*, **123**, 1380 (1976).
16. E.A. Irene and R. Ghez, *J. Electrochem. Soc.*, **124**, 1757 (1977).
17. H.Z. Massoud, J.D. Plummer, and E.A. Irene, *J. Electrochem. Soc.*, **132**, 1745 (1985).

18. H.Z. Massoud, J.D. Plummer, and E.A. Irene, *J. Electrochem. Soc.*, **132**, 2685 (1985).
19. H.Z. Massoud, J.D. Plummer, and E.A. Irene, *J. Electrochem. Soc.*, **132**, 2693 (1985).
20. R.J. Archer, *J. Electrochem. Soc.*, **104**, 619 (1957).
21. R.J. Archer, *J. Optical Soc. America*, **52**, 970 (1962).
22. R.J. Archer and G.W. Gobeli, *Phys. Chem. Solids*, **26**, 343 (1965).
23. F.L. McCrackin, E. Passaglia, R.R. Stromberg, and H.L. Steinberg, *J. Res. Nat. Bureau Standards- A. Physics and Chem*, **67A**, 363 (1963).
24. F.L. McCrackin and J.P. Colson, NBS Publication 256, 61 (1964).
25. F.L. McCrackin, NBS Technical Note 479, April (1969).
26. R.J. Archer and C.V. Shank, *J. Opt. Soc.,* **57**, (1967).
27. J.R. Ligenza and W.G. Spitzer, *J. Phys. Chem. Solids*, **14**, 131 (1960).
28. B.E. Deal and A.S. Grove, *J. Appl. Phys.*, **36**, 3770 (1965).
29. W.A. Pliskin, *IBM J. Res. Dev.*, **10**, 198 (1966).
30. P.J. Burkhardt and L.V. Gregor, *Trans. Metallurgical Soc. AIME*, **236**, 299 (1966).
31. A.G. Revesz, K.H. Zaininger, and R.J. Evans, *Appl. Phys. Lett.*, **8**, 57 (1966).
32. J.R. Ligenza and W.G. Spitzer, *J. Phys. Chem. Solids*, **14**, 131 (1960).
33. W.A. Pliskin and R.P. Gnall, *J. Electrochem. Soc.*, **111**, 872 (1964).
34. D.E. Aspnes and A.A. Studna, *Appl. Optics*, **14**, 220 (1975).
35. Y.J. van der Meulen and N.C. Hien, *J. Opt. Soc. Am*, **64**, 804 (1974).
36. G.E. Jellison, and F.A. Modine, *J. Appl. Phys.*, **76**, 3758 (1994).
37. E.A. Irene, *Thin Solid Films*, **233**, 96 (1993).
38. Y.Z. Hu, J. Joseph, and E.A. Irene, *Appl. Phys. Lett.*, **59**, 1353 (1991).
39. J. Joseph, Y.Z. Hu, and E.A. Irene, *J. Vac. Sci. Technol. B*, **10**, 611 (1992).
40. E.A. Irene, Chapter 3, "In Situ Real Time Characterization of Surfaces and Thin Film Growth via Ellipsometry in "In Situ Real Time Characterization of Thin Films", Eds. O. Auciello and A. Krauss, John Wiley and Sons, Inc., New York, 2001.
41. E.A. Irene, "*In Situ* Real-Time Characterization of Oxide Film Growth and Other Processes Via Spectroscopic Ellipsometry", Chapter for Volume 28 of the *Annual Review of Materials Science*, accepted for 2001.
42. M.A. Hopper, R.A. Clarke, and L. Young, *J. Electrochem. Soc.*, **122**, 1216. (1975).
43. J.W. Andrews, Y.Z. Hu, and E.A. Irene, SPIE 1989 (Oct.) Symposium on Microelectronic Integrated Processing: Growth Monitoring and Control, Proceedings Jan. 1990.
44. Y.Z. Hu, M. Li, Y. Wang, E.A. Irene, M. Rowe, and H.C. Casey Jr., *Applied Physics Lett.*, **63**, 1113 (1993).
45. Y. Wang, Y.Z. Hu, and E.A. Irene, *J. Vac. Sci. Technol.A*, **12**, 1309 (1994).
46. J. Joseph, Y.Z. Hu, and E.A. Irene, "The Physics and Chemistry of SiO_2 and the Si-SiO_2 Interface," Ed C.R. Helms and B.E. Deal, Plenum, New York, (1993).
47. R.J. Jaccodine and W.A. Schlegel, *J. Appl. Phys.*, **37**, 2429 (1966).
48. M.V. Whelan, A.H. Gormans, and L.M. Goossens, *Appl. Phys. Lett.*, **10**, 262 (1967).

49. E.P. EerNisse, *Appl. Phys. Lett.*, **30**, 290 (1977); **35**, 8 (1979).
50. E.A. Irene, E. Tierney, and J. Angillelo, *J. Electrochem. Soc.*, **129**, 2594 (1982).
51. E. Kobeda and E.A. Irene, *J. Vac. Sci. Technol. B*, **4**, 720 (1986); *J. Vac. Sci. Technol. B*, **5**, 15 (1987).
52. E.A. Taft, *J. Electrochem. Soc.*, **125**, 968 (1978).
53. W.A. Tiller, *J. Electrochem. Soc.*, **127**, 619, 625 (1980).
54. G. Lucovsky, M.J. Mantini, J.K. Srivastava, and E.A. Irene, *J. Vac. Sci. Technol. B*, **5**, 530 (1987).
55. E.A. Irene, D. Dong, and R.J. Zeto, *J. Electrochem. Soc.*, **127**, 396 (1980).
56. H. Yao, J.A. Woollam, and S.A. Alterovitz, *Appl. Phys. Lett.*, **62**, 3324 (1993).
57. R. Williams and A.M. Goodman, *Appl. Phys. Lett.*, **25**, 531 (1974).
58. T.W. Sigmon, W.K. Chu, E. Lugujjo, and J.W. Mayer, *Appl. Phys. Lett.*, **24**, 105 (1974).
59. A.M. Stoneham and P.W. Tasker, J. Phys. C: *Solid State Phys.*, **18**, L 543 (1985): A.M. Stoneham and P.W. Tasker, *Semicond. Sci. Technol.*, **1**, 93 (1986).
60. J. Derrien and Commandre, *Surface Science*, **118**, 32 (1982).
61. S.I. Raider and R. Flitsch, *J. Vac. Sci. Technol.*, **13**, 58 (1976).
62. C.R. Helms, *J. Vac. Sci. Technol.*, **16**, 608 (1979).
63. F.J. Grunthaner, P.J. Grunthaner, R.P. Varquez, B.F. Lewis, J. Maserjian, and A. Madhukar, *J. Vac. Sci. Technol.*, 16, 1443 (1979).
64. H. Ibach, H.D. Bruchmann, and H. Wagner, *Appl. Phys. A*, **29**, 113 (1982).
65. P. Chiaradia and S. Nannarone, *Surface Science*, **54**, 547 (1976).
66. E.A. Irene and D. Dong, *J. Electrochem. Soc.*, **125**, 1146 (1978).
67. H.Z. Massoud, J. Plummer, and E.A. Irene, *J. Electrochem. Soc.*, **132**, 1745 (1985).
68. E.A. Lewis and E.A. Irene, *J. Electrochem. Soc.*, **134**, 2332 (1987).
69. E.A. Irene, *J. Electrochem. Soc.*, **125**, 1708 (1978).
70. J.R. Ligenza, *J. Phys. Chem.*, **65**, 2011 (1961).
71. E.A. Irene, H.Z. Massoud, and E. Tierney, *J. Electrochem. Soc.*, **133**, 1253 (1986).
72. K. Ueda and M. Inoue, *Surf. Sci.*, **161**, L578 (1985).
73. R. Oren and S.K. Ghandi, *J. Appl. Phys.*, **42**, 752 (1971).
74. S.A. Schafer and S.A. Lyon, *J. Vac. Sci. Technol.*, **19**, 494 (1981); S.A. Schafer and S.A. Lyon, *J. Vac. Sci. Technol.*, **21**, 422 (1982).
75. Ian W. Boyd, *Appl. Phys. Lett.*, 42, 728 (1983); Ian W. Boyd, "Surface Studies with Lasers," Eds. F.R. Aussenberg, A. Leitner and M.E. Lippitech, Springer-Verlog 1983, p. 193.
76. E.M. Young and W.A. Tiller, *Appl. Phys. Lett.*, **42**, 63 (1983); E.M. Young and W.A. Tiller, *Appl. Phys. Lett.*, **50**, 80 (1987).
77. E.A. Irene and E.A. Lewis, *Appl. Phys. Lett.*, **51**, 767 (1987).
78. L.M. Chanin, A.V. Phelps, and M.A. Biondi, *Phys. Rev.*, **128**, 219 (1962).
79. R. Ghez and Y.J. van der Meulen, *J. Electrochem. Soc.*, **119**, 1100 (1972).
80. E.A. Irene, *Appl. Phys. Lett.*, **40**, 74 (1982).
81. E.A. Irene, *J. Appl. Phys.*, **54**, 5416 (1983).
82. A. Fargeix, G. Ghibaudo and G. Kamarinas, *J. Appl. Phys.*, **54**, 2878 (1983); **54**, 7153 (1983); **56**, 589 (1984).

Next Page

83. G. Camera Roda, F. Santarelli, and G.C. Sarti, J. *Electrochem. Soc.*, **132**, 1909 (1985).
84. R.H. Doremus, *Thin Solid Films*, **122**, 191 (1984).
85. F.J. Grunthaner and P.J. Grunthaner, "Chemical and Electronic Structure of the SiO_2/Si Interface," Materials Science Reports 1987.
86. W.A. Tiller, *J. Electrochem. Soc.*, **128**, 689 (1981).
87. F. Herman, I.P. Batra and R.V. Kasowski, in "The Physics of SiO_2 and Its Interfaces," Ed. S.T. Pantelides, p. 333, Pergamon, N.Y. 1979.
88. B. Agius, S. Rigo, F. Rocket, M. Froment, C. Maillot, H. Roulet, and G. Dufour, *Appl. Phys. Lett.*, **44**, 48 (1984).
89. F. Rochet, B. Agius, and S. Rigo, *J. Electrochem. Soc.*, **131**, 914 (1984).
90. V.A. Yakovlev and E.A. Irene, *J. Electrochem. Soc.*, **139**, 1450 (1992).
91. V. A. Yakovlev, Q. Liu, and E. A. Irene, *J. Vac. Sci. Technol. A*, **10**, 427 (1992).
92. E.A. Taft and L. Cordes, *J. Electrochem. Soc.*, **126**, 131 (1979).
93. A. Kalnitsky, S.P.Tay, J.P. Ellul, S. Chongsawangvirod, J.W. Andrews, and E.A. Irene, *J. Electrochem. Soc.,* **137**, 234, (1990).
94. D.F. Mitchell, K.B. Clark, J.A. Bardwell, W.N. Lennard, G.R. Massoumi, and L.V. Mitchell, *Surface and Interface Analysis*, **21**, 44 (1994).
95. K.H. Gundlach, *Solid State Electron.*, **9**, 949 (1996).
96. M.E. Alferieff and C.B. Duke, *J. Chem. Phys.*, **46**, 938 (1967).
97. J. Maserjian, *The Physics and Chemistry of SiO_2 and Si-SiO_2 Interface,* Ed C.R. Helms and B.E. Deal, Plenum, New York , 505 (1988).
98. S. Zafar, Q. Liu, and E.A. Irene, *J. Vac. Sci. Technol. A*, **13(1)**, 47 (1995); S. Zafar, K.C. Conrad, Q. Liu, E.A. Irene and G.Hames, R. Kuehn, and J.J. Wortman, *Appl. Physics Lett.*, **67**, 1031 (1995).
99. K.J. Hebert, S. Zafar, E.A. Irene, R Kuehn, T.E. McCarthy, and E.K. Demirlioglu., *Appl. Phys. Lett.*, **68**, 266 (1996).
100. K.J. Hebert, T. Labayen, and E.A. Irene, *Physics and Chemistry of SiO_2 and the Si-SiO_2 Interface III*, Eds H.Z. Massoud, C.R. Helms and E.H. Poindexter, 81 (1996), The Electrochemical Soc. Inc., New Jersey, USA.
101. Y. Wang and E.A. Irene, *J. Vac. Sci. Technol. B*, **18(1)**, 279, (2000).
102. L. Spanos and E.A. Irene, *J. Vac. Sci. Technol. A*, **12(5)**, 2646 (1994).
103. L. Spanos, Q. Liu, T. Zettler, B. Hornung, J.J. Wortman, and E.A. Irene, *J. Vac. Sci. Technol. A*, **12(5)**, 2653, (1994).
104. Q. Liu, L. Spanos, C. Zhao, and E.A. Irene, *J. Vac Sci. Technol. A*, **13**, 1977 (1995).
105. C. Zhao, P.R. Lefebvre, and E.A. Irene, *Thin Solid Films*, **313–414**, 286 (1998).
106. L. Lai and E.A. Irene, *J. Appl. Phys.*, **86(3)**, 1729 (1999).
107. L.Lai, and E.A. Irene, *J. Vac. Sci. Technol. B*, **17**, 33 (1999).
108. L.Lai, K.J. Hebert, and E.A. Irene, *J. Vac. Sci. Technol. B*, **17**, 53 (1999).
109. J. Sune, I Placencia, E. Farras, N. Barniol, and X. Aymerich, *Phys. Status Solidi*, **109**, 496 (1988).
110. L. Lai and E.A. Irene, *J. Appl. Phys.*, **87(3)**, 1159 (2000).

9 Theory and Application of Generalized Ellipsometry

Mathias Schubert

Fakultät für Physik und Geowissenschaften
Institut für Experimentelle Physik II
Universität Leipzig, Germany

9.1 Introduction

Spectroscopic ellipsometry has become a standard method among the polarization-dependent optical techniques to explore the optical properties of solids and liquids.[†] However, until today most of the results obtained with ellipsometry are reported from isotropic materials. Investigations on arbitrarily anisotropic layered systems or samples with geometrical structures of high complexity are mainly restricted to data collection and qualitative analysis. This is because of

(1) the complex formulas associated with highly anisotropic systems, and
(2) standard ellipsometry is restricted to isotropic surfaces.

A convenient mathematical formalism is necessary for calculation of the optical response of anisotropic layered structures. Likewise, appropriate experimental setups are required which determine non-redundant optical parameters from a general anisotropic sample. Generalized ellipsometry (GE) comprises theory and experiment of anisotropy in layered samples. Owing to its unique capability to characterize optical and structural properties of general anisotropy, GE enables new insights into physical phenomena of novel thin-film materials. This chapter summarizes recent GE developments and applications. The GE concept and the Jones calculus are outlined. Data acquisition procedures and strategies for measurement of anisotropic sample systems are presented and discussed. The GE theory,

[†]For a comprehensive review of the state-of-the-art data analysis in spectroscopic ellipsometry see the Chapter by G.E. Jellison Jr. in this book. See also *Spectroscopic Ellipsometry,* edited by A.C. Boccara, C. Pickering, J. Rivory (Elsevier, Amsterdam, 1993), and *Spectroscopic Ellipsometry,* edited by R.W. Collins, D.E. Aspnes, E.A. Irene (Elsevier, Lausanne, 1998), Proceedings of the first and second International Conferences on Spectroscopic Ellipsometry.

i.e., a calculus for wave propagation in arbitrary anisotropic layered media, is given. This calculus spans the symmetric and nonsymmetric (magneto-optical) dielectric and twisted symmetric dielectric materials. The GE parameters for layered systems of any combination of the materials within the calculus are obtained using the 4×4 matrix formalism. Recently reported applications of theory and experiment to single-crystal non-cubic solids (e.g., ordered-vacancy compounds), continuously twisted biaxial media (e.g., chiral liquid crystals), mixed-phase solid solutions (e.g., poly-crystalline hard coating materials), free-carrier magneto-optic phenomena, and superlattice-type CuPt order-birefringence in semiconductor materials are included to demonstrate the ability of GE for analysis of complex multilayered samples with inherent and arbitrarily oriented anisotropy.

9.2 The Generalized Ellipsometry Concept

The basic principle of the GE concept is to seek tools that enable the determination *and* analysis of non-redundant optical sample parameters. The keyword between experiment and theory is the so-called Jones matrix. The reader should be aware of the difference between the concept of the Jones algorithm, which describes *measurable* electromagnetic field components, and the 4×4 matrix algebra, which treats the electromagnetic field components *within* the sample. The 4×4 formalism is devoted to details of the internal sample polarization-modifying processes responsible for the external polarizing sample properties. The analytical approach to solving for the light propagation in complex optical systems is to calculate the Jones matrices for arbitrarily anisotropic homogeneous layered systems. This formalism will be outlined in Sections 9.3 and 9.4. A brief introduction of the phrases Jones vector and Jones matrix is given in Section 9.2.2. (See also Chapters 1 and 3, in this book). For a detailed introduction into the Jones vector approach, refer to Azzam and Bashara's textbook. The GE parameters are defined in Section 9.2.3. A feasible experimental setup for acquisition of GE parameters is explained in Section 9.2.4, where the GE extension of rotating-analyzer ellipsometry (RAE) is discussed. To begin with here, the important issue of our GE notation and data format definition is discussed.

9.2.1 Comments on Notations in GE

The issue concerning notation in generalized ellipsometry needs to be addressed here. We need to normalize the ellipsometry equations because the absolute intensity is difficult to measure with the same precision as

the amplitude ratios and their relative phase information. For almost any choice of normalization a solution with a zero denominator can be found for a physical real interesting sample. For instance, any of the Jones reflection or transmission elements may approach zero for certain anisotropic situations. Jellison and others have followed the choice of Azzam and Bashara, and divided the Jones matrix elements by r_{ss} (or t_{ss}).[1, 9, 13] For some common anisotropic samples this definition produces a helpful visualization of the sample symmetry. In this chapter we define the GE quantities $R_{pp} = r_{pp}/r_{ss}$, $R_{ps} = r_{ps}/r_{pp}$, and $R_{sp} = r_{sp}/r_{ss}$ (the transmission parameters follow accordingly). This procedure seems more natural from a data acquisition standpoint. To measure the Jones matrix elements in the "*p*, *s*" notation one can illuminate the sample with *p*- and then once again with *s*-polarized light. For incident *p*-polarized light one can measure what emerges as *p*-, and how "much" does convert into *s*-polarized components. Both "channels" can be tested without changing the input condition. The result is then expressed as "*p*-to-*s*" divided by "*p*-to-*p*", most conveniently to subtract the unknown intensity of the incident beam from the experimental result. The same approach results in the ratio "*s*-to-*p*" divided by "*s*-to-*s*". Similarly, for the use of a rotating-analyzer (polarizer) ellipsometer with adjustable polarizer (analyzer) the ratios of the elements in the Jones matrix columns (rows) can be completely isolated at certain polarizer (analyzer) positions. This provides a better connection between measurement and reported result, which was our initial motivation. The parameter choice also constitutes an excellent experimental access to initial starting values for the GE parameters during the regression analysis of the sets $\{\alpha, \beta\}$. It also implies that these polarizer (analyzer) positions are always included within a GE data acquisition scan.

Our parameter choice further allows direct access to quantities of anisotropic samples when depolarization due to incoherent beam propagation prevents the search of an effective Jones matrix. (See also Section 9.5.2. A simple way to look at the response of such a sample is to imagine that multiple samples with different Jones matrices are illuminated simultaneously.) In this so-called unregressed mode, the sets $\{\alpha, \beta\}$ measured at $P = 0°$, 45°, and 90° transform directly into *effective* GE parameters R^*_{pp}, R^*_{ps} and R^*_{sp}. (Note that, however, the effective GE parameters do not share a common Jones matrix!) This numerical inversion is based on the behavior of the RAE coefficients $(\alpha, \beta) \sim (R_1/R_2)$, which read for an *ideal* Jones matrix ($C = w = 0°$; see Section 9.2.4.)

$$\begin{aligned}
(R_1/R_2)\{P = 0°\} &= (1/R_{ps}),\\
(R_1/R_2)\{P = 90°\} &= (R_{sp}),\\
(R_1/R_2)\{P = 45°\} &= (R_{pp} + R_{sp})/(1 + R_{pp}R_{ps})/\sqrt{2}.
\end{aligned} \tag{9.1}$$

The effective GE parameters obtained by the above equations using sets (α, β) measured at, e.g., thick transparent anisotropic substrates, are still somewhat more similar in their behavior than the GE ratios formed by an ideal Jones matrix. The coefficients R^*_{pp}, and R^*_{ps} result from a measurement where the incident polarization was bound to the *p*-plane, and R^*_{sp} is obtained with the polarizer set to the *s*-plane. There is also weighted information from the sample response when the polarizer is set evenly in between. The effective ("unregressed") GE parameters represent the target data for the subsequent best-fit for the optical constants, crystal orientations and so on, by modeling the effect of the "multiple" Jones matrices at the polarizer positions $P = 0°$, 45°, and 90°.

Our choice of GE parameters is mostly driven by real needs to perform feasible measurement procedures, and to provide reasonable data formats. It has turned out to be very effective during data acquisition and analysis of real anisotropic sample situations (see Section 9.6). The author therefore encourages other users to follow the GE notation defined in this chapter. It is further important to note that Azzam and Bashara defined the off-diagonal ratios using an alternate order of their indices. In Azzam's notation, the first index denotes the outgoing polarization whereas the second index denotes the incident polarization. For instance, the Jones reflection matrix reads by Azzam and Bashara[1]

$$\begin{bmatrix} r_{pp} & r_{ps} \\ r_{sp} & r_{ss} \end{bmatrix}. \tag{9.2}$$

Jellison and others have adopted this notation.[9, 13] Care must be therefore taken when comparing data expressed using the above notation with data presented in the notation used in this chapter. The index order used here, where the first index addresses the incident, and the second index denotes the exit mode was actually adopted from the notation given in Yeh's textbook [P. Yeh, *Optical Waves in Layered Media*, Wiley Series in Pure and Applied Optics (New York: Wiley, 1988).]

9.2.2 The Optical Jones Matrix

A common property of all vector waves is their polarization. This dependence with time of one of the wave field vectors is observed at a fixed point in space. The electric field vector ***E*** of a monochromatic electromagnetic plane wave is commonly used as the polarization vector. Assuming the uniform plane wave propagates along the positive direction of the *z*-axis of an orthogonal, right-handed, Cartesian coordinate system,

the polarization vector can be represented as follows:

$$\vec{E}(z,t) = \left[\tilde{E}_x \cos\left(\omega t - \frac{2\pi}{\lambda} z + \delta_x\right)\right]\vec{x} + \left[\tilde{E}_y \cos\left(\omega t - \frac{2\pi}{\lambda} z + \delta_y\right)\right]\vec{y}. \quad (9.3)$$

$\tilde{E}_x$ and $\tilde{E}_y$ represent the amplitudes of the linear harmonic oscillations of the electric field components along the x and y-axis. δ_x and δ_y are the respective retardations at a fixed spatial position z. A more concise mathematical description is obtained by suppressing temporal information such as the harmonic time dependence. A 2×1 column vector groups the scalar components of the latter equation

$$\vec{E} = \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \hat{E}_x \exp\left\{-i\frac{2\pi}{\lambda} z\right\} \exp\{i\delta_x\} \\ \hat{E}_y \exp\left\{-i\frac{2\pi}{\lambda} z\right\} \exp\{i\delta_y\} \end{pmatrix}, \quad \tilde{E}_{x,y} = Re[\hat{E}_{x,y}]. \quad (9.4)$$

The spatial information may be dropped considering the field over one fixed transversal plane, e.g., at $z = 0$. The remaining vector is the so-called Jones vector. This vector contains full information about amplitude and phase of the field components. The coordinates of the Jones vector depend on the definition of the Cartesian coordinate axes and the plane of representation. Note that the Jones vector renders a plane wave, but does not represent the plane wave itself. As a consequence of the omitted time dependence, the vector $\boldsymbol{E}$ defined in Eq. (9.4) permits the treatment of completely polarized plane waves only. The phenomenon of partially or completely depolarized light may be described by reintroducing the time information of the complex amplitudes E_x and E_y, leading to the concept of the Mueller-Jones matrices. (See also chapters by Jellison in this book).

The polarization of the incident and emerging plane waves represents the fundamental properties that allow for optical characterization of samples by GE. As the wave is processed through the ellipsometer device, each of the device components also modifies the polarization state of the light beam. The Jones calculus effectively describes such device interactions and is used in Section 9.2.4. The following restrictions apply here:

(*i*) The light beam is a monochromatic plane wave.
(*ii*) All polarization modifying interactions are linear and frequency conserving.
(*iii*) The incident light beam is fully polarized.

A single incident plane wave may cause one or more modified plane waves to emerge from a sample (Fig. 9.1). All plane waves are associated

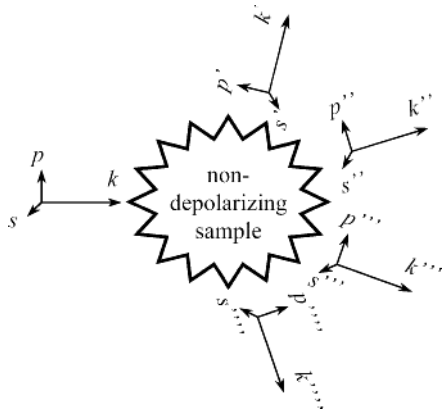


Figure 9.1 Incident and emergent electromagnetic plane waves at a non-depolarizing optical system. A Jones matrix connects the incident mode with each of the emergent plane wave components with respect to a chosen coordinate system, e.g., the *p* and *s* planes in Fig. 9.2.

with right-handed, space-fixed Cartesian coordinate systems. For each of the emergent plane waves a 2 × 2 complex matrix can be found which connects the two vector components (modes) of the incident wave to those from the emergent waves. We choose a Cartesian coordinate system common to both incident and emerging plane waves, with the *p* and *s* planes as in Fig. 9.2. The plane of incidence is defined through the wave vectors of the incident wave and the emerging wave. Their *p* and *s* modes are related by a matrix **j**:

$$\begin{pmatrix} B_p \\ B_s \end{pmatrix} = \mathbf{j}\begin{pmatrix} A_p \\ A_s \end{pmatrix} = \begin{bmatrix} j_{pp} & j_{sp} \\ j_{ps} & j_{ss} \end{bmatrix}\begin{pmatrix} A_p \\ A_s \end{pmatrix}, \tag{9.5}$$

where A_p, A_s, B_p, and B_s denote the *p* and *s* components of the incident and emerging plane waves, respectively. The matrices **j**'s are so-called Jones matrices. These matrices depend on

- The respective optical system,
- The frequency of the incident plane wave (incident photon energy),
- The sample orientation with respect to the incident wave propagation direction,
- The emergent plane wave under consideration, and
- The coordinate systems of both the incident as well as the emergent plane waves.

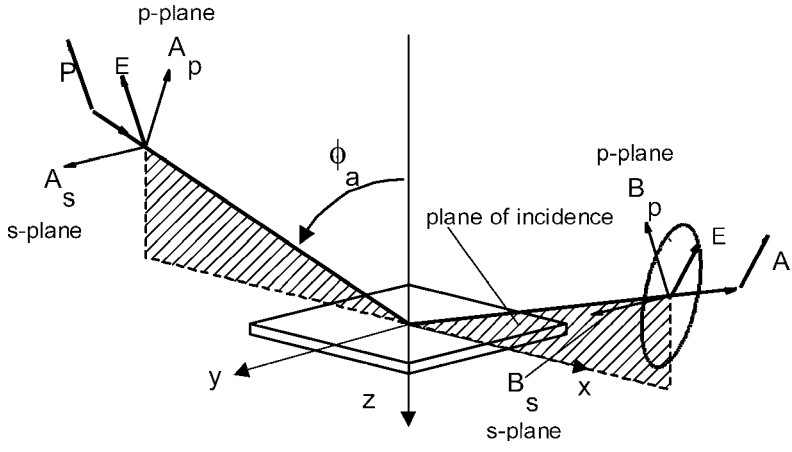


Figure 9.2 Definition of the plane of incidence (*p* plane) and the incidence angle Φ_a through the wave vectors of the incident and emerging (reflected here) plane waves. A_p, A_s, B_p, and B_s, denote the complex amplitudes of the *p* and *s* modes before and after reflection, respectively. *P* and *A* are the azimuth angles of the linear polarizer used, e.g., in the standard arrangement of rotating-analyzer (polarizer) ellipsometer. *P*, or *A* are equal to zero if their preference directions are parallel to the *p* plane. (Both *P* and *A* rotate clockwise with respect to the light propagation.)

Jones matrices describe total changes to the state of polarization caused by an optical system. The term "optical system" can be assigned to any arbitrary optical device or any arbitrary sample as long as the assumptions (*i*) through (*iii*) are true.

9.2.3 The Generalized Ellipsometry Parameters

Classes of samples with plane parallel boundaries are considered only. The sample surface is the (x, y) plane, and x is parallel to the plane of incidence. z points towards the sample. The origin is set at the sample surface (see Figs. 9.2 and 9.4). A single reflected and a single transmitted plane wave are expected upon a single incident beam.[†] The Jones matrix in Eq. (9.5) equals the reflection matrix $\mathbf{r}$

$$\begin{pmatrix} B_p \\ B_s \end{pmatrix} = \mathbf{r}\begin{pmatrix} A_p \\ A_s \end{pmatrix} = \begin{bmatrix} r_{pp} & r_{sp} \\ r_{ps} & r_{ss} \end{bmatrix}\begin{pmatrix} A_p \\ A_s \end{pmatrix}, \tag{9.6}$$

[†]Restriction to plane parallel sample boundaries is not a sufficient condition to have a single pair of transmitted and reflected beams only. Buried layers may have corrugated interfaces and may cause higher-order diffracted beams, which will also emerge from the sample but under different propagation directions.

or the transmission matrix $\boldsymbol{t}$

$$\begin{pmatrix} C_p \\ C_s \end{pmatrix} = \mathbf{t}\begin{pmatrix} A_p \\ A_s \end{pmatrix} = \begin{bmatrix} t_{pp} & t_{sp} \\ t_{ps} & t_{ss} \end{bmatrix}\begin{pmatrix} A_p \\ A_s \end{pmatrix}, \tag{9.7}$$

where C_p, C_s denote the modes of the transmitted beam. To identify the meaning of each matrix element consider for example the elements r_{ss} and t_{sp}. If the incident wave is only s polarized, i.e., $A_p = 0$, the elements r_{ss} and t_{sp} are simply given by:[†]

$$r_{ss} \equiv \left(\frac{B_s}{A_s}\right)_{A_p=0}, \text{ and } t_{sp} \equiv \left(\frac{C_p}{A_s}\right)_{A_p=0}. \tag{9.8}$$

Knowing the respective incident and exit modes would allow for calculation of each individual Jones matrix element. However, the Jones vector components are complex scalars that represent the phase and amplitude of the electromagnetic field components within their reference plane. Single electric field-intensity measurements in general suppress this phase information due to the high angular velocity of the harmonic electric field oscillation. As compared to pure intensity measurements GE can recover some of this phase information, and three relative phases are measurable. Diagonal Jones matrices are related to structures that reflect or transmit p and s polarized light into p and s polarized light only, respectively. In the general case the Jones matrix has non-vanishing off-diagonal elements. These elements provide the cross-polarization properties of the sample, i.e., the ability to progress incident p or s polarized light into emerging s and (or sometimes only) p polarized light, respectively. Standard (isotropic) ellipsometry is restricted to surfaces that do not cause cross-polarization effects. Azzam and Bashara devoted the application of GE to nondepolarizing systems with non-diagonal Jones matrices.[1] They considered Eqs. (9.6) and (9.7) as bilinear transformations between the complex *ratios* of the incident and the emergent plane-wave amplitudes.[2] A linear, nondepolarizing sample can be categorized by use of the coefficients of this bilinear transformation. In fact, these coefficients are equal to Jones matrix elements normalized to a common factor.[3, 4] Let us now introduce the GE parameters. To begin with we define the complex reflectance (transmittance) ratio ρ (τ) as usual ($\chi = A_p/A_s$)

$$\rho \equiv \tan\Psi \exp\{i\Delta\} \equiv \left(\frac{A_p}{A_s}\right)^{-1}\left(\frac{B_p}{B_s}\right), \tau \equiv \tan\Psi \exp\{i\Delta\} \equiv \left(\frac{A_p}{A_s}\right)^{-1}\left(\frac{C_p}{C_s}\right), \tag{9.9}$$

[†]In our notation the first index denotes the incident mode, and the second index refers to the outgoing polarization. See also Sect. 9.2.1.

which depends for non-vanishing off-diagonal matrix elements on the ratio of the incident wave amplitudes χ and the off-diagonal elements themselves

$$\rho = \frac{r_{pp} + r_{sp}\chi^{-1}}{r_{ss} + r_{ps}\chi}, \qquad \tau = \frac{t_{pp} + t_{sp}\chi^{-1}}{t_{ss} + t_{ps}\chi}, \tag{9.10}$$

or in a slightly different form

$$\rho = \frac{r_{pp}/r_{ss} + (r_{sp}/r_{ss})\chi^{-1}}{1 + (r_{ps}/r_{pp})(r_{pp}/r_{ss})\chi}, \qquad \tau = \frac{t_{pp}/t_{ss} + (t_{sp}/t_{ss})\chi^{-1}}{1 + (t_{ps}/t_{pp})(t_{pp}/t_{ss})\chi}. \tag{9.11}$$

The complex ratios ρ and τ in Eqs. (9.11) are bilinear transformations with three complex parameter ratios formed by the elements of the Jones matrices. The idea of GE is to define and to determine three linear independent normalized quantities for a given sample azimuth φ, angle of incidence Φ_a, and incident photon energy E. The sample azimuth φ is an arbitrary angle between the x-axis of the laboratory system and a direction parallel to the sample surface. It is often convenient to choose a crystallographic sample direction parallel to the (x-y) plane. The following sets of normalized Jones matrix elements are defined here as the GE parameters for reflection

$$\begin{aligned}
\frac{r_{pp}}{r_{ss}} &\equiv R_{pp} = \tan\Psi_{pp} e^{i\Delta_{pp}}, \\
\frac{r_{ps}}{r_{pp}} &\equiv R_{ps} = \tan\Psi_{ps} e^{i\Delta_{ps}}, \\
\frac{r_{sp}}{r_{ss}} &\equiv R_{sp} = \tan\Psi_{sp} e^{i\Delta_{sp}},
\end{aligned} \tag{9.12}$$

and for transmission

$$\begin{aligned}
\frac{t_{pp}}{t_{ss}} &\equiv T_{pp} = \tan\Psi_{pp} e^{i\Delta_{pp}}, \\
\frac{t_{ps}}{t_{pp}} &\equiv T_{ps} = \tan\Psi_{ps} e^{i\Delta_{ps}}, \\
\frac{t_{sp}}{t_{ss}} &\equiv T_{sp} = \tan\Psi_{sp} e^{i\Delta_{sp}}.
\end{aligned} \tag{9.13}$$

This results in Eqs. (9.11)

$$\rho = \frac{R_{pp} + R_{sp}\chi^{-1}}{1 + R_{pp}R_{ps}\chi}, \qquad \text{or} \qquad \tau = \frac{T_{pp} + T_{sp}\chi^{-1}}{1 + T_{pp}T_{ps}\chi}. \tag{9.14}$$

Note that the diagonal elements from the same column of the Jones matrix are used to normalize the matrix elements. This choice is convenient for Rotating-Analyzer ellipsometry. For a rotating-polarizer ellipsometer the choice would be to normalize the off-diagonal elements by the on-diagonal elements from the same row (see Section 9.2.1). The real-valued quantities Ψ_{pp}, Ψ_{ps}, Ψ_{sp}, Δ_{pp}, Δ_{ps}, Δ_{sp} now comprise the GE data format. This definition is not an end in itself and may be changed by other users. The three complex quantities chosen here are independent, and all other choices may be expressed as a linear combination of this basic set. In this notation the complex off-diagonal ratios R_{ps} (T_{ps}) and R_{sp} (T_{sp}) behave similarly if the polarization transfer functions for the *s* and *p* polarized light are alike. This situation is observed for, e.g., chiral liquid crystal cells in normal incidence transmission, or in certain configurations from a uniaxial bulk or film. Depending on the sample properties and their orientation during the GE experiment, the off-diagonal normalized Jones matrix elements may be symmetrical or antisymmetrical, Hermitean or anti-Hermitean, completely different or simply zero (see, e.g., Refs. 5–7, and references therein). For small off-diagonal ratios the on-diagonal ratio R_{pp} (T_{pp}) is very similar to the complex ratio determined by standard ellipsometry. In particular, if R_{ps} (T_{ps}) and R_{sp} (T_{sp}) vanish, R_{pp} (T_{pp}) is identical to the standard ellipsometric ratio ρ defined in Eq. (9.9). Note that one can always express ρ or τ in terms of $\tan\Psi$ and $\cos\Delta$ for any set of GE parameters regardless of χ. Note further that Ψ and Δ will depend on χ if the off-diagonal ratios do not vanish. The latter becomes important if the attempt is made to characterize anisotropic surfaces by standard ellipsometry. Some devices automatically perform zone-averaged polarizer measurements at positions mirrored by the plane of incidence to cancel out systematic errors. Because Eqs. (9.14) are not symmetrical in χ, zone-averaged data will be useless for further analysis. Zone-averaging cannot be used to account for slight analyzer or polarizer calibration offsets because such offsets virtually rotate the sample Jones matrix producing virtual off-diagonal elements. Precise calibration of rotating-element ellipsometers is one of the prerequisites to obtain correct GE data. Furthermore, most often in standard ellipsometry the polarizer position is tracked, and set to the actual Ψ value for highest sensitivity to α and β in RAE or rotating polarizer ellipsometry (RPE). Appropriate consideration of this "polarizer tracking" will also entangle subsequent data analysis as

χ was set to $\tan^{-1}\rho$ (or $\tan^{-1}\tau$) during data acquisition, but ρ (or τ) to be modeled also depends on χ, and ambiguous fit results may be obtained.

9.2.4 Generalized Ellipsometry Acquisition Techniques

Many authors reported on extension of standard ellipsometry techniques to GE capabilities.[3, 8–14] Null ellipsometry was used by Azzam and Bashara in their original GE paper.[1] Hauge measured uniaxially anisotropic surfaces with Generalized rotating-compensator ellipsometry (RCE).[3] A rotating-analyzer-based setup was employed in, e.g., Refs. 8, 15–20. Jellison utilized generalized two-modulator ellipsometry.[9, 21–23] En Naciri applied a Fourier-expanded RCE technique,[13] and Thompson *et al*. extended FTIR-ellipsometry to determine the GE parameters for infrared wavelength.[10] The extension of the well-known RAE technique is chosen here for the purpose of demonstration. The fundamentals of RAE are often described and discussed. Nevertheless, a short review is given here so that the reader can follow the extension of RAE to GE. The reflection arrangement is assumed here. The considerations in the following paragraph are valid for the transmission case as well. For this one may simply replace the reflection matrix elements by their respective transmission matrix elements. The RPE technique is in principle identical to RAE. The derivation of the GE extension for RPE is straightforward, and follows the same line as given below for the RAE technique.

A completely non-polarized monochromatic light beam becomes linearly polarized upon passing through an ideal linear polarizer P (Fig. 9.2). An ideal compensator with variable azimuth C and phase retardation w produces an arbitrary ellipticity (The compensator is not included in Fig. 9.2). A second ideal linear polarizer (analyzer) is used to determine the state of polarization of the light beam, which is reflected from the sample surface. The intensity of the light beam after passing through the analyzer can be expressed as a function of the reflection matrix $\mathbf{r}$, and the polarizer, compensator, and analyzer azimuth angles P, C, and A, respectively, and phase retardation w (P, C, or A are equal to zero if their azimuths are parallel to the plane of incidence. P, C, and A move counterclockwise. The axis of rotation is the beam propagation direction.) If $\boldsymbol{E}^{i}$ and $\boldsymbol{E}^{det}$ denote the electric field components of the polarizer incident beam and the detected beam, respectively, the polarization state transfer is described as a successive Jones matrix multiplication

$$\mathbf{E}^{\mathrm{det}} = \mathbf{R}_{A}\Pi\mathbf{R}_{-A}\,\mathbf{R}_{C}\mathbf{w}\mathbf{R}_{-C}\mathbf{r}\mathbf{R}_{P}\Pi\mathbf{R}_{-P}\mathbf{E}^{\mathrm{i}}, \tag{9.15}$$

where Π and $\boldsymbol{w}$ are the linear polarizer and ideal compensator matrices, respectively. $\boldsymbol{R}_P$, $\boldsymbol{R}_C$, and $\boldsymbol{R}_A$ rotate, respectively, P, C, and A within their reference planes

$$\Pi = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{w} \equiv \begin{pmatrix} \exp(iw) & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{R}_\alpha \equiv \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix}. \tag{9.16}$$

We introduce $\boldsymbol{E}'^{\det} = \boldsymbol{R}_{-A}\boldsymbol{E}^{\det}$ and $\boldsymbol{E}'^{\mathrm{i}} = \boldsymbol{R}_{-P}\boldsymbol{E}^{\mathrm{i}}$ for convenience. A rotates with a constant angular frequency Ω. The detected intensity $I = |\boldsymbol{E}'^{\det}|$ has the two real-valued harmonic coefficients α and β

$$I(\Omega t) = I_0\{1 + \alpha\cos 2\Omega t + \beta\sin 2\Omega t\}, \tag{9.17}$$

with

$$\alpha = \frac{|R_1|^2 - |R_2|^2}{|R_1|^2 + |R_2|^2}, \qquad \beta = 2\frac{Re\{R_1\overline{R}_2\}}{|R_1|^2 + |R_2|^2}, \tag{9.18}$$

where the bar denotes the complex conjugate, and $Re\{\}$ the real part. The complex quantities R_1 and R_2 depend on the GE parameters R_{pp}, R_{ps}, and R_{sp}, the polarizer and compensator azimuths P and C, and the phase retardation w

$$\begin{aligned} R_1 = {} & (R_{pp}[\exp(iw)\cos^2 C + \sin^2 C] + R_{pp}R_{ps}\sin C\cos C[\exp(iw) - 1])\cos P \\ & + (R_{sp}[\exp(iw)\cos^2 C + \sin^2 C] + \sin C\cos C[\exp(iw) - 1])\sin P, \end{aligned} \tag{9.19}$$

$$\begin{aligned} R_2 = {} & (R_{pp}R_{ps}[\cos^2 C + \exp(iw)\sin^2 C] + R_{pp}\sin C\cos C[\exp(iw) - 1])\cos P \\ & + ([\cos^2 C + \exp(iw)\sin^2 C] + R_{sp}\sin C\cos C[\exp(iw) - 1])\sin P. \end{aligned} \tag{9.20}$$

Any set $\{\alpha, \beta\}$ can be transformed into a corresponding set $\{\Psi, \Delta\}$

$$\tan\Psi = \sqrt{\frac{1+\alpha}{1-\alpha}}\tan P, \quad \cos\Delta = \frac{\beta}{\sqrt{1-\alpha^2}}, \tag{9.21}$$

with Ψ and $|\Delta|$ being the same parameters as defined in Eq. (9.9). Note that the RAE equations do not instantly provide the handedness of the polarization state of the reflected light beam because Δ appears only as the argument of the even cosine function. However, the compensator, if incorporated into the light path ($w \neq 0$), can overcome this shortcoming

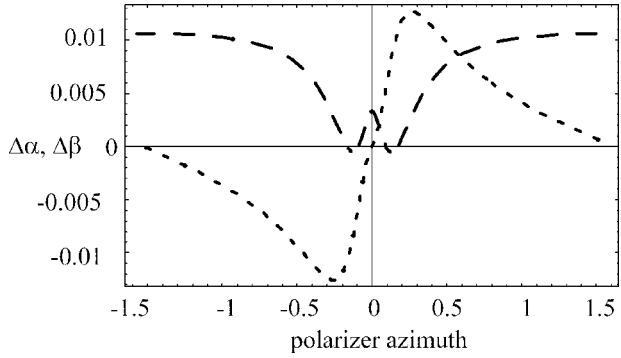


Figure 9.3 Differences between calculated RAE coefficients α, $\beta(P)$ ($-\pi/2 < P < \pi/2$), with and without R_{ps} and R_{sp} as an example ($w = 0$; $\delta\alpha \equiv \alpha[R_{sp}, R_{sp} \neq 0] - \alpha[R_{sp}, R_{sp} = 0]$: short-dashed curve; $\delta\beta$: long-dash curve); $R_{pp} = \tan 19.14° \exp(i50.13°)$, $R_{ps} = \tan 1.33 \exp(i125.4°)$, $R_{sp} = \tan 0.82° \exp(i153.4°)$, after Ref. 8).

by mapping different sets of $\{\alpha, \beta\}$ at different azimuths C and/or retardations w.[24] The extension of RAE to GE consists of the measurement of $\{\alpha, \beta\}$ at multiple sets of $\{P, C, w\}$, and of the determination of the three unknown complex parameters R_{pp}, R_{ps}, and R_{sp} with a regression analysis performed by use of Eqs. (9.18), (9.19), and (9.20). If no compensator is present, RAE will be unable to detect the sign of all phases of the GE parameters. This can be seen by evaluating α and β with one set of R_{pp}, R_{sp}, and R_{ps} and once again using the complex conjugates of the GE parameters. In both cases α and β will be the same.

It was pointed out by Azzam and Bashara that in principle only three different sets of $\{\alpha, \beta\}$ are necessary to determine R_{pp}, R_{sp}, and R_{ps}.[1, 2] However, especially if the off-diagonal coefficients have small amplitudes the number of experimental sets $\{\alpha, \beta\}$ should be large. To make such a critical situation more transparent Fig. 9.3 shows a simulation of the differences in α, β with and without small off-diagonal ratios as a function of the polarizer azimuth. For simplicity the compensator function is turned off for this calculation ($w = 0$). The harmonics α, β possess low and high sensitivity to R_{sp}, and R_{ps} at different polarizer azimuths. Hence, multiple polarizer settings within $-\pi/2 < P < \pi/2$ are necessary to obtain best information for the off-diagonals. The same is true for the compensator azimuth and retardation values. Additional mapping of the compensator azimuth and retardation over large parameter ranges can significantly improve the detection limit for very small anisotropies.

9.3 Theory of Generalized Ellipsometry

9.3.1 Birefringence in Stratified Media

The frequency-domain electromagnetic constitutive relations in Cartesian coordinates can be written as

$$\mathrm{D}(r,\omega) = \epsilon_0 [\epsilon(r,\omega)\mathrm{E}(r,\omega) + \delta(r,\omega)\mathrm{H}(r,\omega)], \tag{9.22}$$

$$\mathrm{B}(r,\omega) = \mu_0 [\mu(r,\omega)\mathrm{H}(r,\omega) + \gamma(r,\omega)\mathrm{E}(r,\omega)], \tag{9.23}$$

where ϵ_0 and μ_0 are the vacuum permittivity and the vacuum permeability, and ϵ and μ are the materials relative permittivity and relative permeability tensors, respectively.[†] The cross-coupling tensors δ and γ are the two magneto electric tensor dyadics. (For definition of the constitutive relations for several anisotropic materials see textbooks by Kong, Lakhtakia, or Landau and Lifshitz). The materials optical response is said to be anisotropic in general if at least one of ϵ or μ differ in more than one of its major elements, whereas δ and γ can be zero. The general case where all four tensors have complex, nonsymmetric, and nonvanishing elements is referred to as a bianisotropic medium (see, e.g., Refs. 25–28). For homogeneous media the four tensors do not depend on the spatial position r. However, an important exception is that of the so-called helicoidal bianisotropic medium (HBM), where the four tensors possess rotational nonhomogeneity along the z-axis. For details of light propagation in HBMs, see Refs. 29–32. The constitutive relations for a general HBM cover chiral liquid crystals (CLCs), ferrocholesteric, ferrosmectic, and supercholesteric media. Discussion of CLC cells will be given further below as these materials have been investigated through the GE formalism. The other classes still await experimental exploitation through the GE approach, and will not be addressed here.

This chapter continues with materials where the magneto electric tensors vanish, and with the relative permeability tensor set to unity. The remaining tensor is the dielectric function tensor ϵ, which can describe isotropic, uniaxial, or biaxial dielectric material properties (Table 9-1). In fact, the optical response of most solids and liquids can be well understood in terms of the major elements of ϵ and rotations, which relate major tensor axes to the laboratory frame of reference. Care should be taken if the "centers-of-gravity" of the real and imaginary parts of ϵ do

[†]Note that in this chapter, the symbol "ϵ" and the symbol "ε" both refer to the same physical quantity, namely the dielectric function.

Table 9-1. Dielectric tensor properties and crystal systems of symmetrically dielectric media.

Optical symmetry	Crystal system	Dielectric tensor
isotropic	Cubic	$\epsilon_x = \epsilon_y = \epsilon_z$
uniaxial	Tetragonal Hexagonal Trigonal	$\epsilon_x = \epsilon_y \neq \epsilon_z$
biaxial	Triclinic Monoclinic Orthorhombic	$\epsilon_x \neq \epsilon_y \neq \epsilon_z$

not coincide. Such angle-dispersive media require wavelength dependent rotations.[†] The above noted CLCs have symmetric dielectric properties with rotational nonhomogeneity. Crystal classes with nonreciprocity effects can be categorized into media with one (uniaxial) or two (biaxial) optical axes. Magneto-optical effects in solids cause antisymmetric dielectric properties, and ϵ becomes nonsymmetric in general. A nonsymmetric tensor decomposes into its symmetric and its antisymmetric part, and both can have three major complex elements. Such materials are said to be bi-biaxial.

To illustrate the physical origin of nonsymmetry in ϵ due to an external magnetic field **H**, the classical picture of the free-carrier magneto-optical birefringence is recalled here. The application of the GE formalism to a highly doped semiconductor film placed within a constant magnetic field is also discussed in Ref. 34. The dielectric function of a noncubic semiconductor with free carriers may be written as

$$\epsilon = \epsilon_L - i\frac{4\pi}{\omega}\sigma. \tag{9.24}$$

The conductivity tensor σ causes ϵ to be nonsymmetric (Pidgeon[33]). The symmetric tensor ϵ_L comprises the anisotropic lattice response and high-frequency behavior. The components of σ follow from the equation of motion for a free carrier with charge q, anisotropic effective mass tensor μ ($\mu_{ij} = \mu_{ji}$), and spatially dispersive but energy-independent inverse relaxation time (scattering rate) tensor γ ($\gamma_{ij} = \gamma_{ji}$) when moving in a constant magnetic field $\mathbf{H} = (H_1, H_2, H_3)$ with velocity $\mathbf{v} = (\partial_t x, \partial_t y, \partial_t z)$. $\mathbf{E} = (E_1, E_2, E_3) exp\{-i\omega t\}$ is the driving field (The tensors μ and γ here are not to be confused with the

[†]For derivation of monoclinic and triclinic dielectric tensors, refer, e.g., to Chapter 2 in M. Schubert, *Infrared Ellipsometry on Semiconductor layer structures: Phonons, Plasmons, and Polaritons* (Springer, Heidelberg, 2005).

permeability and magnetodielectric tensors above.)

$$\mu\frac{1}{\mathrm{q}}(\gamma+\partial_t)\mathbf{v}=\left(\mathbf{E}+\frac{1}{\mathrm{c}}\left[\mathbf{v}\times\mathbf{H}\right]\right). \tag{9.25}$$

A harmonic plane wave solution for **v** is assumed. The axial cross product of the Lorentz force leads to nonsymmetry in σ:

$$\sigma=\frac{1}{4\pi}\boldsymbol{H}^{-1}\langle\omega_p^2\rangle, \tag{9.26}$$

with (**I** is the unit matrix)

$$\boldsymbol{H}=\gamma-\mathrm{i}\omega\mathbf{I}-\frac{\mathrm{q}}{\mathrm{c}}\mu^{-1}\begin{pmatrix}0 & -\mathrm{H}_3 & \mathrm{H}_2\\ \mathrm{H}_3 & 0 & -\mathrm{H}_1\\ -\mathrm{H}_2 & \mathrm{H}_1 & 0\end{pmatrix}, \tag{9.27}$$

and the plasma frequency tensor $\langle\omega_p^2\rangle$ for N carriers per unit volume† is defined as

$$\langle\omega_p^2\rangle=4\pi\mathrm{q}^2\mathrm{N}\mu^{-1}. \tag{9.28}$$

This chapter is about anisotropic materials in layered samples. The problem of light propagation through plane parallel interfaces will be solved, and measurable quantities for the GE formalism will be assigned. The dielectric function tensor follows from numerical inversion of the GE equations. Strategies for data acquisition and analysis will be given and discussed in detail for several applications.

9.3.2 4 × 4 Maxwell's Equations in Matrix Form

Berreman developed a 4 × 4 matrix algorithm which in principle allowed calculation of the Jones matrices of anisotropic layered structures with plane parallel boundaries.[35] But numerical approximations with extensive computational effort were necessary to solve the Booker quartic and the corresponding matrix differential equation.[35, 36] Wöhler *et al.* showed later on a faster way to solve the matrix differential equation within the algorithm using the Sylvester-Lagrange theorem.[6, 37] Solutions of the Booker quartic for several types of linear dielectric birefringent

†The first direct measurement of the free-carrier magneto-optic tensor in a semiconductor layer structure using GE at far-infrared wavelengths was reported recently in "Generalized far-infrared magneto-optic ellipsometry for semiconductor layer structures: Determination of free-carrier effective mass, mobility and concentration parameters in n-type GaAs," M. Schubert, T. Hofmann, and C. M. Herzinger, *J. Opt. Soc. Am. A* **20**, 347–356 (2003).

materials were presented recently.[34, 38] These extensions made the approach analytical throughout, which in turn became fast enough for regression routines to solve for birefringence in large GE data sets. The Berreman concept has been thoroughly discussed in the literature.[7, 35, 37–40] A slightly different concept for a 4×4 matrix approach was developed by Yeh.[41] A comparison between the algorithm developed, respectively, by Berreman and Yeh is given by Lin-Chung and Teitler.[39] The 4×4 matrices follow naturally from reduction of Maxwell equations and the constraint of plane parallel boundaries. Ingenious arithmetic further allows reduction of the 4×4 matrix expression to a series of 2×2 matrices. Such approaches have been presented by Mansuripur[42] and Toussaere and Zyss[44], and will not be discussed here.

The key of the 4×4 matrix algebra consolidates the so-called general transfer matrix $\mathbf{T}$. This matrix provides the Jones reflection and transmission matrix elements of arbitrary anisotropic but homogeneous samples with plane parallel boundaries. One advantage of the 4×4 approach is that a so-called partial transfer matrix $\mathbf{T}_p$ is assigned to each layer. This matrix depends on the optical and structural properties of the respective layer. The challenge consists in finding this matrix for any type of intrinsic or form-induced birefringent layer. The matrices $\mathbf{T}_p$ can be used like building blocks within a unit construction system to model the optical response of layers stacked in any order. Another more technical advantage is that during the regression analysis only the matrix for the layer under consideration is rebuilt for each numerical parameter guess. In this section the 4×4 matrix algebra will be outlined with adequate details. Available solutions for special anisotropic layers will be presented in Section 9.4. The algebra developed here will then be used during Section 9.6 for analysis of GE data.

Consider a layered system with plane parallel interfaces. Assume an incident light wave with wave vector $\boldsymbol{k}_a$ coming from the incident medium (ambient, index a, $-\infty < z < 0$, index of refraction $n_a = \sqrt{\in_a}$) at an angle of incidence Φ_a (Fig. 9.4). All resulting wave vectors lie within the plane of incidence (see also Fig. 9.2), and have no y-components.† The general transfer matrix $\mathbf{T}$ connects wave amplitudes across a "coherent" layer sequence. The coherency of a layer stack is given by its total optical thickness, which shall not exceed the coherence length of the light beam. For the general case depicted in Fig. 9.4 we need two matrices $\mathbf{T}_+$ and $\mathbf{T}_-$

$$(A_s, \quad B_s, \quad A_p, \quad B_p)^T = \mathbf{T}_+ (C_s, \quad D_s, \quad C_p, \quad D_p)^T, \tag{9.29}$$

$$(E_s, \quad F_s, \quad E_p, \quad F_p)^T = \mathbf{T}_- (C'_s, \quad D'_s, \quad C'_p, \quad D'_p)^T, \tag{9.30}$$

†For samples with corrugated surfaces the resulting wavevectors are not bound to the plane of incidence.

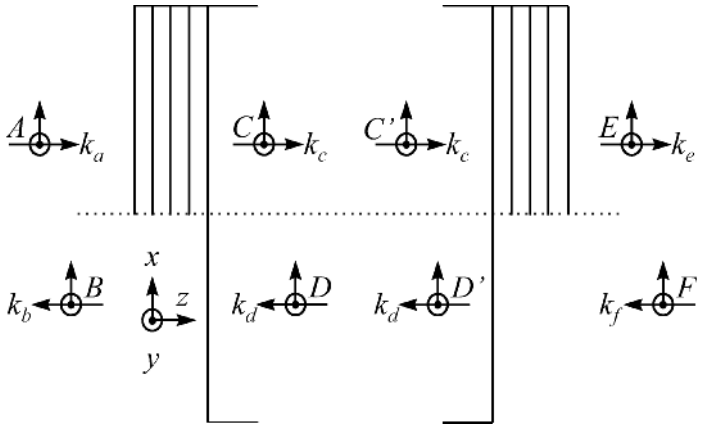


Figure 9.4 Schematic representation of incident, reflected, and transmitted *p* and *s* modes of plane waves *A*, *B* (incident side), *C*, *D* (substrate), and *E*, *F* (exit side). *D* modes are either reflected off the backside of the substrate, or caused by modes *F* incident from the right. If the substrate totally absorbs the *C* modes, *D*, *E*, and *F* are just meaningless. The backside of the sample may support another layer stack. The wave vectors for the *C*, *C*' and *D*, *D*' modes may split if the substrate is anisotropic. *C* and *D* represent modes just inside the left border of the substrate, while *C*' and *D*' represent the modes just inside its right border.

where $(\ )^T$ denotes the transpose of a vector. Across the substrate, coherent modes on the left and right side in Fig. 9.4 can be related by the diagonal matrix $\mathbf{M}$ which depends on the thickness of the substrate d and the eigenvalues q_i of the characteristic matrix Δ built for the substrate material. (See Section 9.4.)

$$(C_s,\ D_s,\ C_p,\ D_p)^T = \mathbf{M}[d, q_i](C'_s,\ D'_s,\ C'_p,\ D'_p)^T, \tag{9.31}$$

$$\mathbf{M}[d, q_i] = \mathrm{diag}[ik_0 d \exp(q_1^+,\ q_1^-,\ q_2^+,\ q_2^-)]. \tag{9.32}$$

For now we will assume the substrate to be optically infinite. We shall later include the case of transmission through the substrate, the case of light reflected off the backside of the sample, and the case where the sample is measured from the backside. We will also be able to treat layers stacked at the backside of the sample. For this we need to connect $(C_s, D_s, C_p, D_p)^T$ and $(C'_s, D'_s, C'_p, D'_p)^T$ within the anisotropic substrate using the eigenvalues and eigenmodes of plane waves within the substrate, which depend on $\boldsymbol{k}_a$ (or $\boldsymbol{k}_f$).[†]

Because we do not include a physical backside yet, only C_p and C_s are considered in Eq. (9.29) ($D_p = D_s = 0$, $E_p = E_s = 0$, $F_p = F_s = 0$). The

[†]See also pages 684–688 in *Introduction to Complex Mediums for Optics and Electromagnetics*, edited by W. S. Weiglhofer and A. Lakhtakia (SPIE, Bellingham, 2003).

tangential components of the electric and magnetic field vectors are continuous across the interface between two media (if the surface current density and the surface charge density vanish.) If d_i denotes the thickness of the i-th layer, a partial transfer matrix $\mathbf{T}_{ip}$ that connects the in-plane wave components at the interface at $z = z_i$ with those at the next interface at $z = z_i + d_i$ can be defined for any homogeneous layer. The ordered product of all such matrices from all N layers transfers the in-plane components at the first interface at $z = 0$ to the last interface at $z = z_N$. We further define the incident matrix $\mathbf{L}_a$ which projects the in-plane wave components of the incident and reflected waves on the left side in Fig. 9.4 to the sample surface. Similarly, the exit matrix $\mathbf{L}_f$ projects the amplitudes at the last interface into the exit medium, which is now the substrate. The general transfer matrix $\mathbf{T}$ defined in Eq. (9.29) is then obtained from the product of all inverted matrices $\mathbf{T}_{ip}$, and the incident and exit matrices in the order of their appearance

$$\mathbf{T} = \mathbf{L_a}^{-1}\left[\prod_{\mathrm{i}=1}^{\mathrm{N}}\mathbf{T_{ip}}(d_i)\right]^{-1}\mathbf{L_f} = \mathbf{L_a}^{-1}\prod_{\mathrm{i}=\mathrm{N}}^{1}\mathbf{T_{ip}}(-d_i)\mathbf{L_f}. \tag{9.33}$$

The inversion of $\mathbf{T}_p$ as indicated in Eq. (9.33) does not require a matrix inversion calculation. The matrices $\mathbf{L}_a$ and $\mathbf{L}_f$ are discussed in Ref. 38. The incident matrix

$$\mathbf{L_a}^{-1} = \frac{1}{2}\begin{bmatrix} 0 & 1 & -1/n_a\cos\Phi_a & 0 \\ 0 & 1 & 1/n_a\cos\Phi_a & 0 \\ 1/\cos\Phi_a & 0 & 0 & 1/n_a \\ -1/\cos\Phi_a & 0 & 0 & 1/n_a \end{bmatrix}, \tag{9.34}$$

depends on the angle of incidence Φ_a and the index of refraction n_a of the ambient material. If the substrate is birefringent the C modes can couple to each other. The exit matrix follows from the assumption that the total transmitted field within the substrate consists of a linear combination of the eigenvectors Ξ_i of the characteristic coefficient matrix Δ. (See Section 9.4.) The coefficients are the amplitudes C_p and C_s. The matrix $\mathbf{L}_f$ is then obtained as follows

$$L_{fjk} = \Xi_{jk}, \quad j, k = 1 \ldots 4, \tag{9.35}$$

and explicit expressions of Ξ_{jk} for nonsymmetric dielectric media are given in Ref. 43. If the major axes of a symmetrically dielectric substrate coincide with the laboratory axes (x, y, z), the C modes decouple, then $\mathbf{L}_f$ reads ($n_x = \sqrt{\epsilon_x}$, $n_y = \sqrt{\epsilon_y}$, $n_z = \sqrt{\epsilon_z}$)

$$\mathbf{L_f} = \begin{bmatrix} 0 & 0 & \cos\Phi_z & -\cos\Phi_z \\ 1 & 1 & 0 & 0 \\ -n_y\cos\Phi_y & n_y\cos\Phi_y & 0 & 0 \\ 0 & 0 & n_x & n_x \end{bmatrix}. \tag{9.36}$$

The angle Φ_j is calculated from Snell's law

$$\cos\Phi_j = \sqrt{1-[(na/nj)\sin\Phi_s]^2}. \tag{9.37}$$

9.3.3 Transmission and Reflection GE

The transmission and reflection coefficients can be expressed in terms of the elements of the general transfer matrix **T**. Consider Eq. (9.29) as a system of four linear relations between the p and s modes across a coherent layer stack (Fig. 9.4). The elements r_{ss} and t_{sp} read for example (See Ref. 41 for the remaining expressions. See also Section 9.2.1.)

$$r_{ss} \equiv \left(\frac{B_s}{A_s}\right)_{A_p=0} = \frac{T_{21}T_{33} - T_{23}T_{31}}{T_{11}T_{33} - T_{13}T_{31}},\ t_{sp} \equiv \left(\frac{C_p}{A_s}\right)_{A_p=0} = \frac{-T_{31}}{T_{11}T_{33} - T_{13}T_{31}}. \tag{9.38}$$

The complex reflectance ratio ρ defined in Eq. (9.9) also follows from **T**

$$\rho = \frac{1}{\chi}\frac{\mathrm{T}_{41}(\mathrm{T}_{33} - \chi\mathrm{T}_{13}) + \mathrm{T}_{43}(\chi\mathrm{T}_{11} - \mathrm{T}_{31})}{\mathrm{T}_{21}(\mathrm{T}_{33} - \chi\mathrm{T}_{13}) + \mathrm{T}_{23}(\chi\mathrm{T}_{11} - \mathrm{T}_{31})}, \quad \chi \equiv \frac{A_p}{A_s}. \tag{9.39}$$

If we substitute the expressions for the transmission and reflection coefficients derived from **T** we obtain for ρ

$$\rho = [r_{pp}/r_{ss} + (r_{sp}/r_{ss})\chi^{-1}]/[1 + (r_{ps}/r_{pp})(r_{pp}/r_{ss})\chi], \tag{9.40}$$

and similarly for the complex transmittance ratio τ

$$\tau = [t_{pp}/t_{ss} + (t_{sp}/t_{ss})\chi^{-1}]/[1 + (t_{ps}/t_{pp})(t_{pp}/t_{ss})\chi], \tag{9.41}$$

which resemble Eqs. (9.14). The reflection GE parameters read in terms of **T** elements

$$R_{pp} = \frac{T_{11}T_{43} - T_{41}T_{13}}{T_{21}T_{33} - T_{23}T_{31}},\ R_{ps} = \frac{T_{11}T_{12} - T_{21}T_{13}}{T_{11}T_{43} - T_{41}T_{13}},\ R_{sp} = \frac{T_{41}T_{33} - T_{43}T_{31}}{T_{21}T_{33} - T_{23}T_{31}}, \tag{9.42}$$

and the transmission GE parameters are

$$T_{pp} = \frac{T_{11}}{T_{33}}, \qquad T_{ps} = -\frac{T_{13}}{T_{11}}, \qquad T_{sp} = -\frac{T_{31}}{T_{33}}. \tag{9.43}$$

ρ and τ simplify in cases where only isotropic materials are included in the layered system or in some special cases where, e.g., the Cartesian principal axes of all layers are collinear with the axes of the laboratory coordinate system[38]

$$\rho^{is} = \frac{T_{43}T_{11}}{T_{33}T_{21}},\ r_{ss} = \frac{T_{21}}{T_{11}},\ r_{pp} = \frac{T_{43}}{T_{33}},\quad \text{and}\quad r_{ps} = r_{sp} = 0,$$
$$\tau^{is} = \frac{T_{11}}{T_{33}},\ t_{ss} = \frac{1}{T_{11}},\ t_{pp} = \frac{1}{T_{33}},\quad \text{and}\quad t_{ps} = t_{sp} = 0. \tag{9.44}$$

9.4 Special Generalized Ellipsometry Solutions

9.4.1 Biaxial Films (Symmetrically Dielectric Materials)

From first-order Maxwell equations Berreman derived the following set of four differential equations for the in-plane components of the electric and magnetic fields in Gaussian units and Cartesian coordinates:

$$\partial_z \Psi(z) = ik_0 \Delta(z)\Psi(z),\quad \Psi(z) = (E_x,\ E_y,\ H_x,\ H_y)^T(z),\quad k_0 \equiv \frac{\omega}{c}, \tag{9.45}$$

with

$$k_0 \equiv \frac{\omega}{c} = 2\pi \frac{e}{hc} E[\text{eV}] = 5.067690488 \times 10^6 E[\text{eV}], \tag{9.46}$$

where ω is the angular frequency, and c is the vacuum velocity of light.[35] Note the sign change of H_x here compared to Eq. (23) in Ref. 35. In this chapter we will assume that the linear response dyadic is zero except for the dielectric tensor $\in$, and the permeability has its scalar vacuum value μ_0. The major dielectric functions $\in_x$, $\in_y$, and $\in_z$ belong to intrinsic axes of the anisotropic material. We further assume without loss of generality that these axes are Cartesian (i.e., orthorhombic). For monoclinic or triclinic systems one may always find a projection of major values onto a

Cartesian system. However, this transformation will then depend on ω. The major crystal axes may generally be different from the laboratory coordinate axes. The Euler angles φ, ψ, and θ can be used to rotate Cartesian coordinate reference frames

$$\in = \mathbf{A}\begin{pmatrix} \varepsilon_x & 0 & 0 \\ 0 & \varepsilon_y & 0 \\ 0 & 0 & \varepsilon_z \end{pmatrix}\mathbf{A}^{-1}, \tag{9.47}$$

and the unitary matrix **A** is the orthogonal rotation matrix

$$\mathbf{A} = \begin{pmatrix} \cos\psi\cos\varphi - \cos\theta\sin\varphi\sin\psi & -\sin\psi\cos\varphi - \cos\theta\sin\varphi\cos\psi & \sin\theta\sin\varphi \\ \cos\psi\sin\varphi + \cos\theta\cos\varphi\sin\psi & -\sin\psi\sin\varphi + \cos\theta\cos\varphi\cos\psi & -\sin\theta\cos\varphi \\ \sin\theta\sin\psi & \sin\theta\cos\psi & \cos\theta \end{pmatrix}. \tag{9.48}$$

(See H. Goldstein, *Classical Mechanics*; see also Fig. 9.5.) Note that $\mathbf{A}^{-1}(\varphi, \psi, \theta)$ is obtained as the transpose of A. {Note that references[38] and[34] contain a typographical error, where it should read $\mathbf{A}^{-1}(\varphi, \psi, \theta) \neq \mathbf{A}(-\varphi, -\psi, -\theta)$!} The Euler angles describe the rotation from the laboratory system into the major crystal system. Note further that for now $\in$ is symmetric, i.e., $\in_{ij} = \in_{ji}$. The coefficient matrix Δ defined in Eq. (9.45) depends on the dielectric tensor $\in$ and the x component k_x of the wave vector $\boldsymbol{k}_a$.

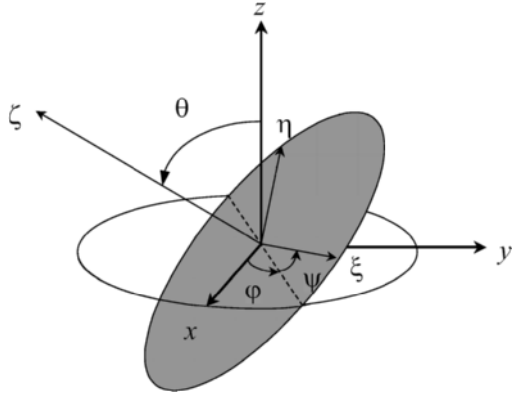


Figure 9.5 Definition of Euler angles φ, ψ, and θ, and the orthogonal rotations as provided by **A**. (ξ, η, ζ, and (x, y, z) refer to the crystal and the laboratory coordinate systems, respectively.

$$(1 = \text{“}x\text{”}, 2 = \text{“}y\text{”}, 3 = \text{“}z\text{”})$$

$$\Delta = \begin{pmatrix} -k_x\dfrac{\varepsilon_{31}}{\varepsilon_{33}} & -k_x\dfrac{\varepsilon_{32}}{\varepsilon_{33}} & 0 & 1-\dfrac{k_x^2}{\varepsilon_{33}} \\ 0 & 0 & -1 & 0 \\ \varepsilon_{23}\dfrac{\varepsilon_{31}}{\varepsilon_{33}}-\varepsilon_{21} & k_x^2-\varepsilon_{22}+\varepsilon_{23}\dfrac{\varepsilon_{32}}{\varepsilon_{33}} & 0 & k_x\dfrac{\varepsilon_{23}}{\varepsilon_{33}} \\ \varepsilon_{11}-\varepsilon_{13}\dfrac{\varepsilon_{31}}{\varepsilon_{33}} & \varepsilon_{12}-\varepsilon_{13}\dfrac{\varepsilon_{32}}{\varepsilon_{33}} & 0 & -k_x\dfrac{\varepsilon_{13}}{\varepsilon_{33}} \end{pmatrix},$$

$$k_x \equiv n_a \sin\Phi_a \tag{9.49}$$

The matrix Δ does not depend on z if the medium is homogeneous. The solution of Eq. (9.45) for a layer with thickness d is the partial transfer matrix $\mathbf{T}_p$

$$\Psi(z+d) = \exp\{ik_0\Delta d\}\Psi(z) = \mathbf{T_p}\Psi(z), \quad \mathbf{T_p} \equiv \exp\{ik_0\Delta d\}. \tag{9.50}$$

This matrix connects the in-plane components of the electric and magnetic fields at layer interfaces separated by d. $\mathbf{T}_p$ includes the effects of all multiple reflections if a part of the wave is traveling along a direction with no or weak absorption. It may also be noted that $\mathbf{T}_p$ is unitary if the medium is loss less in any direction of propagation as a consequence of the conservation law of energy.[7] Otherwise $|\Psi(z+d)|$ and $|\Psi(z)|$ may not be equal. Besides the thickness d, for an orthorhomic medium, $\mathbf{T}_p$ has nine unknowns at each wavelength, including three complex major dielectric function values and three real-valued Euler angles. $\mathbf{T}_p$ is computed with Δ as input. Requiring sufficiently small thickness d for fast convergence the exponential function in Eq. (9.50) can be expanded into a series with the wavelength-dependent factor $dk_0 = 2\pi d/\lambda$. Wöhler *et al.* showed a faster way to calculate the partial transfer matrix applying the theorem of Sylvester-Lagrange[37]

$$\mathbf{T_p} \equiv \exp\{ik_0\Delta d\} = \beta_0\mathbf{E} + \beta_1\Delta + \beta_2\Delta^2 + \beta_3\Delta^3. \tag{9.51}$$

For the latter equation small thickness d is not required. Note the identity $[\mathbf{T}_p(d)]^{-1} = \mathbf{T}_p(-d)$. The scalars β_i following from the linear relations

$$\exp\{ik_0q_kd\} = \sum_{j=0}^{3}\beta_j q_k^j,\ k = 1, \ldots, 4, \tag{9.52}$$

and the q_i are the eigenvalues of Δ associated with one of the four eigenmodes Ξ_i existing within the layer. Two solutions have a positive real part

and constitute the forward traveling plane waves with respect to the chosen laboratory coordinate system. The other solutions with negative real parts are due to the back-traveling wave components. The eigenvalues are

$$q_{1/2}^{+} = \frac{1}{2}\left\{-k_x\frac{\varepsilon_{13}}{\varepsilon_{33}} - \sqrt{-\frac{2}{3}t_1 + \left(k_x\frac{\varepsilon_{13}}{\varepsilon_{33}}\right)^2 + \Sigma} \pm \sqrt{-\frac{4}{3}t_1 + 2\left(k_x\frac{\varepsilon_{13}}{\varepsilon_{33}}\right)^2 - \Sigma - \frac{s_3}{4\sqrt{-\frac{2}{3}t_1 + \left(k_x\frac{\varepsilon_{13}}{\varepsilon_{33}}\right)^2 + \Sigma}}}\right\}$$

$$q_{1/2}^{-} = \frac{1}{2}\left\{-k_x\frac{\varepsilon_{13}}{\varepsilon_{33}} + \sqrt{-\frac{2}{3}t_1 + (k_x\frac{\varepsilon_{13}}{\varepsilon_{33}})^2 + \Sigma} \pm \sqrt{-\frac{4}{3}t_1 + 2\left(k_x\frac{\varepsilon_{13}}{\varepsilon_{33}}\right)^2 - \Sigma + \frac{s_3}{4\sqrt{-\frac{2}{3}t_1 + \left(k_x\frac{\varepsilon_{13}}{\varepsilon_{33}}\right)^2 + \Sigma}}}\right\} \quad (9.53)$$

with the abbreviations

$$\Sigma = \frac{1}{2}\left\{s_1\left(\frac{1}{2}\left\{s_2 + \sqrt{s_2^2 - 4s_1^3}\right\}\right)^{-\frac{1}{3}} + \left(\frac{1}{2}\left\{s_2 + \sqrt{s_2^2 - 4s_1^3}\right\}\right)^{\frac{1}{3}}\right\},$$

$$s_1 = {t_1}^2 + 12\left(k_x\frac{\in_{13}}{\in_{33}}t_2 + t_3\right),$$

$$s_2 = 2{t_1}^3 + 36k_x\frac{\in_{13}}{\in_{33}}t_1t_2 + 108\left({t_2}^2 + \left[k_x\frac{\in_{13}}{\in_{33}}\right]^2 t_3\right) - 72t_1t_3,$$

$$s_3 = -8k_x\frac{\epsilon_{13}}{\epsilon_{33}}\left(\left[k_x\frac{\epsilon_{13}}{\epsilon_{33}}\right]^2 - t_1\right) + 16t_2,$$

and

$$t_1 = \frac{1}{\epsilon_{33}}\left\{\epsilon_{13}^2 + \epsilon_{23}^2 - \epsilon_{33}\left[\epsilon_{11} + \epsilon_{22} - k_x{}^2\left(1 + \frac{\epsilon_{11}}{\epsilon_{33}}\right)\right]\right\},$$

$$t_2 = \frac{k_x}{\epsilon_{33}}(\epsilon_{13}\epsilon_{22} - \epsilon_{12}\epsilon_{23} - \epsilon_{13}k_x{}^2),$$

$$t_3 = \frac{1}{\epsilon_{33}}\{-\epsilon_{11}\epsilon_{23}^2 - \epsilon_{22}\epsilon_{13}^2 - \epsilon_{33}\epsilon_{12}^2 + 2[\epsilon_{12}\epsilon_{13}\epsilon_{23}] + k_x{}^2[\epsilon_{12}^2 + \epsilon_{13}^2 - \epsilon_{11}(\epsilon_{22} + \epsilon_{33} - k_x^2)] + \epsilon_{11}\epsilon_{22}\epsilon_{33}\}. \quad (9.54)$$

Here $q_{1/2}^+$ ($q_{1/2}^-$) refer to forward- (backward-) traveling waves, respectively. Equation (9.50) can now be evaluated using the coefficients β_i which follow from Eqs. (9.52). The eigenvalues reported so far for special orientations of uniaxial media are retained here as particular solutions. (For the eigenvalues from a uniaxial medium see, e.g., Ref. 7; for $\mathbf{T}_p$ from a uniaxial medium with special optical axis orientations see Wöhler *et al*., Ref. 6.)

9.4.2 Bi-Biaxial or Magneto-Optical Films (Non-Symmetrically Dielectric Materials)

We now consider media with non-symmetric dielectric properties, i.e., materials where $\epsilon_{ij} \neq \epsilon_{ji}$. Examples for such materials are magneto-optic compounds of non-cubic symmetry, or the effects of magneto-polariton couplings in non-symmetric semiconductors subjected to an arbitrarily oriented external magnetic field (surface and bulk magneto-plasmons in non-cubic semiconductors).† Other examples are magneto-optic biaxial materials with arbitrary magnetization direction.[34] ϵ is now composed of a symmetric (**s**) and an anti-symmetric (**a**) tensor

$$\epsilon_{ij} = s_{ij} + a_{ij},\ s_{ij} = s_{ji},\ a_{ij} = -a_{ji},$$
$$\text{i.e., } s_{ij} = (1/2)(\epsilon_{ij} + \epsilon_{ji}),\ a_{ij} = (1/2)(\epsilon_{ij} - \epsilon_{ji}). \quad (9.55)$$

†An introduction to electromagnetic wave propagation in magneto-optically biaxial materials in application to surface and bulk related plasmons in semiconductors can be found in R.F. Wallis, "Optical properties associated with surface excitations of semiconductors", *Handbook on Semiconductors,* Vol. 2, edited by M. Balkanski (North-Holland, Amsterdam, 1994), pp. 65–108. See also: M. Schubert, *Infrared Ellipsometry on Semiconductor layer structures: Phonons, Plasmons, and Polaritons* (Springer, Heidelberg, 2005).

The matrix Δ reads then

$$\Delta = \begin{bmatrix} -k_x\dfrac{s_{13}-a_{13}}{s_{33}} & -k_x\dfrac{s_{23}-a_{23}}{s_{33}} & 0 & 1-\dfrac{k_x^2}{s_{33}} \\ 0 & 0 & -1 & 0 \\ \dfrac{(s_{23}+a_{23})(s_{13}-a_{13})}{s_{33}} - (s_{12}-a_{12}) & k_x^2 - s_{22} + \dfrac{(s_{23}-a_{23})(s_{23}+a_{23})}{s_{33}} & 0 & k_x\dfrac{(s_{23}+a_{23})}{s_{33}} \\ s_{11} - \dfrac{(s_{13}+a_{13})(s_{13}-a_{13})}{s_{33}} & (s_{12}+a_{12}) - \dfrac{(s_{23}-a_{23})(s_{13}+a_{13})}{s_{33}} & 0 & -k_x\dfrac{(s_{13}+a_{13})}{s_{33}} \end{bmatrix}, \tag{9.56}$$

and the eigenvalues of Δ are

$$q_{1/2}^{+} = \frac{1}{2}\left\{-k_x\frac{s_{13}}{s_{33}} - \sqrt{-\frac{2}{3}t_1 + \left(k_x\frac{s_{13}}{s_{33}}\right)^2 + \Sigma} \pm \sqrt{-\frac{4}{3}t_1 + 2\left(k_x\frac{s_{13}}{s_{33}}\right)^2 - \Sigma - \frac{S_3}{4\sqrt{-\frac{2}{3}t_1 + \left(k_x\frac{s_{13}}{s_{33}}\right)^2 + \Sigma}}}\right\} \tag{9.57}$$

$$q_{1/2}^{-} = \frac{1}{2}\left\{-k_x\frac{s_{13}}{s_{33}} + \sqrt{-\frac{2}{3}t_1 + \left(k_x\frac{s_{13}}{s_{33}}\right)^2 + \Sigma} \pm \sqrt{-\frac{4}{3}t_1 + 2\left(k_x\frac{s_{13}}{s_{33}}\right)^2 - \Sigma + \frac{S_3}{4\sqrt{-\frac{2}{3}t_1 + \left(k_x\frac{s_{13}}{s_{33}}\right)^2 + \Sigma}}}\right\}$$

with Σ defined as with the symmetrical case, shown immediately after Eq. (9.53), and the new abbreviations

$$s_1 = t_1^{\,2} + 12\left(k_x \frac{s_{13}}{s_{33}} t_2 + t_3\right),$$

$$s_2 = 2t_1^{\,3} + 36k_x \frac{s_{13}}{s_{33}} t_1 t_2 + 108\left(t_2^{\,2} + \left[k_x \frac{s_{13}}{s_{33}}\right]^2 t_3\right) - 72 t_1 t_3,$$

$$s_3 = -8k_x \frac{s_{13}}{s_{33}}\left(\left[k_x \frac{s_{13}}{s_{33}}\right]^2 - t_1\right) + 16 t_2,$$

and

$$t_1 = \frac{1}{s_{33}}\left\{s_{13}^2 + s_{23}^2 - s_{33}\left[s_{11} + s_{22} - k_x^2\left(1 + \frac{s_{11}}{s_{33}}\right)\right] - [a_{23}^2 + a_{13}^2]\right\},$$

$$t_2 = \frac{k_x}{s_{33}}(s_{13}s_{22} - s_{12}s_{23} - s_{13}k_x^2 - a_{23}a_{12}),$$

$$t_3 = \frac{1}{s_{33}}\left\{\begin{array}{l} +s_{11}[a_{23}^2 - s_{23}^2] + s_{22}[a_{13}^2 - s_{13}^2] + s_{33}[a_{12}^2 - s_{12}^2] + s_{11}s_{22}s_{33} \\ +2[s_{12}s_{13}s_{23} - s_{12}a_{13}a_{23} + s_{13}a_{12}a_{23} - s_{23}a_{13}a_{12}] \\ +k_x^2[s_{12}^2 + s_{13}^2 - s_{11}(s_{22} + s_{33} - k_x^2) - (a_{13}^2 + a_{12}^2)] \end{array}\right\}. \tag{9.58}$$

The eigenvalues $q_{1/2}^{+}$ ($q_{1/2}^{-}$) refer again to the forward (backward) traveling waves, respectively. $\mathbf{T}_p$ is then obtained in the same way as done for the biaxial films above. Solutions for magneto-optical ellipsometry in polar, transverse, and longitudinal configuration, as described by Visnovsky for bulk and films, are retained here as well.[45,46] Note that the finite thickness assumption in[46] is no longer a requirement here.

9.4.3 Chiral Biaxial Films (Axially Twisted Symmetrically Dielectric Materials)

Consider a continuously twisted biaxial material (chiral liquid crystals, CLCs). ∈ depends on the spatial position with respect to the z-axis. Such a twisted medium consists of a helical structure with periodicity along the z-direction (Fig. 9.6); the uniaxial molecules are parallel to the (x, y) plane. (The general solution presented here covers biaxial molecules with tilt toward the z-axis.) The complex optical transfer function of

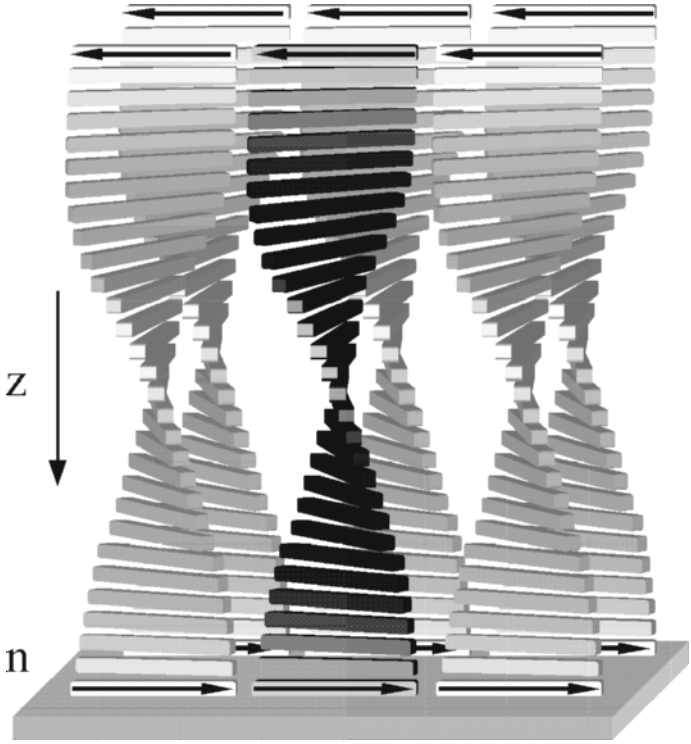


Figure 9.6 Schematic presentation of a right-handed CLC cell ($P > 0$). The pitch P is the distance between one full turn of molecule direction **n**. The sign of P refers to a right (positive) or left (negative) handed helical arrangement of the molecules.

twisted nematic liquid crystals has been the matter of numerous mathematical and experimental approaches.[31, 47–51] However, to calculate the optical response from a cholesteric liquid crystal, especially for light incident at an oblique angle, the layer is usually subdivided into many virtual slices and each slice is treated as a homogeneous biaxial material. This procedure requires immense computational effort and remains an approximation regardless of the number of virtual slices used. An exact solution of $\mathbf{T}_p$ for a CLC layer is shown here. This solution can be used for precise ellipsometric analysis of refractive indices and geometrical properties from liquid crystal devices such as flat panel displays.

The matrix Δ depends on z. Let **n** be a unit vector that is oriented parallel to the z-axis of the crystal coordinate system in each virtual plane formed by the twisted medium (Fig. 9.6). We assign P as the distance along the laboratory z-axis for one full turn of the vector **n** around

the laboratory z-axis. (The pitch P here is not to be confused with the polarizer azimuth P in Section 9.2.) The z-dependence of the dielectric function tensor describes one complete rotation per length P. Similar to Eq. (9.49) the orientation of the biaxial material at the lowest boundary of the slab is described by the unitary matrix $\mathbf{A}$ with φ, ψ, and θ as arguments. The helical rotation of the vector $\mathbf{n}$ along the z-axis is then provided by a rotation matrix $\mathbf{B}$ that depends only on the number of turns ζ per unit length P

$$\mathbf{B} = \begin{pmatrix} \cos\zeta & -\sin\zeta & 0 \\ \sin\zeta & \cos\zeta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \zeta = \frac{2\pi}{P} z. \tag{9.59}$$

The complex elements of the dielectric function tensor at each laboratory z-position are given by

$$\begin{aligned} \epsilon(\zeta) &\equiv \mathbf{B}(\zeta)\mathbf{A}\begin{pmatrix} \epsilon_x & 0 & 0 \\ 0 & \epsilon_y & 0 \\ 0 & 0 & \epsilon_z \end{pmatrix}\mathbf{A}^{-1}\mathbf{B}(\zeta)^{-1} \\ &= \mathbf{B}(\zeta)\begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{12} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{13} & \epsilon_{23} & \epsilon_{33} \end{pmatrix}\mathbf{B}(\zeta)^{-1}. \end{aligned} \tag{9.60}$$

The z-dependence of the matrix Δ needs to be eliminated in order to solve Eq. (9.45).[15] We transform both sides of Eq. (9.45) into the more appropriate elliptical coordinate system applying the unitary matrix Γ which is defined as follows

$$\tilde{\Psi}(\zeta) \equiv \Gamma\Psi(\zeta), \quad \Gamma = \begin{pmatrix} 1 & i & 0 & 0 \\ 1 & -i & 0 & 0 \\ 0 & 0 & 1 & i \\ 0 & 0 & 1 & -i \end{pmatrix}, \tag{9.61}$$

and results in

$$\tilde{\Psi}(\zeta) = (E^+, E^-, H^+, H^-)^T(\zeta), \quad H^{\pm} = H_x \pm iH_y, \quad E^{\pm} = E_x \pm iE_y. \tag{9.62}$$

The vector $\tilde{\Psi}$ contains the in-plane field amplitudes of back- and forward traveling elliptically polarized modes. Next we use the matrix $\mathbf{F}(\zeta)$ to transfer the helical dependence of the vector $\tilde{\Psi}$

$$\mathbf{F}(\zeta) \equiv \mathbf{diag}(\exp\{-i\zeta\}, \exp\{+i\zeta\}, \exp\{-i\zeta\}, \exp\{+i\zeta\}), \tag{9.63}$$

where **diag**( ) indicates the diagonal 4 × 4 matrix, introducing a new vector $\tilde{\Phi}$

$$\tilde{\Phi}(\zeta) = \mathbf{F}(\zeta)\tilde{\Psi}(\zeta). \tag{9.64}$$

If we substitute $\tilde{\Phi}$ (ζ) into Eq. (9.45) we obtain a new coefficient matrix $\tilde{\Delta}$

$$\partial_\zeta \tilde{\Phi}(\zeta) = in^{-1}\tilde{\Delta}(\zeta)\tilde{\Phi}(\zeta), \tag{9.65}$$

$$\tilde{\Delta}(\zeta) = \mathbf{F}(\zeta)\Gamma\Delta(\zeta)\Gamma^{-1}\mathbf{F}(\zeta)^{-1} + \partial_\zeta \mathbf{F}(\zeta), \tag{9.66}$$

$$\tilde{\Delta}(\zeta) = \mathbf{U} + e^{i\zeta}\mathbf{V}_+ + e^{-i\zeta}\mathbf{V}_- + e^{i2\zeta}\mathbf{W}_+ + e^{-i2\zeta}\mathbf{W}_-, \tag{9.67}$$

with

$$\mathbf{U} = \begin{pmatrix} -n & 0 & -i\kappa & 0 \\ 0 & n & 0 & i\kappa \\ f & s_+ & -n & 0 \\ s_- & -f & 0 & n \end{pmatrix}, \tag{9.68}$$

$$\mathbf{V}_+ = \begin{pmatrix} 0 & 0 & 0 & 0 \\ i\pi_+ & -i\pi_- & 0 & 0 \\ 0 & 0 & -i\pi_- & 0 \\ 0 & 0 & -i\pi_+ & 0 \end{pmatrix}, \quad \mathbf{V}_- = \begin{pmatrix} i\pi_+ & -i\pi_- & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & i\pi_- \\ 0 & 0 & 0 & i\pi_+ \end{pmatrix}, \tag{9.69}$$

$$\mathbf{W}_+ = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & ik_x{}^2/(2\epsilon_{33}) & 0 \\ 0 & 0 & 0 & 0 \\ -ik_x{}^2/2 & 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{W}_- = \begin{pmatrix} 0 & 0 & 0 & -ik_x{}^2/(2\epsilon_{33}) \\ 0 & 0 & 0 & 0 \\ 0 & ik_x^2/2 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{9.70}$$

and

$$n = \frac{2\pi}{k_0 P}, \tag{9.71}$$

$$\kappa = \left(1 - \frac{k_x^2}{2\epsilon_{33}}\right), \qquad \pi_\pm = \frac{k_x}{2\epsilon_{33}}(\epsilon_{23} \pm i\epsilon_{13}),$$

$$f = -\frac{i}{2\varepsilon_{33}}(\epsilon_{13}^2 + \epsilon_{23}^2 - \epsilon_{33}[\epsilon_{11} + \epsilon_{22} - k_x^2]),$$

$$s_\pm = \frac{1}{2\varepsilon_{33}}[2(\epsilon_{13}\epsilon_{23} - \epsilon_{12}\epsilon_{33}) \pm i(\epsilon_{23}^2 - \epsilon_{13}^2 + \epsilon_{33}\{\epsilon_{11} - \epsilon_{22}\})]. \tag{9.72}$$

Note that the substitutions $cos\varphi = (1/2)(exp[i\varphi] + exp[-i\varphi])$ and $sin\varphi = i(1/2)(exp[i\varphi] - exp[-i\varphi])$ were used here. Note further that the sign of P determines the handedness of the CLC structure. Solution for $\mathbf{T}_p$ is immediate at $\Phi_a = 0°$. For $k_x \neq 0$ the solution given below is exact but implicit.

9.4.3.1 Normal Incidence ($k_x = 0$)

For $k_x = 0$ the eigenvalues of $\tilde{\Delta} = \mathbf{U}$ are ($\mathbf{V}_+ = \mathbf{V}_- = \mathbf{W}_+ = \mathbf{W}_- = \mathbf{0}$)

$$q_\pm \equiv \sqrt{n^2 - if} \pm \chi, \quad \chi = \sqrt{s_+ s_- - i4n^2 f}, \tag{9.73}$$

which refer to both elliptical eigenmodes within the twisted medium. The coefficients β_i follow from Eqs. (9.52)

$$\beta_0 = \frac{1}{2\chi}[q_+^2 \cos\kappa_- - q_-^2 \cos\kappa_+],$$

$$\beta_1 = i\frac{1}{2\chi}[(q_+^2/q_-)\sin\kappa_- - (q_-^2/q_+)\sin\kappa_+],$$

$$\beta_2 = \frac{1}{2\chi}[\cos\kappa_+ - \cos\kappa_-],$$

$$\beta_3 = i\frac{1}{2\chi}[(1/q_+)\sin\kappa_+ - (1/q_-)\sin\kappa_-], \tag{9.74}$$

with the phase thicknesses for the elliptical eigenmodes

$$\kappa_\pm \equiv \zeta n^{-1} q_\pm = k_0 d q_\pm. \tag{9.75}$$

The partial transfer matrix in elliptical coordinates is calculated from β_i and $\mathbf{U}$

$$\tilde{\mathbf{T}}_\mathbf{p}(\zeta = \frac{2\pi}{P}d) = (\beta_0\mathbf{E} + \beta_1\mathbf{U} + \beta_2\mathbf{U}^2 + \beta_3\mathbf{U}^3). \tag{9.76}$$

After transformation into the laboratory coordinate system we obtain $\mathbf{T}_p$

$$\mathbf{T}_\mathbf{p}(d) = \Gamma^{-1}\mathbf{F}(\zeta = \frac{2\pi}{P}d)\tilde{\mathbf{T}}_\mathbf{p}(\zeta = \frac{2\pi}{P}d)\Gamma. \tag{9.77}$$

Note that the inverse of $\mathbf{T}_p$ is needed in Eq. (9.33). Because of the identities $[\tilde{\mathbf{T}}_p(\zeta)]^{-1} = \tilde{\mathbf{T}}_p(-\zeta)$ and $(\mathrm{AB})^{-1} = \mathrm{B}^{-1}\mathrm{A}^{-1}$ the inverse of $\mathbf{T}_p$ in the laboratory system is

$$\mathbf{T}_\mathbf{p}^{-1}(d) = \Gamma^{-1}(\beta_0\mathbf{E} - \beta_1\mathbf{U} + \beta_2\mathbf{U}^2 - \beta_3\mathbf{U}^3)\mathbf{F}(\zeta = -\frac{2\pi}{P}d)\Gamma. \quad (9.78)$$

9.4.3.2 Oblique Incidence ($k_x \neq 0$)

The Ansatz to calculate the field amplitudes within CLCs is similar to that used by Lakhtakia and Weiglhofer.[32] Exact solutions for $\mathbf{T}_p$ at off-axis propagation are obtained as follows. According to Eq. (9.67) the matrix $\tilde{\Delta}$ is analytical in ζ, ζ_0

$$\tilde{\Delta}(\zeta, \zeta_0) = \sum_{m=0}^{\infty} \delta_m(\zeta_0)(\zeta - \zeta_0)^m. \quad (9.79)$$

We assume that $\tilde{\mathbf{T}}_p$ is also analytical, and the 4×4 coefficient matrices $\mathbf{t}_\nu$ do not depend on ζ

$$\tilde{\mathbf{T}}_p(\zeta, \zeta_0) = \sum_{\nu=0}^{\infty} \mathbf{t}_\nu(\zeta_0)(\zeta + \zeta_0)^\nu. \quad (9.80)$$

The elements of the matrices $\mathbf{t}_\nu$ follow straightforwardly from $\tilde{\Delta}$. A differential equation for $\tilde{\mathbf{T}}_p$ is obtained from Eq. (9.50)

$$\partial_\zeta \tilde{\mathbf{T}}_\mathbf{p}(\zeta) = in^{-1}\tilde{\Delta}(\zeta)\tilde{\mathbf{T}}_\mathbf{p}(\zeta). \quad (9.81)$$

If we substitute $\tilde{\Delta}$ and $\tilde{\mathbf{T}}_p$ by the respective series expansions, and carry out the derivative with respect to ζ, we obtain a recurrence relation for the coefficient matrices of the series expansion for $\tilde{\mathbf{T}}_p$

$$\mathbf{t}_{m+1}(\zeta_0) = \frac{i}{n(m+1)} \sum_{k=0}^{m} \delta_{m-k}(\zeta_0)\mathbf{t}_k(\zeta_0), \quad (9.82)$$

where $\mathbf{t}_0 = \mathbf{T}_p(\zeta_0, \zeta_0)$ is the unit matrix $\mathbf{E}$. The coefficients for $\tilde{\Delta}$ follow straightforwardly from the Taylor series expansion

$$\tilde{\Delta}(\zeta, \zeta_0) = \sum_{m=0}^{\infty} \frac{\partial^m}{\partial m}\tilde{\Delta}(\zeta)\Big|_{\zeta=\zeta_0} \frac{(\zeta - \zeta_0)^m}{m!}$$

and read

$$\delta_0 = \mathbf{U}(\zeta_0) + \mathbf{V}_+(\zeta_0) + \mathbf{V}_-(\zeta_0) + \mathbf{W}_+(\zeta_0) + \mathbf{W}_-(\zeta_0),$$

$$\delta_n = \frac{1}{n!}[i^n\mathbf{V}_+(\zeta_0) + (-i)^n\mathbf{V}_-(\zeta_0) + (2i)^n\mathbf{W}_+(\zeta_0) + (-2i)^n\mathbf{W}_-(\zeta_0)], \quad n > 0. \tag{9.83}$$

The matrices δ_n depend on the orientation of the biaxial molecules at the boundary ζ_0. For convenience, and without loss of generality, we can set ζ_0 to zero. Note that the matrices $\mathbf{V}_+$ and $\mathbf{V}_-$ describe the effect of the tilt of the molecules towards the z-axis. $\mathbf{V}_+$ and $\mathbf{V}_-$ vanish if the CLCs are oriented parallel within the (x, y) plane. For $k_x = 0$ $\mathbf{W}_+$ and $\mathbf{W}_-$ also vanish, and the recurrence relation becomes the series expansion of $exp\{in^{-1}\mathbf{U}\}$, which is then calculated more efficiently through Eq. (9.76). Note that the inverse of $\mathbf{T}_p$ is needed in Eq. (9.33). $\mathbf{T}_p^{-1}$ in the laboratory system is

$$\mathbf{T}_\mathbf{p}^{-1}(d) = \Gamma^{-1}\tilde{\mathbf{T}}_\mathbf{p}\left(\zeta = -\frac{2\pi}{P}d\right)\mathbf{F}\left(\zeta = -\frac{2\pi}{P}d\right)\Gamma. \tag{9.84}$$

Figure 9.7 depicts the squared reflection Jones matrix elements of a usual CLC device cell on top of a transparent substrate (typically BK7, $\in_{\text{substrate}} = 2.25$ was chosen here for simplicity). A well-known drawback of standard CLC devices is their color distortion at large view angles. This effect is traced in Fig. 9.7 upon the change of the center frequency of the selective reflection band to higher wavelengths as the angle of incidence increases. Today's liquid crystal displays possess specific alignment layers that fix the liquid crystal molecules at the substrate-liquid crystal interface within small domains with nonunidirectional alignment [see, e.g., "Liquid Crystal Materials and Liquid Crystal Displays," M. Schadt, *Annual Review of Materials Sciences*, Vol. 27, 305–379 (1997)].†

9.4.4 Isotropic Dielectric Films

Although isotropic materials can be treated more easily and by simpler equations than through the 4×4 approach, the isotropic solution within the 4×4 matrix formalism is necessary in order to include isotropic layers in anisotropic samples. The eigenvalues of Δ for isotropic materials are well known

$$q_1 = q_2 = -q_3 = -q_4 \equiv q = \sqrt{\in - k_x^2} = \sqrt{n^2 - n_a^2 \sin^2 \Phi_a}, \tag{9.85}$$

†See also: M. Schubert, C. M. Herzinger, *phys. stat. sol. (a)*, **188**, 1563 (2001).

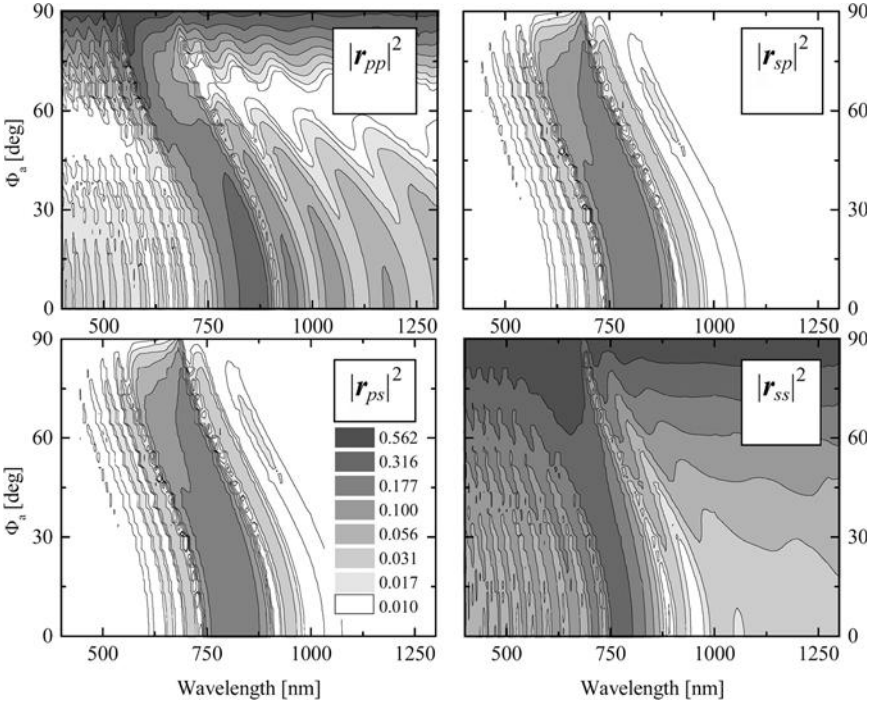


Figure 9.7 Squared Jones reflection matrix elements (intensities) of a chiral liquid crystal layer on a transparent substrate as a function of the angle of incidence and incident wavelength ($\epsilon_{\parallel} = 2.89$, $\epsilon_{\perp} = 2.25$, $\epsilon_{\text{substrate}} = 2.25$, $d = 3$ μm, $P = -0.5$ μm). The band of selective reflection is centered at $\lambda_0 \sim |P(\sqrt{\epsilon_{\parallel}} + 2\sqrt{\epsilon_{\perp}})/3| \sim 783$ nm for normal incidence, and moves to lower wavelengths at oblique angles of incidence.

with ϵ being the isotropic dielectric function. Equations (9.52) are degenerate and do not provide a solution for $\mathbf{T}_p$. However, the expansion of the exponential series in Eq. (9.50) and the separation of the associated sums for the cosine and sine functions result directly in

$$T_p = E\left\{1 + \frac{(ik_0dq)^2}{2!} + \frac{(ik_0dq)^4}{4!} + \cdots\right\} + i\frac{\Delta}{q}\left\{\frac{(k_0dq)}{1!} - \frac{(k_0dq)^3}{3!} + \cdots\right\}, \tag{9.86}$$

$$T_p = E\cos(k_0dq) + i\frac{\Delta}{q}\sin(k_0dq), \tag{9.87}$$

$$T_p = \begin{pmatrix} \cos k_0 dq & 0 & 0 & i\frac{q}{\varepsilon}\sin k_0 dq \\ 0 & \cos k_0 dq & -\frac{i}{q}\sin k_0 dq & 0 \\ 0 & -iq\sin k_0 dq & \cos k_0 dq & 0 \\ i\frac{\varepsilon}{q}\sin k_0 dq & 0 & 0 & \cos k_0 dq \end{pmatrix}. \quad (9.88)$$

where we have used the identities $\Delta^2 = q^2\mathbf{E}$ and $\Delta^3 = q^2\Delta$. The eigenvalue approach to find the coefficients β_i breaks down for extremely small anisotropy. In particular, if the differences $k_0 d|q_1^{\pm} - q_2^{\pm}|$ are too small, the attempt to invert Eqs. (9.52) fails. The determinant of the right side of Eqs. (9.52) can be used to check whether the optical response can be treated as anisotropic or isotropic during a regression analysis of experimental GE data.

9.4.5 Further Solutions: [1 1 1] Superlattice Ordering in III-V Compounds (CuPt-Ordering)

Special GE solutions for samples with small anisotropies, small layer thickness, or special optical axes configuration help to decrease the computational effort. Such solutions also help to shed light on typical features within the GE data. Sample symmetries or amplitude and strength of the birefringence read more easily from the experimental data. A solution is presented here for the special anisotropy in CuPt-ordered zinc-blende semiconductor films. Other anisotropy problems require different equations, and the line of derivation here can be transformed to other situations. Such equations have been recently presented by Lecourt and coworkers to determine orientation and birefringence in ultrathin Langmuir-Blodgett layers.[11]

Long-range chemical ordering has been widely observed during the last decade in ternary or quaternary zinc-blende III-V semiconductor compounds.[52–54] The CuPt-ordered phase consists of a 1,1-superlattice of monolayer planes stacked along one of the [111] directions. Similar to (1 1 1)-biaxial strain, CuPt ordering splits and shifts the zinc-blende states, e.g., near the valence-band maximum, and leads to polarized interband transitions and uniaxial optical birefringence.[55] This birefringence is very small and difficult to detect. However, we will demonstrate in Section 9.4 that GE can accurately determine the ordering-induced birefringence in CuPt-ordered semiconductor films. The present section derives special GE solutions for CuPt-ordered films. Although the solutions presented in Section 9.4.1 cover the case of this uniaxial dielectric material,

the present approach is more lucid, and permits better understanding of the relation between the uniaxial dielectric function values and the GE parameters. The latter are derived to first-order approximation in $\Delta\epsilon$, i.e., we assume $|\Delta\epsilon| \ll |\epsilon|$. Second-order and higher terms in $\Delta\epsilon$ are omitted.†

The inclination θ of the ordering direction from the sample normal amounts to $\cos^{-1}(1/\sqrt{3})$ for (001) surfaces. For this type of anisotropy the GE parameters have most the sensitivity if the sample is aligned with the [1 0 0] direction parallel to the x-axis of the laboratory coordinate system. Accordingly, the [1 $\bar{1}$ 0] direction is then rotated away from the plane of incidence by $\varphi = \pi/4$, but still parallel to the sample surface. (Equations for x || [110] are further discussed in Ref. 17.) The uniaxial dielectric function tensor ϵ in laboratory coordinates is then expressed as follows ($\theta = \cos^{-1}(1/\sqrt{3}) \sim 54.7°$)

$$\epsilon = \begin{pmatrix} \epsilon & -\Delta\epsilon & \Delta\epsilon \\ -\Delta\epsilon & \epsilon & -\Delta\epsilon \\ \Delta\epsilon & -\Delta\epsilon & \epsilon \end{pmatrix}, \tag{9.89}$$

where $\Delta\epsilon = 1/3\epsilon_{||} - 1/3\epsilon_{\perp}$, $\epsilon = 1/3\epsilon_{||} + 2/3\epsilon_{\perp}$, and $\epsilon_{||}$, $\epsilon_{\perp}$ refer to the dielectric functions for polarizations parallel and perpendicular to the ordering direction, respectively. (The symbol "Δ" used here in "$\Delta\epsilon$" is not to be confused with the matrix Δ or the ellipsometric parameter Δ). In Eq. (9.49), the coefficient matrix $\hat{\Delta}$ simplifies

$$\hat{\Delta} = \frac{1}{\epsilon}\begin{pmatrix} -k_x\Delta\epsilon & k_x\Delta\epsilon & 0 & \epsilon - k_x^2 \\ 0 & 0 & -\epsilon & 0 \\ \epsilon\Delta\epsilon & -\epsilon(\epsilon - k_x^2) & 0 & -k_x\Delta\epsilon \\ \epsilon^2 & -\epsilon\Delta\epsilon & 0 & -k_x\Delta\epsilon \end{pmatrix}. \tag{9.90}$$

The eigenvalues of $\hat{\Delta}$ are $q_{1,2} = \pm\sqrt{(\epsilon - k_x^2 - \Delta\epsilon)}$, and $q_{3,4} = -k_x\Delta\epsilon/\epsilon \pm \sqrt{(\epsilon - k_x^2 + \Delta\epsilon)}$. We define $q^2 = \epsilon - k_x^2$, and decompose $\hat{\Delta}$ into $\hat{\mathbf{D}} + (\Delta\epsilon)\hat{\delta}$ with

$$\hat{\mathbf{D}} = \begin{pmatrix} 0 & 0 & 0 & \frac{q^2}{\epsilon} \\ 0 & 0 & -1 & 0 \\ 0 & -q^2 & 0 & 0 \\ \epsilon & 0 & 0 & 0 \end{pmatrix}, \quad \hat{\delta} = \frac{1}{\epsilon}\begin{pmatrix} -k_x & k_x & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \epsilon & 0 & 0 & -k_x \\ 0 & -\epsilon & 0 & -k_x \end{pmatrix}. \tag{9.91}$$

†It is understood that $|\Delta\epsilon| = |Re\{\Delta\epsilon\}| + |Im\{\Delta\epsilon\}| \ll |Re\{\epsilon\}| + |Im\{\epsilon\}| = |\epsilon|$

The coefficients β_i follow from Eqs. (9.52) and do not depend on $\Delta\epsilon$. (Use q_i of $\hat{\Delta}$, and the rule of L'Hospital. Consider $\Delta\epsilon \to 0$.)

$$\beta_0 = \cos(k_0 dq) - \frac{1}{2} k_0 dq \sin(k_0 dq),$$

$$\beta_1 = (i/2q)\{-k_0 dq \cos(k_0 dq) - 3\sin(k_0 dq)\},$$

$$\beta_2 = (1/2q^2) k_0 dq \sin(k_0 dq),$$

$$\beta_3 = (i/2q^3)\{k_0 dq \cos(k_0 dq) + \sin(k_0 dq)\}. \tag{9.92}$$

$\mathbf{T}_p$ can be decomposed into $\hat{\mathbf{I}} + \Delta\epsilon\hat{\mathbf{B}}$ with

$$\hat{\mathbf{I}} = \hat{\mathbf{E}} \cos(k_0 dq) - \frac{i}{q} \hat{\mathbf{D}} \sin(k_0 dq), \tag{9.93}$$

$$\hat{\mathbf{B}} = (\gamma_1 + 2q^2\gamma_3)\hat{\delta} + \gamma_2\hat{\gamma} + \gamma_3\hat{\sigma}, \tag{9.94}$$

and $\hat{\gamma} = \hat{\mathbf{D}}\hat{\delta} + \hat{\delta}\hat{\mathbf{D}}$, $\hat{\sigma} = \hat{\mathbf{D}}\hat{\delta}\hat{\mathbf{D}}$, which result in

$$\hat{\gamma} = \begin{pmatrix} 0 & -\frac{q^2}{\epsilon} & -\frac{k_x}{\epsilon} & -2k_x\frac{q^2}{\epsilon^2} \\ -1 & 0 & 0 & \frac{k_x}{\epsilon} \\ -k_x & 0 & 0 & \frac{q^2}{\epsilon} \\ -2k_x & k_x & 1 & 0 \end{pmatrix},$$

$$\sigma = \begin{pmatrix} -k_x\frac{q^2}{\epsilon} & 0 & \frac{q^2}{\epsilon} & 0 \\ k_x & 0 & 0 & -\frac{q^2}{\epsilon} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -k_x & -k_x\frac{q^2}{\epsilon} \end{pmatrix}. \tag{9.95}$$

The identities $\hat{\mathbf{D}}^2 = q^2\hat{\mathbf{E}}, \hat{\mathbf{D}}^3 = q^2\hat{\mathbf{D}}$ are used. Note that $T_{p22} = \hat{\boldsymbol{I}}_{22}$, $T_{p23} = \hat{\boldsymbol{I}}_{23}$, $T_{p32} = \hat{\boldsymbol{I}}_{32}$, and $T_{p33} = \hat{\boldsymbol{I}}_{33}$ do not depend on $\Delta\epsilon$. T_{p11}, T_{p14}, T_{p41}, and T_{p44} are of type $T_{pij} = \hat{\boldsymbol{I}}_{ij} + \Delta\epsilon\hat{B}_{ij}$, whereas the remaining elements $T_{pij} = \Delta\epsilon\hat{B}_{ij}$ are linear in $\Delta\epsilon$. Let $\epsilon_f = n_f^2$ denote the isotropic substrate

dielectric function. The matrix **T** has non-vanishing elements in the first and third column only:

$$
\begin{aligned}
T_{11} &= [T_{p22} - T_{p23}N_{ff}] - [T_{p32} - T_{p33}N_{ff}]/N_{aa},\\
T_{21} &= [T_{p22} - T_{p23}N_{ff}] + [T_{p32} - T_{p33}N_{ff}]/N_{aa},\\
T_{31} &= [T_{p42} - T_{p43}N_{ff}]/n_a + [T_{p12} - T_{p13}N_{ff}]/\cos\Phi_a,\\
T_{41} &= [T_{p42} - T_{p43}N_{ff}]/n_a - [T_{p12} - T_{p13}N_{ff}]/\cos\Phi_a,\\
T_{13} &= [T_{p21}\cos\Phi_f + T_{p24}n_f] - [T_{p31}\cos\Phi_f + T_{p34}n_f]/N_{aa},\\
T_{23} &= [T_{p21}\cos\Phi_f + T_{p24}n_f] + [T_{p31}\cos\Phi_f + T_{p34}n_f]/N_{aa},\\
T_{33} &= [T_{p41}\cos\Phi_f + T_{p44}n_f]/n_a + [T_{p11}\cos\Phi_f + T_{p14}n_f]/\cos\Phi_a,\\
T_{43} &= [T_{p41}\cos\Phi_f + T_{p44}n_f]/n_a - [T_{p11}\cos\Phi_f + T_{p14}n_f]/\cos\Phi_a,
\end{aligned}
\tag{9.96}
$$

with

$$
N_{ij} \equiv n_i\cos\Phi_j = n_i\sqrt{1 - [(n_a/n_j)\sin\Phi_a]^2}. \tag{9.97}
$$

We will label the elements of **T** as T'_{ij} if we set $T_{pij} = \hat{\boldsymbol{I}}_{ij}$, and T''_{ij} if we set $T_{pij} = \hat{B}_{ij}$. Note that T_{11} and T_{21} do not depend on $\Delta\epsilon$, whereas T_{31}, T_{41}, T_{13}, and T_{23} are proportional to $\Delta\epsilon$. The elements T_{33} and T_{43} are of the type $a + b\Delta\epsilon$. The GE parameters in Eqs. (9.42) and (9.43) are expanded in $\Delta\epsilon$ according to $[c/(a + b\Delta\epsilon)]\Delta\epsilon \sim (c/a)\Delta\epsilon$, or $[(A + \Delta\epsilon B)/(C + \Delta\epsilon D)] \sim (A/C) + \Delta\epsilon[B/C - (A/C)(D/C)]$. This allows factorization of R_{ps}, R_{sp}, and R_{pp} into

$$
R_{ps;sp} = \Delta\epsilon\beta_{ps;sp} + O((\Delta\epsilon)^2), \tag{9.98}
$$

$$
R_{pp} = (\boldsymbol{r}_p/\boldsymbol{r}_s)[1 + \Delta\epsilon\beta_{pp} + O((\Delta\epsilon)^2)], \tag{9.99}
$$

with the coefficients β_{ps}, β_{sp}, and β_{pp}

$$
\beta_{ps} = T''_{23}/T'_{43} - \boldsymbol{r}_s T''_{13}/T'_{43}, \tag{9.100}
$$

$$
\beta_{sp} = T''_{41}/T'_{21} - \boldsymbol{r}_p T''_{31}/T'_{21}, \tag{9.101}
$$

$$
\beta_{pp} = T''_{43}/T'_{43} - T''_{33}/T'_{33}, \tag{9.102}
$$

and similarly for T_{ps}, T_{sp}, and T_{pp}

$$
T_{ps;sp} = \Delta\epsilon\alpha_{ps;sp} + O((\Delta\epsilon)^2), \tag{9.103}
$$

$$
T_{pp} = (\boldsymbol{t}_p/\boldsymbol{t}_s)[1 + \Delta\epsilon\alpha_{pp} + O((\Delta\epsilon)^2)], \tag{9.104}
$$

with the coefficients α_{ps}, α_{sp}, and α_{pp}

$$\alpha_{ps} = -T''_{13}/T'_{11}, \tag{9.105}$$

$$\alpha_{sp} = -T''_{31}/T'_{33}, \tag{9.106}$$

$$\alpha_{pp} = -T''_{33}/T'_{33}, \tag{9.107}$$

The isotropic ratios $\boldsymbol{r}_p$ ($\boldsymbol{t}_p$) and $\boldsymbol{r}_s$ ($\boldsymbol{t}_p$) are those for the complex p and s polarized reflectance (transmittance) of an isotropic ambient-film-substrate system with $\in$ as the isotropic dielectric constant of the film. The GE parameters R_{pp} and T_{pp} evolve linearly from the isotropic reflectance and transmittance ratios ρ_r^{is}, and ρ_t^{is}, and determine $\in$. The ratios R_{ps} (T_{ps}) and R_{sp} (T_{sp}) are most sensitive to $\Delta\in$, but depend still on $\in$. Simultaneous analysis of the GE parameters can provide both $\in$ and $\Delta\in$ unambiguously (see Sect. 9.4). The coefficients α and β are not to be confused with the coefficients β_i in Eqs. (9.52). Note that α and β possess similarity with the well-known Seraphin coefficients for isotropic semiconductor surfaces.[57] Note finally that the approximate equations are correct as long as $|\Delta\in| \ll |\in|$ regardless of the thickness d.

9.5 Strategies in Generalized Ellipsometry

This section surveys strategies for treatment of several types of anisotropies. Issues of depolarization due to incoherent backside effects are discussed. Model strategies are given for absorbing and transparent substrates and films. Combinations of anisotropy situations discussed here may be treated by joining the strategies given here. More complicated situations will require other specific procedures.

No universal rule exists for best treatment of general anisotropy situations, except to take as much data as necessary from as many independent directions as possible. The strategies for measurement and analysis will depend on the actual sample properties. If the nature of the anisotropy is unknown, several strategies need to be tested in order to find out which type of anisotropy belongs to the sample under investigation. For known types of anisotropy qualitative simulations prior to actual data acquisition have been shown to be very useful,[8, 15, 19, 34] and the Jones matrix elements should be surveyed as a function of the sample azimuth and the angle of incidence to learn about propitious conditions for data acquisition (see Section 9.5.1).

Further difficulties arise if the substrate material is transparent. In that case depolarization due to multiple discrete polarization states needs to be

considered. Depolarization may also occur due to nonuniform thickness, finite optical bandwidth, angular spread of the incident beam and nonuniformities in the optical constants. Such situations are discussed in Ch. 3 by Jellison, and are not included here. Depolarization due to multiple discrete polarization states with simultaneous cross coupling of eigenmodes affects the ellipsometric data acquisition procedure. For the latter there is no effective Jones matrix to be obtained through regression analysis of (α, β) sets in Eqs. (9.18) through (9.20). Strategies for treatment of incoherent backside effects in thick anisotropic layers are given in Section 9.5.2.

For data analysis of anisotropic samples it is highly desired to reduce the number of fit parameters. This can be done by employing parameterized dielectric function models (see also Ch. 3 by Jellison). Sometimes, however, point-by-point fits are more straightforward and may reveal spectral features, which are otherwise subsumed into the lineshape of the model functions used. Additional information such as crystal axis orientations, sample thickness, and estimates for optical constants data are then extremely valuable for the initial parameter guess. Issues concerning anisotropic model strategies are discussed in Section 9.5.3.

9.5.1 Data Acquisition Strategies for Anisotropic Samples

Absorbing substrate with optical axis not parallel to the sample normal (uniaxial material): GE data should be taken at multiple sample azimuth orientations φ, and angles of incidence Φ_a. If the in-plane orientation of the optical axis is unknown, the spacing between different orientations should be less than 90°. Data should also be measured at opposite φ. Measurements at near-normal incidence can help to isolate the difference $\Delta\epsilon = \epsilon_{||} - \epsilon_{\perp}$. The off-diagonal elements may be difficult to detect if $\Delta\epsilon$ is small, or if the optical axis is close to the sample normal. Many polarizer and compensator settings should be used, and the number of data points (α, β) included into the regression analysis for the GE parameters should be large. Standard ellipsometry measurements at the pseudo-Brewster angle can establish the absolute values for $\epsilon_{||}$ and $\epsilon_{\perp}$. For the standard ellipsometry measurements the position of the polarizer should be held fixed, most conveniently at 45°. Coherent beams reflected off the backside help to isolate the index difference in uniaxial films. It is generally good advice to measure GE data over a large range of incidence angles.

Absorbing film on an isotropic but absorbing substrate with optical axis not parallel to the sample normal (uniaxial material): This case is similar to that of the absorbing substrate above. However, helpful thickness

oscillations may occur if the optical thickness of the film is not too small. The off-diagonal ratios may become very large, especially for wavelengths at which the *p*-to-*p* or *s*-to-*s* reflection coefficients vanish. GE parameters at near normal incidence help to differentiate between $\in_{||}$ and $\in_{\perp}$. The pseudo-Brewster angle is no longer a well-defined quantity because of the wavelength-dependent thickness oscillations. GE parameters at large angles of incidence will provide $\in_{||}$ and $\in_{\perp}$. Note that the error bars on R_{ps} and R_{sp} will be naturally large when the *p*-to-*p* or *s*-to-*s* reflection coefficients vanish (Ψ_{ps} or Ψ_{sp} reach 90°).

Substrate is absorbing and optical axis is parallel to the sample normal (uniaxial material): This situation usually indicates proper determination of $\in_{||}$. The off-diagonal ratios are zero regardless of $\Delta\in$, and standard ellipsometry parameters can be measured as usual. Ellipsometric measurements near normal incidence will be of no value. For small anisotropy it is commonly observed that $\Delta\in$ cannot be measured in substrates.[4, 8, 20, 56] However, if $\Delta\in$ is large, as seen for non-cubic solids within spectral regions of strong lattice absorptions, both $\in_{||}$ and $\in_{\perp}$ are accessible from standard ellipsometry measurements upon appropriate parameterization.[58, 59]

Transparent film with optical axis parallel to the sample normal on isotropic substrate (uniaxial material): The off-diagonal ratios are zero regardless of $\Delta\in$, and standard ellipsometry parameters can be measured. Ellipsometric measurements near the pseudo-Brewster angle will be useful. In order to detect the index-difference measurements away from and on either side of the Brewster angle are necessary. If measured at multiple angles of incidence the dependence of the interference oscillations on the angle of incidence reveal both $\in_{||}$ and $\in_{\perp}$.[60, 61] For infrared wavelengths $\Delta\in$ can be large and complex due to strong lattice absorptions. Simultaneous extraction of $\in||$ and $\in\perp$ is possible only if model functions for the dielectric response are employed. This situation is discussed in Refs. 62 and 63 for $Al_xGa_{1-x}N$ films on (0001) sapphire.

Anisotropic transparent substrate: The thickness of the substrate is a significant sample parameter to be known or to be varied during data analysis. Independent measurement of the thickness is of great value. The general anisotropy situation may be very complicated, even for a uniaxial material. If the differences between the major dielectric function values are large, beams of co-propagating eigenmodes, split by the anisotropy, may become incoherent after traveling over some distance within the anisotropic substrate (temporal beam separation due to strong birefringence). We will not consider this case here, although it could in principle be handled within the 4×4 matrix formalism. Large spatial separation may also occur. The difficulty is to unify the splitted beams, which emerge from the sample in order to guide them within the same light path through the ellipsometer into the same detector area. Here we assume small spatial

separation of the eigenmodes (compared to the size of the beam). Two co-propagating eigenstates are further assumed to maintain internal coherence. Eigenstates of different internal bounces are treated as incoherent with respect to other beams on different bounces at the front or backside of the substrate. In reflection, this "backside" effect can sometimes be suppressed by roughening the backside or by wedge-shaped samples. In transmission this effect cannot be suppressed, however, the effect will be smaller because the second-order eigenmodes undergo two lossy reflections at the front- and backside whereas the primary beam does not. When internal coherence of the eigenmodes is maintained, and higher-order internal reflections can be put last, the transmitted beam emergent from the sample is not depolarized, and the Jones matrix approach is fully valid. GE parameters can be acquired as for the situations discussed above. If reflection type data are required, proper treatment of depolarization is necessary. For the case of depolarization there is no effective Jones matrix to describe the sample. The regression for the GE parameters fails in general. Unregressed ellipsometry data, possibly converted into Ψ and Δ values for storage, are necessary. Useful unregressed data are pairs of α and β obtained at polarizer positions, e.g., $P = 0°$, $45°$ and $90°$ (to begin with the compensator function should be turned off, i.e., $C = w = 0$), and converted for convenience into Ψ and Δ with Eqs. (9.21). The unregressed data are the only data commensurable with model calculations in order to consider the "backside" induced depolarization effects appropriately. Note again that the transmitted data may also be affected by depolarization if the secondary beam is not insignificant in intensity as compared to the first one. Unregressed data are then required for transmission ellipsometry measurements as well. See also Sections 9.2.1 and 9.5.2.

Transparent substrate with optical axis not parallel to the sample surface but far away from the sample normal (uniaxial material): Polarized transmission intensity (PTI) measurements yield the squared transmission matrix elements. PTI data at normal incidence can provide a quick way to check for anisotropy before ellipsometric data are acquired.[19] If anisotropy is known to exist, PTI measurements can also provide a quick way to locate an optical axis. GE and PTI data should be taken at multiple sample azimuth orientations φ and $-\varphi$. The spacing between different orientations should be less than $90°$ for unknown optical axis orientations. GE parameters in transmission are extremely useful for determining the index differences, if depolarization is not an issue. Include positive and negative angles of incidence for transmission data, if possible, to avoid sample rotation from φ to $-\varphi$. The transmission ellipsometry data are of less use for the absolute index values. Angle-of-incidence scans of unregressed reflection ellipsometry data are very helpful for determining the absolute

index values (see Section 9.6.2). The co-propagating eigenstates produce a characteristic interference pattern according to the effective birefringence, which the propagating beams experience within the substrate. The effective birefringence depends on the sample azimuth, the index difference, the angle of incidence, the substrate thickness, and to some extent the absolute index values. The optical path difference between co-propagating eigenstates, i.e., the effective birefringence times the thickness divided by the vacuum wavelength, may be very large. The period of the interference pattern may become small, and depolarization due to smearing of polarization states by the finite spectral resolution of the ellipsometric equipment may occur. Unregressed data are then necessary as well. However, the latter effect makes transmission ellipsometry a very helpful tool to study uniaxial transparent substrates cut with the *c*-plane as the surface. Such *c*-plane cuts almost always have a slight misorientation of the *c*-axis away from the sample normal by a few tenths of a degree.

Transparent substrate with optical axis slightly off the sample normal (uniaxial material): Such misorientation will cause subtle anisotropy effects within the off-diagonal ratios (if beams are not depolarized), or within the unregressed data (if beams are depolarized). The transmission data will allow precise localization of the optical axis upon data analysis. If the sample can be oriented such that the refracted beam runs along the optical axis, a zero-crossing will occur in Δ_{pp}. If the angle of incidence and Euler angle φ are varied slightly around this orientation, the change in Δ_{pp} is extremely sensitive to $\Delta\in$ and θ, φ_0 (where φ_0 denotes the in-plane offset orientation of the optical axis).

Transparent substrate with optical axis exactly parallel to the sample normal (uniaxial material): Transmission ellipsometry is of no use. The co-propagating eigenstates do not split regardless of $\Delta\in$. Standard ellipsometry in reflection will provide $\in_{\perp}$. If $\Delta\in$ is small there is not much information available for $\in_{||}$. Note that depolarization due to backside reflections can be traced using standard ellipsometry data because the co-propagating eigenstates do not split (no cross coupling of modes), and the ellipsometry data do not depend on the actual polarizer setting.

9.5.2 Strategies for Treatment of Sample Backside Effects

The eigenstates, maintaining internal coherence with negligible spatial separation, are treated as incoherent with respect to other beams on different bounces at the front- or backside of the substrate. We do not

consider the case where co-propagating eigenmodes within the anisotropic substrate split due to strong anisotropy.

Anisotropic backside effects can be conveniently treated within the mathematical formalism developed in this chapter. Let $\mathbf{T}_+$ and $\mathbf{T}_-$ denote the general transfer matrices for the "coherent" layer stacks at the front and backside of the sample, respectively. The matrices $\mathbf{r}_+$ ($\mathbf{t}_+$) and $\mathbf{r}_-$ ($\mathbf{t}_-$) are Jones reflection (transmission) matrices, and follow from $\mathbf{T}_+$ and $\mathbf{T}_-$, respectively

$$\begin{pmatrix} B_p \\ B_s \end{pmatrix} = \mathbf{r}_+ \begin{pmatrix} A_p \\ A_s \end{pmatrix}, \qquad \begin{pmatrix} C_p \\ C_s \end{pmatrix} = \mathbf{t}_+ \begin{pmatrix} A_p \\ A_s \end{pmatrix}, \qquad D_p = D_s = 0, \tag{9.108}$$

$$\begin{pmatrix} D'_p \\ D'_s \end{pmatrix} = \mathbf{r}_- \begin{pmatrix} C'_p \\ C'_s \end{pmatrix}, \qquad \begin{pmatrix} E_p \\ E_s \end{pmatrix} = \mathbf{t}_- \begin{pmatrix} C'_p \\ C'_s \end{pmatrix}, \qquad F_p = F_s = 0. \tag{9.109}$$

To obtain their elements consider Eq. (9.29) and Eq. (9.30) as systems of four linear relations between the modes A, B, C, D, and E. See also Section 9.3.1. Note that modes D and F are zero when we ask for transmission and reflection coefficients for light incident from left to right in Fig. 9.8. To include the beams which are internally reflected off the backside of the sample we need the reflection and transmission coefficients for light incident from right to left onto the backside of the top layer stack in Fig. 9.8. These reflection and transmission matrices are denoted by $\mathbf{r}_{+i}$ and $\mathbf{t}_{+i}$, respectively, and follow from $\mathbf{T}_+$

$$\begin{pmatrix} B_p \\ B_s \end{pmatrix} = \mathbf{r}_{+i} \begin{pmatrix} D_p \\ D_s \end{pmatrix}, \qquad \begin{pmatrix} C_p \\ C_s \end{pmatrix} = \mathbf{t}_{+i} \begin{pmatrix} D_p \\ D_s \end{pmatrix}, \qquad A_p = A_s = 0. \tag{9.110}$$

For the latter case the A modes incident from the left to the right are not permitted and set to zero. The modes within the substrate change upon traveling over distance d

$$\begin{pmatrix} D_p \\ D_s \end{pmatrix} = \mathbf{m}_- \begin{pmatrix} D'_p \\ D'_s \end{pmatrix}, \qquad \begin{pmatrix} C'_p \\ C'_s \end{pmatrix} = \mathbf{m}_+ \begin{pmatrix} C_p \\ C_s \end{pmatrix}, \tag{9.111}$$

$$\mathbf{m}_\pm = \begin{pmatrix} \exp[ik_0 dq_i^\pm] & 0 \\ 0 & \exp[ik_0 dq_2^\pm] \end{pmatrix}. \tag{9.112}$$

where $q_i^\pm$ are the eigenvalues of the characteristic matrix Δ, which follows from Eq. (9.49), with $\in$ built according to the orientation of the substrate. The zero-order reflected beam amplitudes follow from the front reflection matrix $\mathbf{r}_+$ (modes A are set by the ellipsometer)

$$B^{(0)} = \mathbf{r}_+ A. \tag{9.113}$$

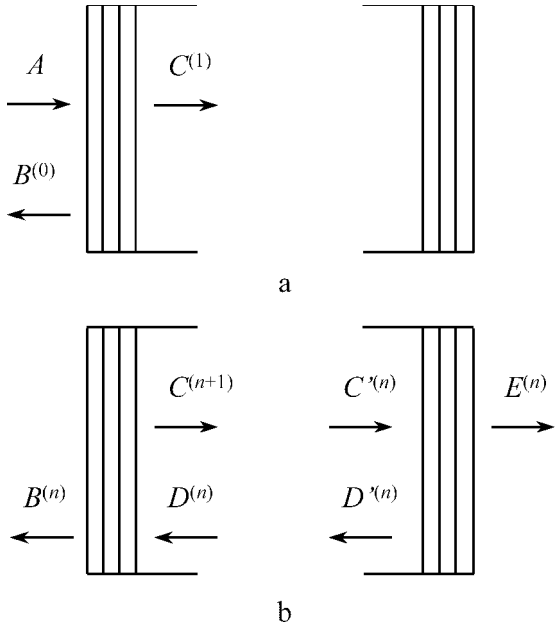


Figure 9.8 Schematic presentation of multiple beams reflected off internally at the front and backside of the substrate. The incident beam A propagates coherently through the top layer stack, and the zero-order beam $B^{(0)}$ is reflected from the front side (Fig. 9.8a). The transmitted beams $C^{(n)}$ reach the backside ($C'^{(n)}$) and cause the n^{th}-order beams $E^{(n)}$, which are transmitted out of the backside (Fig. 9.8b). The beams $D'^{(n)}$ ($D^{(n)}$) then produce the n^{th}-order front side transmitted beams $B^{(n)}$, as well as the next order of internally reflected beams $C^{(n+1)}$.

The first-order transmitted beam is obtained from

$$E^{(1)} = \mathbf{t}_{-}\mathbf{m}_{+}\mathbf{t}_{+}A, \tag{9.114}$$

and the first-order reflected beam reads

$$B^{(1)} = \mathbf{t}_{+i}\mathbf{m}_{-}\mathbf{r}_{-}\mathbf{m}_{+}\mathbf{t}_{+}A. \tag{9.115}$$

Higher-order amplitudes are computed accordingly. For analysis of transmission-type data it is often sufficient to consider the first-order beam only because the higher-order reflected beams undergo lossy internal reflections twice per round trip. In that case Eqs. (9.18) through (9.20) are still valid because only one beam is entering the detector. For analysis of backside effects in reflection-type data consideration of higher-order

beams is unavoidable. In that case Eqs. (9.18) through (9.20) are valid to some extent only. Multiple incoherent beams contribute to the Fourier coefficients in Eq. (9.17). (Each beam brings another set of quantities R_1 and R_2, which depend on the normalized Jones matrix parameters of the beam.) There is no effective Jones matrix to describe the sum effect of many beams for different polarizer, compensator, or retardation settings. The attempt fails to fit one set of GE parameters to Eqs. (9.18) through (9.20). Unregressed data are the only values which are comparable to model calculations in a straightforward manner. Unregressed ellipsometry data are the coefficients α and β taken at prearranged settings of polarizer and compensator azimuths and retardation values. (See Sections 9.2.1 and 9.2.4.) These settings are used later during data analysis as input values to model the target data α and β. Higher-order beams bear a certain danger of not fully propagating within the same path as the zero-order beam (partially spatial separation). Higher-order beams contribute to the ellipsometry result as long as parts enter the detector, but their absolute intensity is used to normalize the Fourier coefficients. An additional fit parameter, which mimics the actual beam intensity during measurement, needs to be included for data analysis.

9.5.3 Model Strategies

Best results are obtained if data from many samples with different orientations of their major axes are available. At least four azimuth orientations should be used for uniaxial samples with their *c*-axis not parallel to the surface normal. Include data from more than six azimuths and large ranges of incidence angles for modeling biaxial samples with arbitrary axis orientations. The azimuths should span one full sample turn. Orientations opposite in φ help to locate the in-plane orientation of the optical axes. Data from all samples and from all sample azimuths are best included into a multiple-sample, multiple-data-type regression analysis. To begin with, data at a single wavelength should be selected to locate the Euler angles for each data set, and to guess the initial major index values. Use also Fig. 9.5.

Euler angles: Observe the behavior of Ψ and Δ at opposite sample positions. It is helpful to perform numerical simulations with estimated major index values, and vary the Euler angles and the sample azimuth. Pay attention to the off-diagonal Ψ and Δ values. In particular, the pair Δ_{ps} and Δ_{sp} will help to locate the crystal coordinate system with respect to the (x, y) plane. The pair Ψ_{ps} and Ψ_{sp} provides hints for the Euler angle θ. Carefully inspect the measured ellipsometry data to guess what the actual orientation of the sample was during each measurement. Note that the

Euler angle θ is the same for a uniaxial sample at different sample azimuths. Use the amount and handedness from each azimuth change to constrain the Euler angles during the fit. For example, if data were taken at azimuths φ_0, $\varphi_0 + 90°$, $\varphi_0 + 180°$, and $\varphi_0 + 270°$ only the offset φ_0 needs to be fitted. In this case the sample was rotated clockwise.

Optical constants: A Cauchy model is usually sufficient to parameterize the major refractive indices within spectral regions of insignificant absorption. For uniaxial materials it is sometimes helpful to fit $\Delta\epsilon = (\epsilon_{\parallel} - \epsilon_{\perp})/3$ and $\epsilon' = (\epsilon_{\parallel} + 2\epsilon_{\perp})/3$ instead of $\epsilon_{\parallel}$ and $\epsilon_{\perp}$. Similar substitutes for biaxial materials may work too. Data from high-symmetry sample orientations provide a good estimate to the value of the major refractive indices. Such orientations occur when the crystal coordinate system coincides with (x, y, z). Care should be taken not to reverse the order of ϵ_x, ϵ_y, and ϵ_z. This can happen if near-normal incidence transmission data are available only. A point-by-point fit is a good procedure to begin with for spectral regions of high absorption. The point-by-point fitted optical constants can later be analyzed employing parametric models. It is very helpful if the optical axis orientations are known from data analysis within the transparent region prior to the point-by point fit into the high absorption region. The overlap region between transparency and high absorption is often difficult to fit. Scattering and non-homogeneous absorption within the film or substrate may occur in this spectral region. If the nature of the onset of absorption is known (e.g., direct or indirect band-to-band transitions; see also Ch. 3 by Jellison) appropriate parametric model approximations can help to differentiate between the small major absorption indices. (See paragraph on the CuPt order-induced birefringence.)

9.6 Generalized Ellipsometry Applications

The GE formalism has been employed by many workers to measure optical birefringence in, e.g., anisotropic bulk materials,[8, 10, 20–24] stretched plastic sheets[19], ultrathin Langmuir-Blodgett films,[11] sculptured films,[12] chiral-liquid crystal (CLC) cells,[15] porous silicon Fabry-Perot resonators,[18] magneto-optical films,[65] and birefringent III-V semiconductor films[17, 66]. Optical constant spectra of well-known materials were updated or extended into unknown spectral regions revising existent n and k spectra for crystals such as Rutile or sapphire. Electronic band-structure parameters were derived from the order-induced birefringence in III-V compound materials, and order parameter, structure, and intrinsic refractive indices were determined for CLCs. The following section provides examples for such materials studied by GE. Cases of transparent and absorbing substrates are included. Film examples are discussed further. Physical origins of birefringence are outlined for the chemically ordered cubic semiconductors,

and for the optical response of chiral liquid crystal cells, and demonstrated upon real data obtained from previous experiments. The selection of anisotropy situations given here is by far not complete, and the field of GE application continues to grow.[†]

9.6.1 Anisotropic Bulk Materials

Anisotropy in reasonably sized crystals is traditionally investigated by standard reflectometry or transmission techniques. The most accurate tool is the prism refractometry method in which the angular deviation of the probe beam is converted into the difference between a known and an unknown index of refraction. This method is more precise than the ellipsometry technique, but fails for spectral regions of absorption, and is not applicable to films. The prism method also becomes entangled for biaxial materials with arbitrary axis orientation. For such cases the GE approach uniquely resolves crystal axes and anisotropic optical constants. The precision of the extracted index data depends on the accuracy of the ellipsometer components, and may be less than for those obtained for isotropic materials by the prism method. However, the n and k spectra can be traced into the opaque spectral region. A general comment regarding data analysis of bulk materials must be added. It is well known that dielectric overlayers or surface roughness with thickness of a few nanometers cause deviation from the true bulk spectra. This overlayer may consist of naturally grown oxide or contaminants. Surface roughness within the nm scale produces similar effects. A common procedure for numerical roughness overlayer removal is to include a thin film with thickness d_{rough}. The roughness-layer optical constants are modeled through an effective medium approach (EMA) by averaging "void" to the bulk optical constants. Naturally grown oxides or contaminants can be removed numerically if their optical constants are known. Chemical treatment or polishing may induce new overlayer reactants or surface-damage induced strain. (See, e.g., discussion of surface aging effects to the anisotropic optical constants of HgI_2 by A. En Naciri[64].)

9.6.1.1 TiO_2 (Rutile)

Rutile (TiO_2) has tetragonal symmetry with uniaxial anisotropy. Single crystals of TiO_2 cut at different crystallographic orientations are

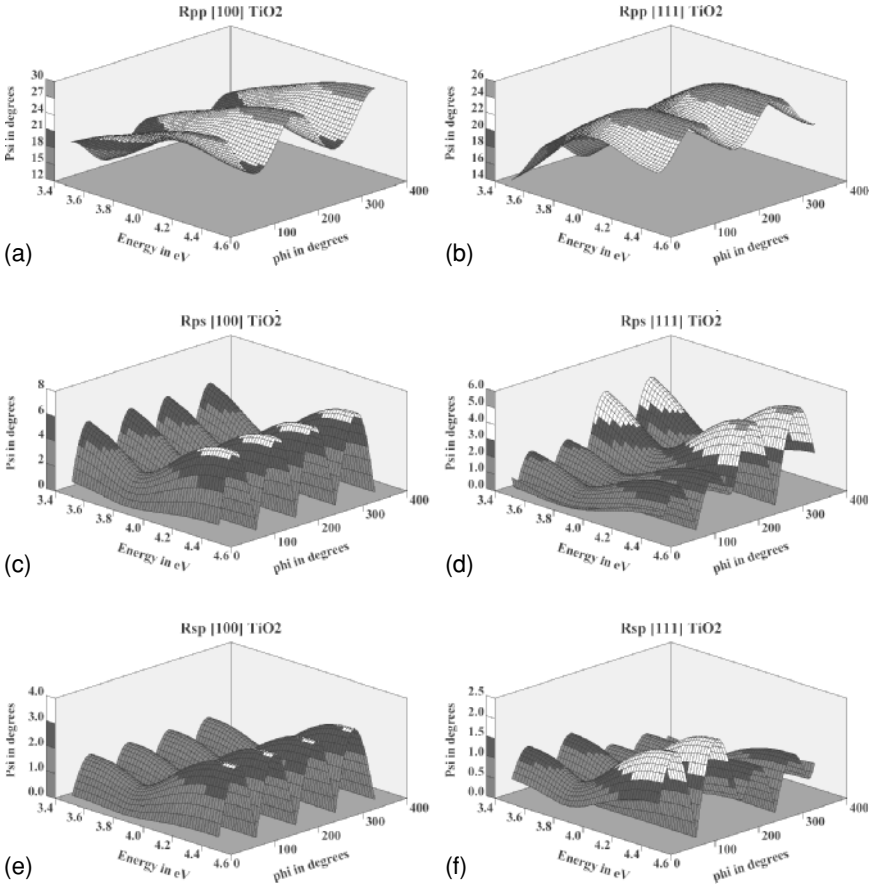
[†]Applications now also include biaxial absorbing minerals, as demonstrated for orthorhombic ore minerals such as stibnite in "Generalized ellipsometry for biaxial absorbing materials: determination of crystal orientation and optical constants of Sb2S3," M. Schubert and W. Dollase, *Opt. Lett.*, **27**, 2073–2075 (2002).

chosen for demonstration of the GE formalism to determine optical constants and crystal axis orientations from uniaxial substrate materials. The backsides of the samples were roughened to suppress incoherent backside effects. Standard substrates (Commercial Crystal Lab., Florida) with high-symmetry cuts ([100], [110], [111], [001]) were studied. Typical sample thickness is ~ 900 μm. For the [100] and [110] samples the *c*-axis is parallel to the surface, whereas for the [001] sample the *c*-axis is parallel to the surface normal. The optical axis coincides with the *c*-axis. The inclination θ of the optical axis from the surface normal ([111]) is ~ 40° for the [111] cut surface.

For calculation of the GE parameters the general transfer matrix $\mathbf{T}$ follows from the product of $\mathbf{L}_a$ and $\mathbf{L}_f$. The exit matrix $\mathbf{L}_f$ is built from the eigenmodes $\Xi_i(q_i)$ of the characteristic matrix Δ, which is the same as in Eq. (9.49). The eigenvalues q_i are obtained from Eqs. (9.53), and the $\Xi_i(q_i)$ can be easily computed from there.

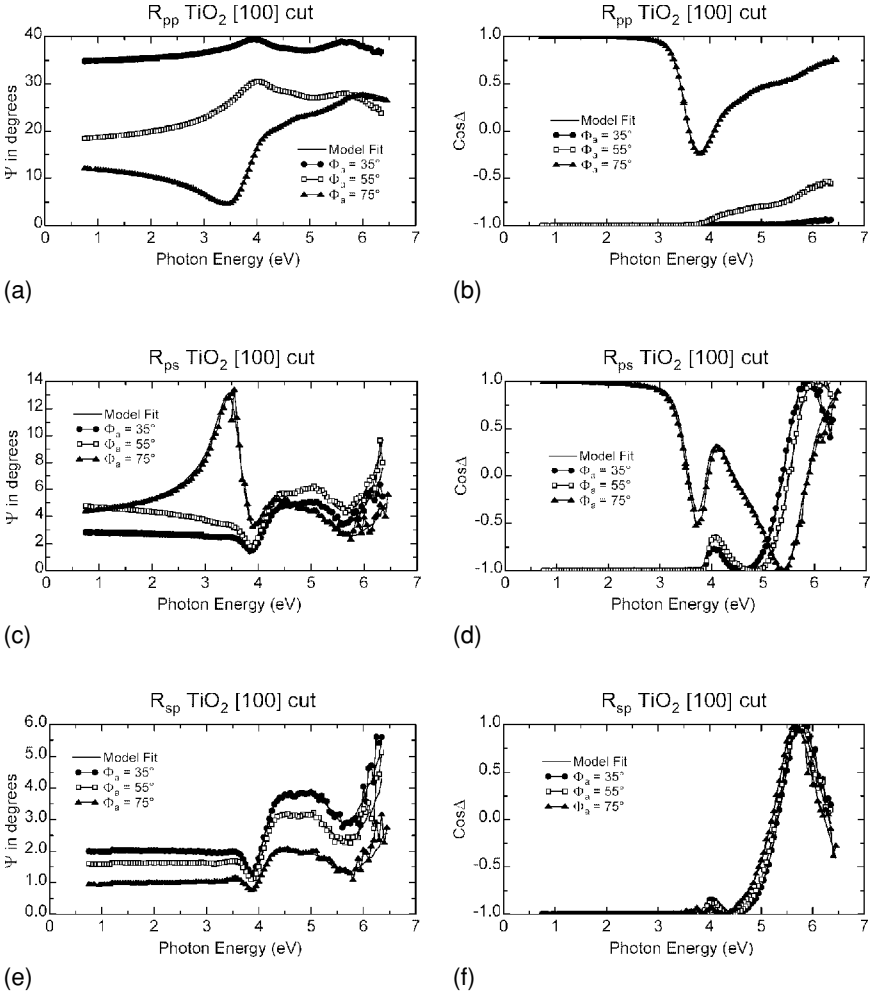
To begin with here, numerical simulations of GE parameters for *a*-plane ([100]-cut) and *r*-plane ([111]-cut) TiO_2 are shown in Figs. 9.9 a through f. Data for $\in_{||}$, $\in_{\perp}$ are taken from Ref. 8. For the [100] cut the off-diagonal GE parameters Ψ_{ps} and Ψ_{sp} are equal for opposite sample azimuths φ. Note that Ψ_{ps} and Ψ_{sp} vanish when the optical axis is oriented parallel to either one of the *x* or *y*-axes of the laboratory coordinate system. Ψ_{ps} and Ψ_{sp} are fourfold degenerate because the optical axis is parallel to the sample surface. The fourfold symmetry in Ψ_{ps} and Ψ_{sp} reduces when the optical axis is not within the (*x*, *y*) plane (e.g., for the [111] cut). Ψ_{ps} and Ψ_{sp} are no longer equal at opposite sample azimuths φ. Note that Ψ_{ps} and Ψ_{sp} are zero when the optical axis is parallel to the (*x*, *z*) plane, i.e., the plane of incidence. Ψ_{ps} and Ψ_{sp} vanish completely and regardless of $\Delta\in$ if the *c*-axis becomes collinear to the sample normal (*c*-plane TiO_2, not shown here). As seen from Figs. 9.9 a through f the GE parameters provide sensitivity to sample azimuth and orientation of the *c*-axis. These figures indicate the ability of GE to locate optical axis coordinates from anisotropic materials. (Note that for uniaxial materials we cannot differentiate between crystal directions perpendicular to the optical axis.)

According to the 3D plots, except for the [001] sample, GE data were taken at four different sample azimuths φ separated by 90°. The azimuths were chosen such that the optical axis had a 45°-type in-plane azimuth (45°, 135°, and so forth), where the off-diagonals are maximized. At each sample position further spectra were measured at multiple angles of incidence. This provided excellent sensitivity to the Euler angles and optical constants for both ordinary and extraordinary direction. (Note that the *c*-plane sample here provides sensitivity to $\in_{\perp}$ only. This situation is different for the optical constants within the IR-spectral range. See Refs. 58 and 59.)



Figures 9.9 a–f Simulated Ψ_{pp}, Ψ_{ps} and Ψ_{sp} for *a*-plane (a, c, e) and *r*-plane TiO_2 (b, d, f) for 3.5 eV $\leq E \leq$ 4.5 eV versus azimuth φ. $0° < \varphi < 360°$, $\Phi_a = 65°$; $\theta = \psi = 0°$ for [100] TiO_2; $\psi = 0°$ and $\theta = 50°$ for [111] TiO_2; $\in_{||}$, and $\in_{\perp}$ are taken from Ref. 8. As discussed in Sec. 9.4 the Euler angles φ, ψ, and θ provide the orthogonal rotation matrices that diagonalize the dielectric function tensor $\in$. Here φ and θ are the azimuth and the inclination angles of the optical axis with respect to *x* and *z*, respectively, whereas ψ is meaningless. See also Fig. 9.5. The Euler angle φ in Eq. (9.47) is equal to the sample azimuth because $\in_x = \in_{||}$, $\in_y = \in_z = \in_{\perp}$.

Examples for experimental and best-fit calculation of GE reflection parameters are presented in Figs. 9.10 a through f. Data belong to the [100] cut sample, and analysis resulted in Euler angle coordinates of $\varphi = -212.5°$ (modulo 90° for every other data set), and $\theta = 89.7°$ ($\psi = 0$), i.e., this sample was miscut by ~0.3° (φ was chosen as the angle between *x* and the



Figures 9.10 a–f Experimental and best-fit calculations of R_{pp}, R_{ps}, R_{sp} for *a*-plane TiO_2 ($\epsilon_x = \epsilon_y = \epsilon_\perp, \epsilon_z = \epsilon_\parallel$; $\varphi = -212.5°$, $\theta = 89.7°$, $\psi = 0$; $\hbar\omega = 0.75 - 6.5$ eV).

projection of the *c* axis onto the sample surface). GE data from [100], [110], [111], and [001] cuts were included into the best-fit analysis. A surface roughness layer for each sample was included within the model calculation. The optical properties of the overlayer were approximated by linear fractions of $(1 - f_{\text{void}})(2/3\epsilon_\perp + 1/3\epsilon_\parallel)$ and f_{void}. A point-by-point fit was then performed to obtain the ordinary and extraordinary index values. Figure 9.11 contains the resulting *n* and *k* spectra (symbols). A parametric

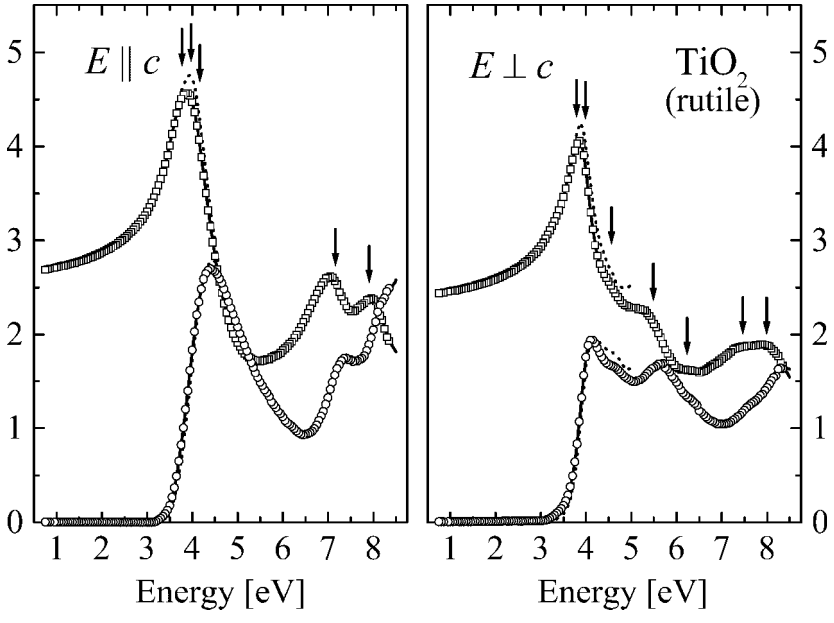


Figure 9.11 Ordinary ($N_o = n_o + ik_o = \sqrt{\epsilon_\perp}$) and extraordinary ($N_e = n_e + ik_e = \sqrt{\epsilon_\parallel}$) refractive ($n_{o,e}$: squares) and extinction ($k_{o,e}$: circles) indices of TiO_2. Symbols refer to data obtained from simultaneous point-by-point fit analysis of GE reflection parameters from [100], [110], [111], and [001] cut bulk crystals ($\hbar\omega = 0.75-8.35$ eV) corrected for surface-overlayer effects (roughness). The solid lines result from lineshape fit of N_o and N_e using sets of Lorentz oscillators. The vertical arrows indicate the center energies of the individual critical-point contributions for $E \parallel c$ (left panel) and $E \perp c$ (right panel).[58] Dotted lines depict data obtained by Jellison using two-modulator GE ($\hbar\omega = 1.44-5$ eV).[9, 21]

dielectric function model was further used to analyze the point-by-point fitted spectra. The solid lines in Fig. 9.11 refer to a Lorentz oscillator fit where the arrows indicate the center energies of the individual harmonic critical-point contributions.[58] The dotted lines depict data from Jellison obtained from analysis of two-modulator GE.[9, 21] Note that the optical constants presented in Ref. 8 were obtained without consideration of surface roughness effects as this was not in focus. The *n* and *k* spectra there vary from sample to sample, and demonstrate well the necessity to consider the overlayer issue for best results. It is stressed, however, that multiple-sample fits as performed here increase only the *consistency* of optical constants spectra. The *accuracy* of the true bulk *n* and *k* values is still not really

secure as the real optical effects of roughness, contaminants, and/or surface oxides are actually unknown. The EMA procedure mimics the overlayer effect but cannot remove it completely, unless the optical constants for the overlayer are known. The only way out is to measure crystals *in situ* at the clean and as-grown surface.

9.6.1.2 α-Al_2O_3 (Sapphire)

Sapphire is an excellent insulator with good thermal conductivity. Due to its wide band gap the material is transparent for photons with energies from the far infrared to the deep-ultraviolet, i.e., from the lattice reststrahlen band (~0.12 eV) to the onset of the electronic band-to-band transitions (~9 eV).[20, 57, 67] Sapphire (α-Al_2O_3) consists of a hexagonal closely packed (*hcp*) lattice. Because of its rombohedral structure, sapphire exhibits uniaxial anisotropy. High-quality synthetic sapphire is widely used as substrate material for solid-state device applications. Because of its similarities in crystal structure and lattice constants, compound materials such as $Al_xGa_{1-x}N$ or $In_xGa_{1-x}N$ can be grown on sapphire using metal-organic vapor phase epitaxy. Successful growth of heterostructures, multilayers, and superlattices from group III-nitride materials on sapphire promises large-scale fabrication of GaN-based blue-light emitting diodes and laser devices.[68] Double-side polished sapphire substrates with nominal thickness of ~300 μm are chosen here to demonstrate treatment of uniaxial materials when multiple backside reflected beams interfere incoherently. For the calculation of transmission and reflection quantities during data analysis the general transfer matrix $\mathbf{T} = \mathbf{T}_+$ follows again from the product of $\mathbf{L}_a$ and $\mathbf{L}_f$. Note that we also need $\mathbf{T}_-$, and $\mathbf{T}_+$ is simply the inversion of $\mathbf{T}_-$. (We do not consider films or roughness at the front or backside here.) $\mathbf{L}_f$ is calculated using the eigenmodes $\Xi_i(q_i)$ as done within the last paragraph.

Figure 9.12 demonstrates the effect of incoherent backside reflected beams on single-wavelength reflection-type ellipsometry data taken from *r*-plane (1102) sapphire. If we chose $\epsilon_z = \epsilon_{||}$, $\epsilon_x = \epsilon_y = \epsilon_\perp$, the *r*-plane situation is $\theta \sim 60$, $\psi = 0°$, and φ according to the in-plane sample azimuth. In Fig. 9.12 the [0001] direction is rotated by $\varphi = +135°$ (circles) and $\varphi = -135°$ (triangles) around the *z*-axis, and tilted from the sample normal (*z*) by $\theta = 60.5°$. The actual values for θ and φ are best-fit analysis results. The measured ellipsometric parameters Ψ and Δ were transformed from α and β using Eqs. (9.21). The Fourier coefficients were measured at fixed polarizer azimuth $P = 45°$. The compensator function was turned off. ($w = 0$, unregressed data acquisition mode. (See Sections 9.2.1 and 9.5.2.)

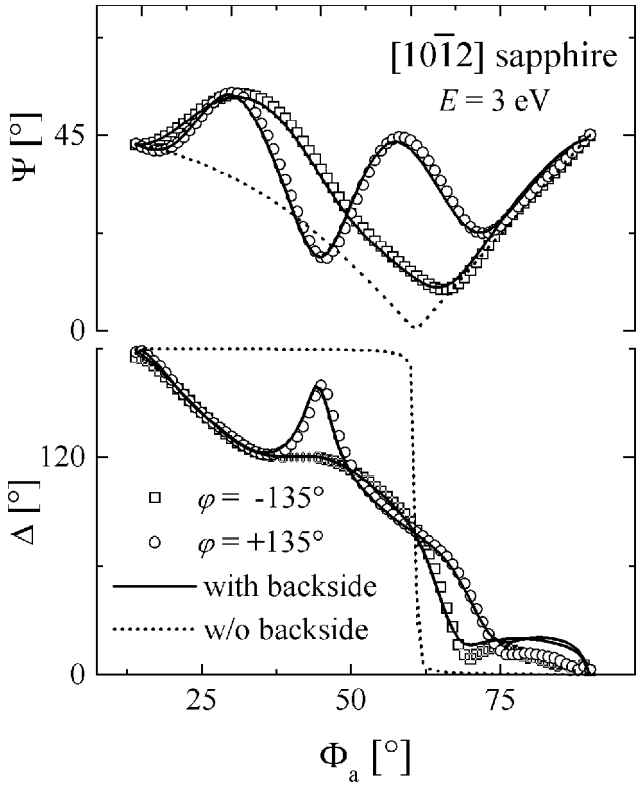


Figure 9.12 Influence of incoherent backside reflected beams on Ψ and Δ in reflection. The thickness of the sample is 279 μm. The experimental (symbols) Ψ and Δ values were measured with $\chi = \tan P = 1$ ($w = 0$; polarizer P set at 45°) at double side polished sapphire for incident photon energy $E = 3$ eV. The crystal is nominally cut with the [10 $\bar{1}$ 2] direction normal to the surface (*r*-plane sapphire; $\in_x = \in_y = \in_\perp$, $\in_z = \in_{||}$, $\psi = 0°$). Data analysis resulted in $\theta = 60.5°$. The in-plane azimuth of the optical axis with respect to x is $\varphi = -135°$ (squares) and $\varphi = +135°$ (circles). The dotted lines refer to treatment of the zero-order reflected beam only (half-infinite substrate approach). The solid lines are the best-fit calculations including up to third-order incoherent backside reflected beams.

The angle-of-incidence scans produce Ψ and Δ values which differ substantially from those observable for optically thick bulk materials (half-infinite substrates). The dotted lines in Fig. 9.12 are simulated Ψ and Δ values for reflection from an ideal half-infinite sapphire medium. Ψ would vanish at the Brewster angle, and Δ would switch by π. This Brewster-angle effect is extremely useful for finding accurate optical

constants values for bulk materials and films with finite thickness, because the actual incidence angle for which the *p*-polarized light is suppressed depends crucially on the materials *n* and *k* values. However, for the case of incoherent backside reflected beams propagating through anisotropic substrates, the Brewster angle is no longer a well-defined quantity as can be clearly seen from the experimental data in Fig. 9.12. Although the *p*-polarized light reflected off the front side is still suppressed at the Brewster angle, there will be in general some *p*-polarized amplitudes from waves entering through the front side of the sample after bouncing off its backside. This effect is also present if the material is isotropic, however, the Ψ and Δ data would be identical for different sample azimuths. Figure 9.12 is an impressive demonstration of how backside reflected beams can actually help to identify orientation, amount, and sign of anisotropy in birefringent media: the calculated zero-order Ψ and Δ values for the two sample azimuths are virtually indistinguishable despite the present anisotropy whereas the real data, affected by incoherent beam propagation, differ substantially for different sample azimuths. The intensities of the backside reflected beams depend on the thickness, the optical constants, and the crystal orientations of the anisotropic slab. But spatial separation of front- and backside reflected beams might cause nonequal detection of multiple beams propagating into the detector. This effect can become more significant for large Φ_a. A quantity that accounts for the Φ_a-dependence of the intensity of the n^{th}-order reflected beam is therefore usually needed within the data regression algorithm. The solid lines in Fig. 9.12 present the best-fit calculations where backside reflected beams up to the 3rd order were considered. Higher-order beams do not contribute any further as the intensity of the back-reflected beams vanishes rapidly due to the lossy reflections at the front and backside of the sample.

For the transmission setup in sapphire it is sufficient to consider the first-order transmitted beam only because higher-order beams are of too little intensity. The symmetry of the transmission GE parameters upon sample rotation is similar to the reflection parameters presented for Rutile within the preceding paragraph. Figures 9.13 a and b show experimental transmission GE and *p*-to-*p* ($|t_{pp}|^2$) and *p*-to-*s* ($|t_{ps}|^2$) polarized transmission intensity (PTI) data taken from *r*-plane sapphire at the same sample azimuths as in Fig. 9.12. Data were acquired at $\varphi = +135°$, and $\Phi_a = +10°$ and $\Phi_a = -10°$. (Note that for a uniaxial material $\Phi_a = -10°$ and $\varphi = +135°$ is identical to $\Phi_a = +10°$ and $\varphi = -135°$.) The PTI data provide information for the average values $\epsilon' = (\epsilon_{||} + 2\epsilon_{\perp})/3$. ($\epsilon'$ follows basically from the difference between 1 and the maxima in $|t_{pp}|^2$, $|t_{ss}|^2$. Note that $|t_{pp}|^2$ and $|t_{ps}|^2$ are always less than unity at their maxima in Fig. 9.13b.) The off-diagonal GE and PTI data are extremely sensitive to $\Delta\epsilon = (\epsilon_{||} - \epsilon_{\perp})/3$, and to the crystal orientation of the *r*-plane substrate.

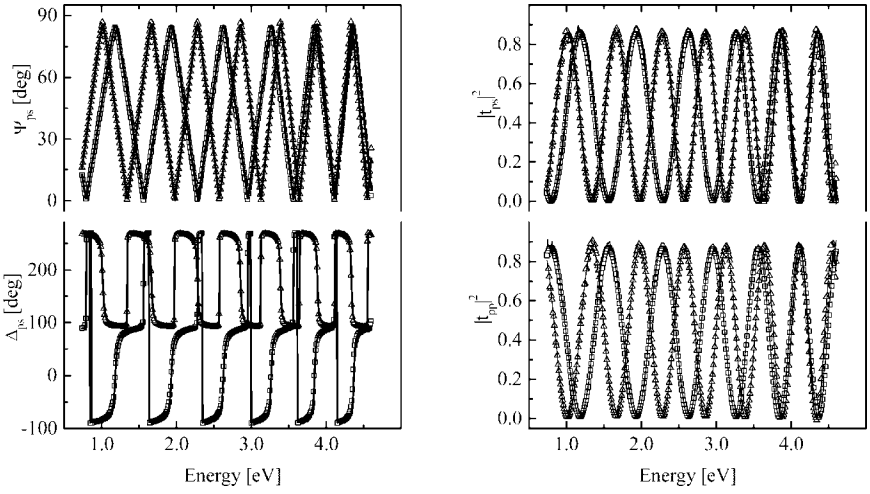


Figure 9.13 Same as Fig. 9.12 for transmission GE and polarized intensity parameters. First-order beam propagation is considered here for data analysis (solid lines are best-fit calculations). Fig. 9.13a: Ψ_{ps} and Δ_{ps}; Fig. 9.13b: $|t_{ps}|^2$ and $|t_{pp}|^2$. (squares: $\Phi_a = +10°$; triangles: $\Phi_a = -10°$; $\varphi = 135°$, $\theta = 60.5°$, $\psi = 0°$).

The GE and PTI data in Figs. 9.13 a and b reveal that the *c*-axis of the sample is oriented parallel to either one of the (*x*, *z*) or (*y*, *z*) planes, but not parallel to either one of the *x*, *y*, or *z*-axes. This is because

(*i*) The off-diagonal elements are zero regardless of $\Delta\in$ if $c \parallel z$, or $c \parallel x$, or $c \parallel y$.

(*ii*) For $c \perp z$ and $\varphi = 45°$, 135°, 275°, or 315°, the off-diagonal elements are symmetrical in Φ_a, except for a sign change in Δ_{ps} and Δ_{sp}. The off-diagonals and on-diagonals are conjugated at the same angle of incidence, e.g., $|t_{ps}|^2 + |t_{pp}|^2 =$ constant. Ψ_{ps} and Ψ_{sp} span 0° to 90° because $|t_{pp}|^2$ and $|t_{ss}|^2$ become zero at their interference minima.

(*iii*) For $c \perp z$ and $\varphi \neq 45°$, 135°, 275°, or 315°, the off-diagonal elements are still symmetrical in Φ_a. The off-diagonals and on-diagonals are also still conjugated at the same angle of incidence, e.g., $|t_{ps}|^2 + |t_{pp}|^2 =$ constant. But Ψ_{ps} and Ψ_{sp} do not extend to 90° because $|t_{pp}|^2$ and $|t_{ss}|^2$ do not vanish anymore at their interference minima.

(*iv*) For $\theta \neq 0°$, and $\theta \neq 90°$ [*c* is not parallel to either *z* or (*x*, *y*)], and $\varphi = 45°$, 135°, 275°, or 315°, the off-diagonal elements are not symmetrical in Φ_a anymore (The spectral positions

of minima/maxima are shifted with respect to Φ_a and $-\Phi_a$). The difference of the period of the interference pattern in Ψ_{ps} or Ψ_{sp} at Φ_a and $-\Phi_a$ can be used to estimate the *c*-axis in-plane orientation within the (*x*, *y*) plane. The amplitudes of the PTI data are still the same at Φ_a and $-\Phi_a$. The off-diagonal and on-diagonal PTI elements are still conjugated at Φ_a, e.g., $|t_{ps}|^2 + |t_{pp}|^2 = \text{constant}$, but $|t_{pp}|^2$ and $|t_{ss}|^2$ do not vanish at their interference minima, and $0° \leq \Psi_{ps}, \Psi_{sp} < 90°$. However, as for (*i*)–(*iii*), $|t_{ps}|^2 = |t_{sp}|^2$.

(*v*) For $\theta \neq 0°$, and $\theta \neq 90°$, and $\varphi \neq 45°$, 135°, 275°, or 315°, amplitude and period of the PTI data are not symmetrical in Φ_a. But we still have $|t_{ps}|^2 = |t_{sp}|^2$, and $|t_{ps}|^2 + |t_{pp}|^2 = const.$, $|t_{sp}|^2 + |t_{ss}|^2 = const.$ and also $|t_{pp}|^2$, $|t_{ss}|^2 > 0$.

In Figs. 9.13 a and b the maxima/minima in both Ψ_{ps} and $|t_{ps}|^2$ are not symmetrical with respect to Φ_a. Because of $0° \leq \Psi_{ps} \leq 90°$ and $max\{|t_{ps}(\Phi_a)|^2\} = max\{|t_{ps}(-\Phi_a)|^2\}$, the sample orientation corresponds to case (*iv*). A Cauchy model is sufficient to parameterize $\epsilon_{||}$ and $\epsilon_{\perp}$ of sapphire within the spectral range shown in Figs. 9.13 a and b. Data from multiple samples cut at different crystallographic orientations (*r*-, *m*-, *a*-, and *c*-plane), and scans from multiple sample azimuths were simultaneously analyzed to find a common set of $\epsilon_{||}$ and $\epsilon_{\perp}$. It is stressed that for this uniaxial situation $\Delta\epsilon$, ϵ', and the thickness *d* are strongly correlated. Even the unregressed reflection ellipsometry data at multiple sample azimuths and multiple angles of incidence can be reproduced with different sets of $\Delta\epsilon$, ϵ', and *d*. Only the PTI data contain independent information for ϵ'. But intensity data have usually larger error bars than the ellipsometry data, and it is good advice to caliper the thickness of the sample to estimate an initial parameter value. Figure 9.14 presents the results for the ordinary and extraordinary refractive indices obtained from analysis of the ellipsometry and transmission intensity data. Data from Refs. [67, 69, 70] are included for comparison.

9.6.2 Anisotropic Films

9.6.2.1 CuPt-Order-Induced Birefringence in III-V Compounds

This paragraph focuses on the application of GE to very small anisotropy with arbitrary orientation in thin films where traditional transmission or reflection techniques fail to measure the spectral dependence of the major dielectric functions. An example of small birefringence is that of the CuPt-order-induced anisotropy in semiconducting III-V

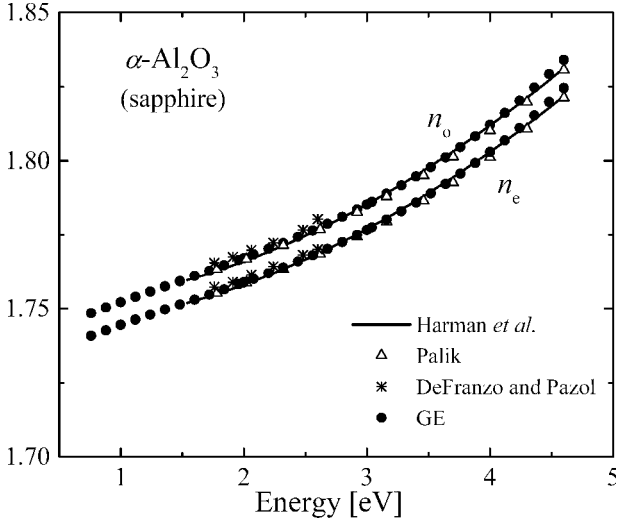


Figure 9.14 Ordinary ($n_o = \sqrt{\epsilon_\perp}$) and extraordinary index of refraction ($n_e = \sqrt{\epsilon_{||}}$) of α-Al_2O_3 obtained from simultaneous best-fit analysis of reflection and transmission-type ellipsometry and intensity data taken from *m*-, *r*-, *a*-, and *c*-plane sapphire substrates.

compounds. Such long-range chemical ordering (LRO) is widely observed in ternary or quaternary III-V compounds grown by vapor-phase epitaxy.[52–54] The spectral form of the birefringence $\Delta\epsilon$ depends to the degree of ordering η within the films. The birefringence is very small, and $||\Delta\epsilon|| \ll ||\epsilon||$ is valid regardless of the incident photon energy. Electronic band structure parameters such as the Brillouin-zone-center transition energies and broadening values, and the valence band splitting parameters are accessible from $\Delta\epsilon$ and ϵ. The GE approach was applied recently to measure the order-induced birefringence in AlGaInP thin films grown on GaAs. The procedures used there are recalled here to demonstrate the treatment of such small anisotropy. Further details can be found in Refs. 17 and 66.

Numerous theoretical and experimental efforts have been undertaken to predict and study the effect of LRO on the physical properties of semiconductor materials.[52,55,71–73] In zinc-blende compounds such as $Al_yGa_{1-y}InP_2$ or $Al_yGa_{1-y}InAs_2$ the spontaneous arrangement of column-III elements in alternating $\{111\}_B$ planes constitutes ordering of the $CuPt_{(B)}$ type. As a consequence the symmetry reduces from F43*m* (zinc-blende structure) to R3*m* (CuPt structure).[53] The stacking direction of the superlattice (SL) planes within the $(A_yB_{1-y})^{III}C^{III}D_2{}^{V}$ compounds is usually

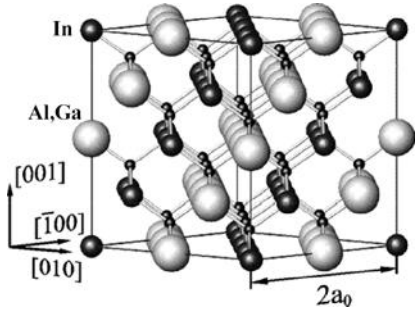


Figure 9.15 Crystal structure of perfectly $CuPt_B$ ordered $Al_yGa_{1-y}InP_2$ for ordering along the [1 $\bar{1}$ 1] direction. See also A. Zunger, "*Spontaneous atomic ordering in semiconductor alloys*", *MRS. Bull.,* **22**, 27−32 (1997).

found according to ordering within either one of the $\{111\}_B$ planes, i.e., parallel to the [$\bar{1}$ 11] or the [1 $\bar{1}$ 1] direction. ($CuPt_B$-type ordering, Fig. 9.15. Two other possible but rarely observed $CuPt_A$ variants are due to LRO within the $\{111\}_A$ planes.) The ordered phase consists of a 1,1-SL of monolayer planes $(A_yB_{1-y})_{x+\eta/2}C_{x-\eta/2}$ and $(A_yB_{1-y})_{x-\eta/2}C_{x+\eta/2}$, where x and y are the composition parameters of the random $(A_yB_{1-y})_xC_{1-x}D_2$ alloy, and the order parameter η is the difference between the compositions x of two subsequent sublattice planes within the CuPt SL. The SL planes are usually not completely ordered, and the parameter η specifies the degree of LRO.†

CuPt LRO reduces the fundamental band gap by ΔE_g. This gap reduction can be as large as 430 meV or 270 meV for $\eta = 1$ in $GaInP_2$ or $AlInP_2$, respectively.[73] CuPt LRO also splits the Brillouin-zone Γ-point valence band states of the disordered compound into states $|\Gamma_{4,5v}>$, $|\Gamma_{6(1)v}>$, and $|\Gamma_{6(2)v}>$, and leads to polarized interband transition matrix elements. Sometimes treated as "chemical" strain, the effects of the CuPt LRO are similar to biaxial strain along the LRO direction. As a consequence $\in$ is anisotropic. The space group R3m has trigonal symmetry, and $\in$ is uniaxial. The CuPt-LRO birefringence (order-birefringence) is very small. However, the spectral dependencies of $\Delta\in$ and $\in$ reveal the transitions energies from the splitted valence band states to the conduction band states (Γ_{6c}) at the center of the Brillouin zone ($E_0^{(1)}$: $\Gamma_{4,5v} - \Gamma_{6c}$, $E_0^{(2)}$: $\Gamma_{6(1)v} - \Gamma_{6c}$, $E_0^{(3)}$: $\Gamma_{6(2)v} - \Gamma_{6c}$).

†If the random alloy is nonstoichiometric ($x \neq 0.5$), η can only vary between $0 \leq \eta/2 \leq \min\{x, 1-x\}$. Our samples were grown lattice matched to GaAs in which case $x = 0.48$. The lattice matching condition is necessary to avoid strain-induced birefringence. For derivation of the LRO parameter η, however, we assume the ideal composition $x = 0.5$. The $(Al_{0.48}Ga_{0.52})_{0.5}In_{0.5}P$ is then referred to as $Al_{0.48}Ga_{0.52}InP_2$. One may consider the remaining 0.02 Ga as being located at a Group III site randomly.

The states $|\Gamma_{4,5v}>$, $|\Gamma_{6(1)v}>$, and $|\Gamma_{6(2)v}>$ (figuratively "heavy-hole" like, "light-hole" like, and "spin-split" like) are further related to the spin-orbit (Δ_{so}), and crystal-field splitting parameters (Δ_{cf}) within the quasicubic perturbation model. As shown by Wei and coworkers η can then be derived from ΔE_g, Δ_{so} and Δ_{cf}, as these quantities scale with η: $Q(x, \eta) = Q(x, 0) + \eta^2 Q(x, 1)$. (See Ref. 52 and references therein. See also Ref. 17.)

Polarization selection rules for optical transitions cause dichroic material properties as the strength of photonic absorption will depend on the direction of polarization.† For analysis of stress-induced birefringence (piezo-birefringence) in zinc-blende semiconductors Higginbotham, Cardona, and Pollak developed a one-electron interband model for $\in$.[74] This model contains the transition energies, their strength, and broadening values as adjustable parameters. Quantitative analysis of anisotropy measurements then allows for determination of level splitting and transition strength as a function of strain. This piezo-birefringence model is adopted here as the order-birefringence model, and the strain dependence is replaced by the LRO order parameter η.

Let the index j refer to the component of momentum polarized parallel to either one axis of a right-handed Cartesian coordinate system. (The matrix elements are assumed to be independent of $\boldsymbol{k}$). The contribution of the direct 3D M_0-type (E_0) transition i to the major complex dielectric function $\in(\omega)_j$ may be written as (j = "x", "y", "z")[74]

$$[\in(\omega)]_j \propto \in_{\infty j} + \sum_i A_j^{(i)} f^{(i)}(\chi^{(i)}), \tag{9.116}$$

with $\chi_0^{(i)} = (E + i\Gamma_0{}^{(i)})/E_0{}^{(i)}$ and

$$f_0^{(i)}(\chi_0^{(i)}) = [E_0^{(i)}]^{-1.5}(\chi_0^{(i)})^{-2}[2 - \sqrt{1 + \chi_0^{(i)}} - \sqrt{1 - \chi_0^{(i)}}]. \tag{9.117}$$

The quantities $A_j{}^{(i)}$, $\Gamma_0{}^{(i)}$, and $E_0{}^{(i)}$ are treated as, respectively, strength, broadening, and energy of transition i due to incident electromagnetic radiation polarized along j. The transition $E_0{}^{(1)} = \Gamma_{4,5v} - \Gamma_{6c}$ is forbidden for polarization parallel to the ordering direction ($\in_{||}$).[85] The transitions $\hat{E}_0{}^{(2)} = \Gamma_{6(2)v} - \Gamma_{6c}$, and $E_0{}^{(3)} = \Gamma_{6(2)v} - \Gamma_{6c}$ may be polarized parallel as well as perpendicular to the ordering direction ($\in_{\perp}$).‡

A special solution for the GE parameters of CuPt-ordered films is given in Section 9.4.5. It was assumed that $\Delta\in \ll \in$, and that the sample in-plane azimuth is set to $\varphi = 45°$. The latter was chosen because the 45°-type position is when the p-to-s coefficients can be measured with

†See, e.g., G.E. Pikus and G. E. Bir, "*Symmetry and strain-induced effects in Semiconductors*" (Wiley, New York, 1974).

‡The ordering direction is here [1 $\bar{1}$ 1].

best accuracy. The advantage of the approximate solutions given in Section 9.4.5 is that the link between $\Delta\epsilon$, ϵ and the GE parameters is most obvious. In particular, we learned that the off-diagonal ratios evolve linearly in $\Delta\epsilon$ whereas the on-diagonal ratios are proportional to ϵ'. The accuracy in $\Delta\epsilon$ is therefore related to the error bars on the off-diagonal GE parameters, and the on-diagonal ratios are important as they will provide the isotropic "background" dielectric function. The approach to obtaining the solution for this specific sample orientation and anisotropy situation may inspire similar derivations for samples with different types of birefringence. It goes without saying that the full approach for biaxial films given in Section 9.4.1 can be used just as well for the CuPt-ordered films. It is worth noting that the approximate solutions and the exact solution are virtually the same for values of $\Delta\epsilon$ and ϵ obtained from CuPt-ordered samples.

A thin film of $Al_{0.48}Ga_{0.52}InP$ grown by MOVPE is studied here. The degree of LRO ordering in this sample is $\eta = 0.47$, and the band gap is red-shifted by 84 meV ($E_g \sim 2.195$ eV for disordered $Al_{0.48}Ga_{0.52}InP$[88]). The zone-center transition energies are located at $E_0^{(1)} = 2.116$ eV, $E_0^{(2)} = 2.141$ eV, and $E_0^{(3)} = 2.22$ eV. The film thickness is 1210 nm. For the optical experiments one piece of the sample was placed upside down on a thin glass slide (BK7) to remove the GaAs substrate material chemically.[86,87] This was necessary for the transmission measurements because the band gap of GaAs (1.42 eV) is smaller than that of the $Al_{0.48}Ga_{0.52}InP$ sample. The ellipsometric spectra were measured at multiple angles of incidence ($\Phi_a = 0°$ in transmission; $\Phi_a = 15°$, and 65° in reflection arrangement), and the samples were aligned to the laboratory coordinate system with [100] || x, [0 $\bar{1}$ 0] || y, and [0 0 $\bar{1}$] || z. Note that y and z invert accordingly for the sample upside down on glass. (See Fig. 9.15). Figure 9.16 shows Ψ_{sp} data measured in transmission (O; $\Phi_a = 0°$) and reflection (□; $\Phi_a = 15°$) from the sample on BK7. The insets enlarge the zone-center transition range. Figure 9.17 presents real and imaginary parts of R_{ps} and R_{sp} measured in reflection at $\Phi_a = 15°$ from the as-grown film on GaAs. The data sets reveal the spectral features of the near-band-gap CuPt order-birefringence:

- Below the first zone-center transition energy $E_0^{(1)}$, the birefringence is real-valued and undergoes a sign change indicated by C. At this spectral position the epilayer is isotropic, and both $\mathbf{t}_{sp}/\mathbf{t}_{ss}$ and $\mathbf{r}_{sp}/\mathbf{r}_{ss}$ vanish. Below C, the birefringence is positive, and pronounced interference oscillations occur in Ψ_{sp}. Above C, the order-birefringence is negative with strong dispersion.
- The absorption of incident electric field polarization along the (1 $\bar{1}$ 1) sublattice planes entails the strong rise in $\Psi_{sp} = tan^{-1}|\mathbf{t}_{sp}/\mathbf{t}_{ss}|$ at $E_0^{(1)}$. There, the reflection-type data are still dominated by interference effects.

- At $E_0^{(2)}$ the incident radiation polarized along the ordering direction excites electrons from the $\Gamma_{6(1)v}$ valence-band state to the Γ_{6c} conduction band state damping the *p-s* mode-coupled wave amplitudes for photon energies above $E_0^{(2)}$.
- At $E_0^{(3)}$ the $\Gamma_{6(2)v} - \Gamma_{6c}$ transition is activated accompanied again by a change of dispersion of the birefringence. Above $E_0^{(3)}$ the epilayer becomes opaque. The transmission-type data pass into noise as the transmitted electric field intensities vanish. The exponential damping of the wave amplitudes reflected from the backside of the epilayer causes the decay of the interference oscillations within the reflection type data. Note that the $\Gamma_{6(2)v} - \Gamma_{6c}$ transition can be identified by the pronounced shoulders at $E_0^{(3)}$ in the reflection data.

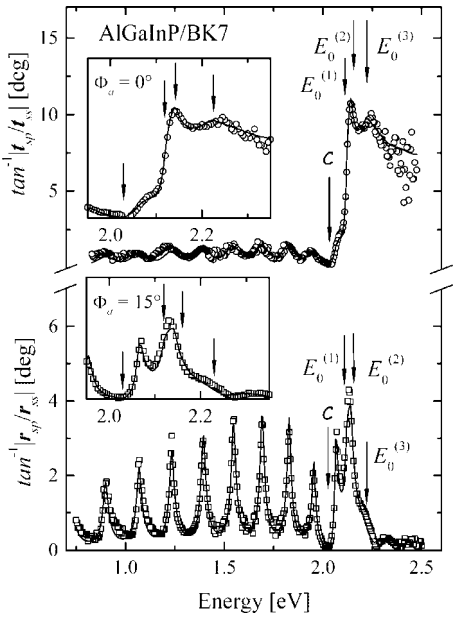


Figure 9.16 Transmission (circles; $\Phi_a = 0°$) and reflection (squares; $\Phi_a = 15°$) GE data Ψ_{sp} (—best fit) from CuPt-ordered $Al_{0.48}Ga_{0.52}InP_2$ on BK7 (sample 3 in Ref. 17). The insets enlarge the zone-center transition region. The data sets are fingerprints of the near-band-gap CuPt order-birefringence. Absorption of incident electric field polarizations along the sublattice planes entails the strong rise in $tan^{-1}|\,t_{sp}/t_{ss}\,|$. At $E_0^{(2)}$, the incident radiation polarized along the ordering direction excites electrons from the $\Gamma_{6(1)v}$ valence-band state to the Γ_{6c} conduction band state bleaching the *p-s* mode coupling effects above $E_0^{(2)}$. The $\Gamma_{6(2)v} - \Gamma_{6c}$ transition is activated at $E_0^{(3)}$. Above $E_0^{(3)}$, the film becomes opaque. See also Fig. 9.17.

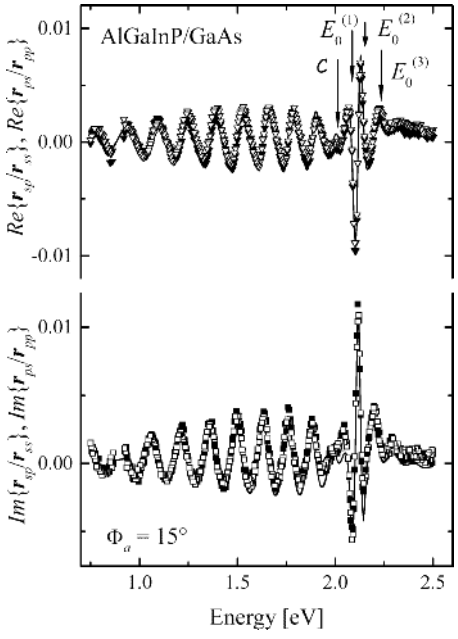


Figure 9.17 Real and imaginary parts of R_{ps} (solid symbols) and R_{sp} (open symbols) from CuPt-ordered $Al_{0.48}Ga_{0.52}InP_2$ on GaAs (sample 3 in Ref. 17;——best fit).

Figure 9.17 depicts $\in_{||}$ ($||$ $[1\bar{1}1]$) and $\in_{\perp}$ ($\perp$ $[1\bar{1}1]$) in the near-band-gap spectral range. The piezo-birefringence model above is used for parameterization of $\in_{||}$ and $\in_{\perp}$. The reflection and transmission GE data from the film on BK7 and the as-grown film on GaAs were analyzed simultaneously. R_{pp} (not shown here) determines $\in$. Note that due to the symmetry of the sample orientation T_{pp} is unity at $\varphi = 45°$ ($\Phi_a = 0°$: $\alpha_{pp} = 0$ and $\boldsymbol{t}_p = \boldsymbol{t}_s$, see Section 9.4.5). An unbroadened Lorentz oscillator was employed to model the isotropic contribution to $\in$ from higher-energy critical points. (For details see Ref. 17.) The locations of the zone-center transitions are indicated by arrows in Figs 9.16 through 9.18. The crossover between $\in_{||}$ and $\in_{\perp}$ is indicated by *C*.

The data in Figs. 9.16 refer to very small off-diagonal coefficients. Here, the off-diagonal transmission and reflection *intensity* coefficients are within the order of ~1% of the on-diagonal coefficients. For convenient and nondestructive analysis it is desirable to circumvent the GaAs substrate removal. This restricts measurement to reflection GE. Figure 9.17 shows the off-diagonal ratios obtained from the as-grown sample. Due to the closer index matching between AlGaInP and GaAs compared to

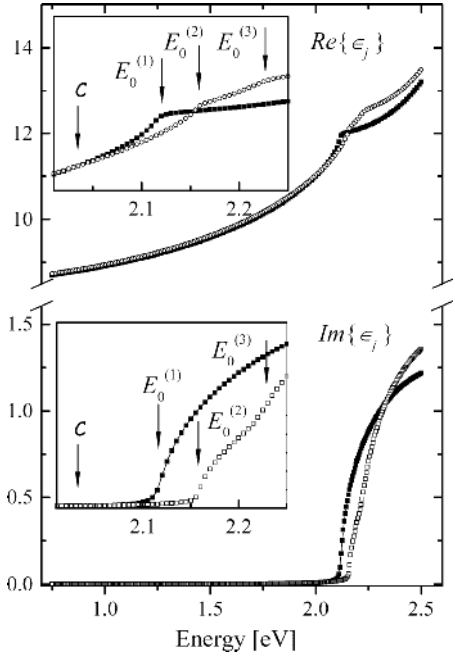


Figure 9.18 $\epsilon_\perp$ (solid symbols), and $\epsilon_{||}$ (open symbols) of CuPt-ordered $Al_{0.48}Ga_{0.52}InP_2$ (sample 3 in Ref. 17). The insets enlarge the zone-center transition range. Below *C*, the order-birefringence is positive and real-valued, and changes sign at the isotropy point *C*. Above the first zone-center transition $\Gamma_{4,5v} - \Gamma_{6c}$ ($E_0^{(1)}$), the onset of absorption for polarizations perpendicular[1 $\bar{1}$ 1], and activation of the $\Gamma_{6(1)v} - \Gamma_{6c}$, and $\Gamma_{6(2)v} = \Gamma_{6c}$ transitions by polarizations parallel [1 $\bar{1}$ 1] cause strong dispersion in $\Delta\epsilon = 1/3\epsilon_{||}$ (|| [1 $\bar{1}$ 1]) $- 1/3\epsilon_\perp$ ($\perp$ [1 $\bar{1}$ 1]). The order-birefringence allows simultaneous identification of the zone-center transitions of the CuPt-ordered alloy.

BK7 the *p-s* cross-coupling effects are much smaller than for the film on the glass substrate (Note the scales to the left in Fig. 9.17.) However, the off-diagonals provide $\Delta\epsilon$ and the optical fingerprints of the ordered state just as well as the data sets in Fig. 9.16.

To obtain accurate measurement results for such small off-diagonal ratios perfect alignment of all optical components within the ellipsometer is a prerequisite. A valuable check for the "goodness" of the calibration is to measure the Jones transmission elements in straight-through-air arrangement. In this configuration T_{ps} and T_{sp} should be zero. Because there is no ideal surface without anisotropy per se, measuring "air" is the best available choice. Even a bare silicon waver may reveal subtle anisotropy at this

level. A slight misalignment of polarizer, compensator, or analyzer azimuth will virtually rotate the sample Jones matrix regardless to the transmission of the reflection setup. In the straight-through-air arrangement T_{ps} and T_{sp} will reveal non-zero values. Such misalignments will cause constant offsets within the real and imaginary parts of the *p*-to-*s* and *s*-to-*p* ratios determined later from an anisotropic surface. Note that the real and imaginary parts of R_{ps} and R_{sp} shown in Fig. 9.17 match each other naturally because R_{ps} and R_{sp} are almost symmetric at this uniaxial high-symmetry sample situation. It was often encountered during measurements that the loss of calibration constants resulted in offsets between $Re\{R_{ps}\}$ and $Re\{R_{sp}\}$, and between $Im\{R_{ps}\}$and $Im\{R_{sp}\}$. Straight-through-air measurements taken immediately afterwards then revealed misalignment of polarizer calibration constants by tenths of a degree.

9.6.2.2 Chiral Liquid Crystal Cells

Chiral liquid crystals are fundamental for flat panel display technologies. Because of the variety of optical phenomena such as polarization, selective reflection, and huge optical activity, liquid crystals are still under extensive study today. (See, e,g., S.Chandrasekhar, *Chiral Liquid Crystals*, Cambridge University Press.) A special property of CLCs is that they provide Bragg-reflectivity conditions for left- $(-)$ or right- $(+)$ circularly polarized light within spectral regions of the so-called selective reflection. Depending on the handedness of the CLC the wave with the circular polarization that matches the sense of the helix is transmitted whereas the other polarization is totally reflected. At normal incidence this band of high reflectivity for either left or right handed polarization waves is centered at $\lambda_0 = P\sqrt{\epsilon}$, and the width of this band is approximately $\Delta\lambda_0 \sim P\Delta\epsilon$.[†] CLCs with different pitches can be used to build devices that will pass or reflect light of different wavelengths (color). Obviously, pitch and optical constants are crucial parameters for device designs. Traditional measurements of CLC refractive indices require techniques that align the rod-like molecules parallel within a volume large enough to study the birefringent materials like a homogeneous biaxial crystal. Such techniques employ high magnetic or electric fields to break the intrinsic twisted nematic phase, and will likely affect the optical properties of the molecules. But even if the optical constants are known, the handedness of the helix formed by the liquid crystals along the *z*-axis in their nematic state is difficult to obtain without the phase information of the light emerging from the sample. Traditional techniques that determine

[†]Here, P is the length of the pitch. See also Section 9.5.3.

sign and amount of P require special sample geometries such as wedge-shaped cells.[75,76,78] The GE approach is the only technique known so far which allows simultaneous measurement of refractive indices and pitch P from CLC materials within the twisted nematic phase. Applications of transmission GE to CLCs have been reported recently.[15] For the latter, the exact solution presented in Section 9.4.3 was used to analyze the GE data. This paragraph summarizes the GE procedures and results for the CLCs obtained in Ref. 15, where mixtures of 4-cyano-4'-penthylbiphenyl (5CB) and 4-cyano-4'(2methyl)-butylbiphenyl (CB15) were discussed as examples. Because the optical properties of CLCs depend crucially on the sample temperature, all studies were performed as a function of temperature.

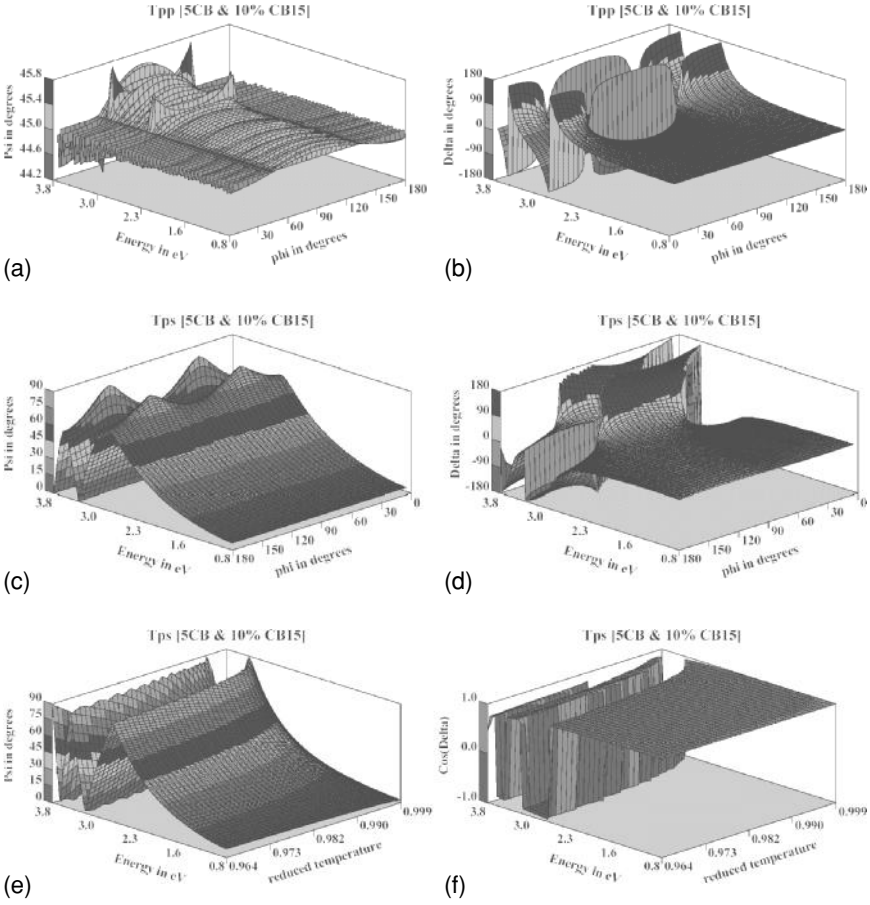
Mixtures of 5CB and CB15 form short-range ordered helices, whereas the molecules in pure 5CB arrange parallel throughout. When such mixtures are placed into a cell where the inner cell interfaces are prepared with a structural preference direction, the molecule directions **n** form a cooperative helix with pitch P through the whole cell (See Fig. 9.6. **n(r)** is a unit vector field, or the so called director field, which is oriented parallel to the average direction of the long molecular axis at point **r**. For pure 5CB $P = \infty$.) The pitch P depends on the CB15 concentration. The CLCs were placed between parallel glass slides (sandwich cells). The cell-glass interfaces were coated with polyimid, and rubbed parallel in order to force the same directional preference of the adjacent liquid crystal layers. This provided a well-defined orientation of the local director field at the incident and exit interfaces with respect to the cavity boundaries. Because of the parallel alignment of the preference directions of both glass-liquid crystal interfaces, integer numbers d/P of helical turns adopted by the molecules within the cell are expected. Mixtures of 5CB and CB15 produce CLCs with **n**(**r**) oriented parallel to the (x, y) plane throughout the cell.

The CLCs are uniaxial, i.e., the average dielectric function tensor of the rod-like molecules within any plane at $z = const.$ has two major values, $\epsilon_{||}$ and $\epsilon_{\perp}$. $\mathbf{T}_p$ for the CLC cells is calculated according to Eq. (9.78). The Euler angles ψ and θ in Eq. (9.47) are zero. The tensor elements of ϵ follow from Eq. (9.47). The in-plane orientation φ of the director **n** at the opposite cell interface follows from P and d. Besides $\epsilon_{||}$ and $\epsilon_{\perp}$, the unknown cell parameters are the pitch P, the cell thickness d, and the in-plane azimuth φ of the director $\mathbf{n}(z = 0)$ at the glass-cell interface on the incident side of the sample. Similar to the case of the transparent bulk material, the cell thickness correlates with the pitch and the CLC optical constants. (See paragraph on sapphire.) The thickness of the unfilled cell can be obtained from standard reflectometry experiments prior to placing the CLC within the sandwich cell. Fused glass (BK7) is used as incident and exit medium, and BK7 optical constants are used to calculate $\mathbf{L}_f$ and $\mathbf{L}_a$. Incoherent backside reflected beam propagation within the glass slides

(~1 mm thick) is insignificant, and ignored here. Note that at normal incidence the glass-air interfaces do not affect the polarization state of the transmitted beam, and consideration of the glass-air interfaces during sample analysis is unnecessary. The glass-air interface effects need to be modeled for oblique angle-of-incidence data only. As will be shown below the effective anisotropy $\Delta\epsilon$ depends crucially on the sample temperature. Accordingly the GE parameters are functions of the reduced sample temperature t_r. The reduced temperature $t_r = T[K]/T_c[K]$ is defined by the so-called clearing temperature T_c at which the CLC turns into an optically isotropic solution. At T_c the average kinetic energy exceeds the total energy associated with the parallel alignment of the molecules, which then start to become randomly oriented within the cell. The order-disorder phase transition causes isotropic optical properties for temperatures $T > T_c$.

To begin with here, Figs. 9.19 a throughf show simulations of transmission GE parameters at $\Phi_a = 0°$ for a uniaxial CLC cell with no tilt of the molecules towards the z-axis ($d = 5.8$ µm; $P = 1.45$ µm; $t_r = 0.985$). These simulations may act as maps in which the reader can find and verify the experimental results discussed in this paragraph. The optical constants refer to the mixture 5CB and 10% CB15. Figures 9.19 a through d show 3D-plots of Ψ_{pp}, Δ_{pp} and Ψ_{ps}, Δ_{ps} versus photon energy and φ. Small oscillations in Ψ_{pp}, Δ_{pp} are due to anisotropic interference effects between forward and backward traveling waves inside the CLC cell. At high-symmetry orientations, i.e., at $\varphi = 45°$ or $\varphi = 135°$, T_{pp} is unity, and Ψ_{ps} (not shown here) presents the amount of rotation of the exit polarization with respect to the incident polarization, and exceeds multiples of 90°. Note that $\Psi_{ps} = \Psi_{sp}$, $\Delta_{ps} = \pi - \Delta_{sp}$, and $\Delta_{ps}(+P) = \Delta_{sp}(-P)$. The latter rule allows identification of the handedness of P. Note further that $|t_{ps}|^2$ and $|t_{sp}|^2$ do not vanish at $\varphi = 0°$, 90°, 180° or 270°. This behavior is very different from that of a homogeneous uniaxial material. The amount of rotation of linearly polarized light upon transmission through the CLC cell, i.e., the optical rotary power, can be obtained modulo 2π from the GE parameters.[77] If this rotation exceeds $k\pi/2$ ($k = 1, 2, \ldots$) the phase Δ_{ps} switches by $\pm\pi$, and Ψ_{ps} is bound within $0° \ldots 90°$. The optical activity in Fig. 9.19c increases with photon energy because $\Delta\epsilon$ increases for shorter wavelengths (see below). The spikes in Ψ_{pp} ($\varphi = 45°$ and $\varphi = 135°$) are at wavelengths where t_{pp} and t_{ss} vanish. These wavelengths are identical to the spectral position at which the optical rotary power exceeds the next multiple of 90°. Figures 9.19 e and f show Ψ_{sp}, Δ_{ps} for $\varphi = 45°$ versus photon energy and t_r. Ψ_{sp} is now a direct measure for the optical activity. Because $\Delta\epsilon$ decreases for $t_r \to 1$ ($\Delta\epsilon = 0$ at $t_r = 1$) the optical rotary power decreases, and disappears at $t_r = 1$.

Figures 9.19 a through f help to set an optimum experimental sample orientation. At 45°-type positions of φ the on-diagonal transmission ratio



Figures 9.19 Simulated T_{pp} and T_{ps} for the twisted nematic mixture 5CB & 10% CB15 versus incident photon energy and sample azimuth φ (Figs. 9.19 a–d: 0.75 eV $\leq E \leq$ 3.75 eV, $0° \leq \varphi \leq 180°$, $t_r = T[K]/T_c[K] = 0.985$, after Ref. 15). The rod-like molecules are parallel to the (x, y) plane. ($\psi = \theta = 0°$, $P = +1.48$ µm, $d = 5.8$ µm, $T_c = 301.4$ K). Figs. 9.19 e–f show simulated T_{ps} for the same cell versus incident photon energy and reduced sample temperature t_r ($0.964° \leq t_r \leq 0.999$, $\varphi = 45°$).

is unity. The off-diagonals are symmetric, and provide all necessary information. (The latter does not hold for arbitrary ratios P/d.) In fact, a single sample orientation is sufficient for data acquisition. No further information is available at different sample azimuths. Rotation of the sample just rotates the Jones transmission matrix. Figures 9.20 a through f show experimental (symbols) and best-fit GE transmission data (T_{ps} and T_{sp}) for

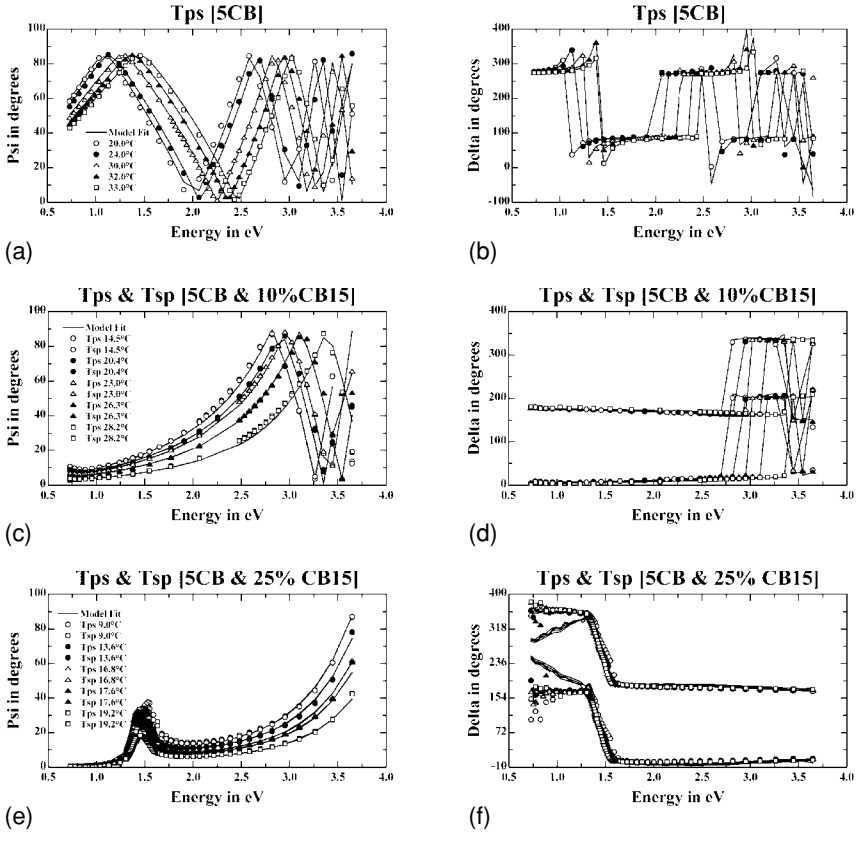


Figure 9.20 Experimental and generated transmission GE data (T_{ps} only) for three CLC cells with different mixtures of 5CB and CB15 as a function of the reduced sample temperature and photon energy. The generated data (solid lines) correspond to the best fit for $\in_{\parallel}$ and $\in_{\perp}$, the director orientation at the incident glass-CLC interface φ, the cell thickness d, and the pitch P. (Figs. 9.20a, b: pure 5CB, $P = \infty$, $d = 3.29$ µm, $T_c = 308.65$ K; Figs. 9.20c, d: 5CB & 10% CB15, $P = +1.45$ µm, $d = 5.80$ µm, $T_c = 301.45$ K; Figs. 9.20e, f: 5CB & 25% CB15, $P = +0.53$ µm, $d = 3.19$ µm, $T_c = 293.15$ K) Note how the length of the pitch P affects the optical behavior of the cells, and the "spectral window" is virtually red-shifted from Figs. 9.20a, b to Figs. 9.20e, f which reveal the spectral region of selective reflection (~1.3−1.7 eV).

three mixtures of 5CB and CB15 as a function of photon energy and sample temperature ($d = 3.29$ µm, $P = +\infty$, $T_c = 308.6$ K, $v = 0.145$; for more details see Ref. 15). Figures 9.20 a and b contain data for pure 5CB. This material does not form cooperative helices, and behaves like a homogeneous uniaxial crystal instead. However, the orientation disorder of the

molecules increases as the sample temperature goes up, and the solution becomes isotropic at T_c. The shift of the off-diagonal parameters Ψ_{ps} and Δ_{ps} with temperature is clearly seen within the data. A Cauchy model was used to parameterize $\epsilon_{||}$ and $\epsilon_{\perp}$[79]

$$\epsilon(\lambda[\text{nm}]) = \left(c_0 + \frac{c_1}{\lambda^2}\right)^2, \tag{9.118}$$

$$\Delta\epsilon(\lambda[\text{nm}], T[K] \leq T_c[K]) = 3\left(e_0 + \frac{e_1}{\lambda^2} + \frac{e_2}{\lambda^4}\right)^2\left\{1 - \frac{T}{T_c}\right\}^{\nu}, \tag{9.119}$$

where $\epsilon_{||} = \epsilon' + 1/3\,\Delta\epsilon$, and $\epsilon_{\perp} = \epsilon' - 2/3\,\Delta\epsilon$. [Eq. (31) in Ref. 15 misses the factor 3 in Eq. (9.119) above.] The wavelength λ is given in units of nm, and the sample temperature T as well as the clearing temperature T_c are given in units of Kelvin. The parameters c_0 and c_1 describe the spectral dependence of the disordered and therefore isotropic phase of the liquid crystals. The temperature-dependent term is related to the degree of order in the liquid crystalline phase, and e_0, e_1, and e_2 describe $\Delta\epsilon$. Reflection data from isotropic 5CB were used to fit c_0 and c_1. The in-plane orientation φ for the pure 5CB cell was ~45°. Note the phase switch in Δ_{ps} at photon energies where the optical rotary power exceeds the next multiple of 90°, and Ψ_{ps} is mapped back into the interval 0° to 90°. These switching wavelengths, especially, shift as the sample temperature changes.

Figures 9.20 c and d present GE data (T_{ps} and T_{sp}) for the mixture 5CB and 10% CB15 in the twisted nematic phase (d = 5.8 μm, P = +1.45 μm, T_c = 301.4 K, ν = 0.126; see also Ref.[15]). The in-plane azimuth φ was set at ~135°, and T_{pp} was unity throughout. The region of selective reflection for this sample is centered $\lambda_0 \sim 2.25$ μm ($\sqrt{\epsilon} \sim 1.53$) and is not seen within the spectral range investigated. Only the data at the long wavelength end in Fig. 9.20c hint at this reflection band. (Note how Ψ_{ps} and Ψ_{sp} bend upwards when the sample is at lower temperatures because the selective reflection bandwidth increases due to the increase in $\Delta\epsilon$.) Figures 9.20e and f show experimental (symbols) and best-fit GE data for 5CB and 25% CB15 (d = 3.19 μm, P = +0.53 μm, T_c = 293.2 K, ν = 0.156;[15]). The center of selective reflection is at $\lambda_0 \sim 825$ nm. This band can now be clearly seen within the data. It is also obvious that the width of the reflection band increases for lower sample temperatures. For incident wavelengths $\lambda > \lambda_0$ (photon energies below ~1.3 eV in Fig. 9.20e) the CLC cell does not support guided circularly polarized waves anymore, and the sample acts as an optical rotator with no in-plane reference direction. The off-diagonal transmission elements t_{ps} and t_{sp} become physically meaningless, and vanish rapidly regardless of the sample azimuth φ. The

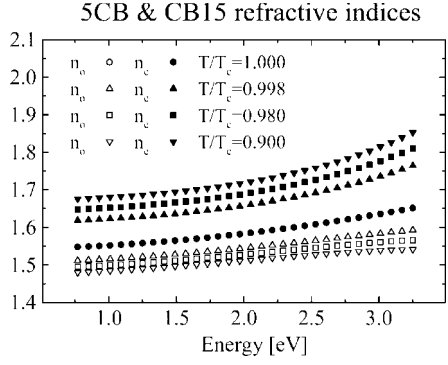


Figure 9.21 Ordinary and extraordinary refractive indices from mixtures of $(5CB)_{1-x}(CB15)_x$ as a function of photon energy and the reduced sample temperature ($x \leq 25\%$; Ref. 15).

phases Δ_{sp} and Δ_{ps} are therefore undetectable, and data in Fig. 9.20f pass into noise.

Figure 9.21 summarizes the index of refraction data obtained from analysis of the three samples described above. Ordinary and extraordinary spectra are shown at different reduced sample temperatures. The best-fit regression calculations for d, P, φ, and $\Delta\in$ were performed for each sample separately. The Cauchy parameters for $\in$' were obtained from analysis of reflectivity data taken at isotropic 5CB. Because $\in$' was known, d, P, φ, and $\Delta\in$ have no correlation. If $\in$' is unknown, the cell thickness d should be measured by standard techniques to decorrelate the remaining parameters. A "multiple-temperature" data set was fit to the calculated data, and the sample temperature T was the input parameter for Eq. (9.119). A detailed discussion of the results obtained on the CLC cells is given in Ref. 15 . The refractive indices shown in Fig. 9.21 at reduced temperatures are valid for all mixtures studied there, whereas the optical response of the three samples is completely different at the same reduced temperature. The latter is a vivid demonstration of the strength of optical anisotropy effects in CLC cells.

9.6.2.3 Kerr Effect in Obliquely Magnetized Co/Au Multilayer Systems

Magneto-optic (MO) materials have gained research and technology interest for mass data storage devices. Standard ellipsometry is widely used to determine thin-film complex index of refraction and MO coupling constants $Q = Q_r + iQ_i$. Usually two different experiments are performed:

an oblique angle-of-incidence measurement to resolve the sample layer structure and layer refractive indices, and a MO measurement of the complex Kerr angle Θ at normal incidence.[80–82] Prerequisites are isotropic sample refractive indices, and special magnetization directions parallel or perpendicular to the sample surface. (For introduction to MO materials and optics in layered MO films, see the textbook by Mansuripur.)

The GE solution for MO films presented in Section 9.4.2 is used here to demonstrate the use of the GE approach for vector magnetometry, i.e., to solve for the magnetization direction and amplitude of magnetized media. Optical and MO constants are taken from realistic multilayer samples. Although not truly realistic for the sample structure assumed here, the optical constants data are sufficient for demonstration. The sample is a silicon substrate [$(n + ik) = (3.8752 + i0.016)$], with 60 nm gold [0.1991 + i3.0829] and an embedded cobalt (Co) MO layer ($d = 20$ nm; Fig. 9.22). The non-zero elements of the Co dielectric function tensor are $\epsilon = \epsilon_{xx} = \epsilon_{yy} = \epsilon_{zz} = (2.6109 + i2.0348)^2$, and $\epsilon_{xy} = -\epsilon_{yx} = -i\epsilon Q = (0.07746 + i0.646024)^2$, i.e., $Q = (0.00079331 - i0.19656)^2$.[84] The MO layer is coated with 2 nm Au.† The Co tensor ϵ refers to the polar Kerr-effect (PK), i.e., the magnetization direction is perpendicular to the sample surface. Oblique magnetization directions are obtained by rotating ϵ using the Euler angles

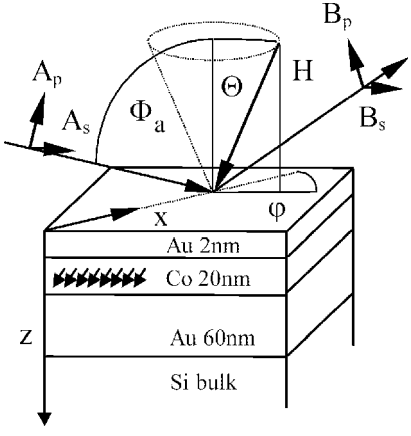


Figure 9.22 Sample structure for the oblique magnetization Kerr-effect calculations of a Co/Au system. θ and φ rotate the magnetization direction **H**. In our notation ($\Theta = 0°$, $\varphi = 0°$), ($\Theta = 90°$, $\varphi = 0°$), and ($\Theta = 90°$, $\varphi = 90°$) refer to the polar (PK), transverse (TK), and longitudinal (LK) Kerr-effect, respectively.

†The optical response of this structure symbolizes that of a 10-period {2 nm Co/2 nm Au} multilayer system, embedded by metal bottom and top layers. The magneto-optical quantities for the Co layer were taken from experimental studies of Cu/Pt multilayers.[84]

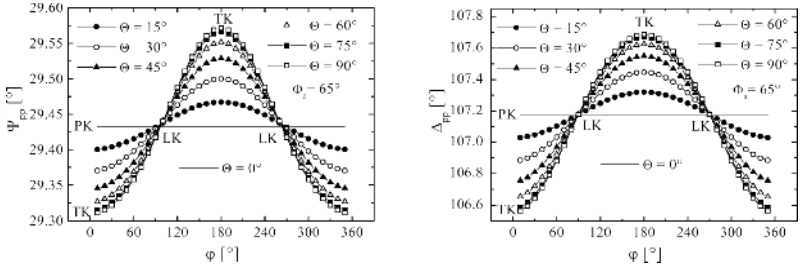


Figure 9.23 MO Kerr-effect for the structure in Fig. 9.22 at $\Phi_a = 65°$ for ψ_{pp} and Δ_{pp} vs. magnetization azimuth φ and inclination $\theta(\hbar\omega = 1.97$ eV, $\lambda = 630$ nm).

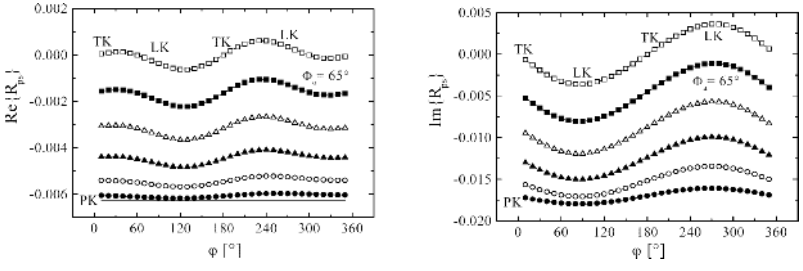


Figure 9.24 Same as Fig. 9.23 for Re{R_{ps}} and Im{R_{ps}}.

θ and φ. At $\theta = 90°$ and $\varphi = 0°$, $\theta = 90°$ and $\varphi = 90°$ the magnetic film has in-plane magnetization referring, by definition, to the transversal (TK), or the longitudinal (LK) Kerr-effect, respectively.

Figures 9.23 and 9.24 show the MO Kerr-effect for R_{pp} (Ψ_{pp}, Δ_{pp}) and R_{ps} ($Re\{R_{ps}\}$, $Im\{R_{ps}\}$) as a function of the sample magnetization azimuth angle φ at various magnetization inclinations θ ($\Phi_a = 65°$; spectra at $\Phi_a = 0°$, and also R_{sp} are discussed in detail in Ref. 34). Note that in our notation of the GE parameters, $Re\{R_{ps}\}$ and $Im\{R_{ps}\}$ are identical to the usual Kerr rotation and ellipticity parameters, Θ_K and η_K, respectively, as defined for example in Ref. 80. (This is one more reason why the present GE notation was chosen in this chapter. See also comments in Section 9.2.1.) The in-plane orientation φ causes modulation in both R_{pp} and R_{ps}, and the modulation amplitude increases for $\theta \to 90°$. The optical response at PK, i.e., $\theta = 0°$, is independent of φ as expected. Note that R_{ps} vanishes at TK. The magnetization-induced anisotropy causes small but yet measurable GE parameters. The amplitude change in Ψ_{pp} vs. φ is detectable when the standard deviation in Ψ is at least $\leq 0.01°$. The off-diagonal elements are small, but within the order of those as measured from the CuPt-ordered birefringent sample shown in Fig. 9.17. This simulation is encouraging as it proves the capability of GE to

determine direction and amplitude of **H** in thin films with an arbitrary amount and orientation of magnetization. It is further observed (see Ref. 34) that GE when performed at oblique rather than at normal incidence is more sensitive to MO effects, and the sample magnetization direction. GE may therefore become the standard tool for analyzing dielectric and MO functions simultaneously. Berger and Pufall reported recently on GE measurement of magnetization direction in polycrystalline Co films demonstrating the feasibility of MO-GE as a magneto-optical vector magnetometer.[14,65]

9.7 Conclusions

GE is a powerful tool for anisotropic optical metrology. Optical constants spectra and orientation "center-of-gravity" relations for anisotropic thin film and bulk materials can be measured accurately. Equations for analysis of GE experiments are now existent for a large variety of anisotropy situations. Physical relevant parameters such as the band-gap reduction and valence-band splitting in chemically ordered III-V compounds, or the degree of ordering, can be obtained through analysis of GE data. A precise index of refraction spectra resulted from GE measurement of chiral liquid crystal cells. As the field of GE continues to grow, more exciting applications will be reported for novel materials and film structures. GE will always remain a challenge because the anisotropy situations are manifold. A general rule for treatment of anisotropic materials does not exist, and the experimenter needs to find best conditions for any new situation. A further task is to find solutions for treatment of general bianisotropic materials, or materials with corrugated surfaces. For example, index gratings may find usage in surface emitting laser diodes, and grating dimensions and index differences are crucial parameters for device performance. Ellipsometric investigations of corrugated surfaces is possible by the GE formalism as demonstrated by Azzam and Bashara through their very first application of GE in 1972.[1] However, the experimental results there remained unexplained, and to analyze GE data convenient analytical grating solutions still have to be incorporated into the formalism. Solid-state sculptured films promise smart materials applications, and investigation of their anisotropy is a future task of the GE formalism.

Acknowledgements

The author wishes to thank many colleagues for their numerous contributions and valuable comments: C.M. Herzinger, B. Johs, C. Cramer, I. Pietzonka, E. Franke, T. Hofmann, A. Kasic, T.E. Tiwald,

H. Schmiedel, and V. Gottschalch. B. Rheinländer and J.A. Woollam have continuously supported research presented in this Chapter. I am indebted to C.M. Herzinger, T.E. Tiwald and G.E. Jellison for critical reading of the manuscript. The author also acknowledges a former fellowship from the Studienstiftung des Deutschen Volkes. This work was in part supported under DFG grants Rh 28/1-1, Rh 28/1-2, Rh 28/3-1, and NSF grant DMI-9901510, and also by the Center for Microelectronic and Optical Materials Research at the University of Nebraska, Lincoln.

9.8 Further Reading and References

9.8.1 General Reading

- A.C. Boccara, C. Pickering and J. Rivory, Eds. *Spectroscopic Ellipsometry* (Elsevier, Amsterdam, 1993).
- R.W. Collins, D.E. Aspnes and E.A. Irene, Eds. *Spectroscopic Ellipsometry 1997* (Elsevier, Lausanne, 1997).
- R.M.A. Azzam and N.M. Bashara, *Ellipsometry and Polarized Light* (North-Holland, Amsterdam, 1977).
- H.A. Macleod, *Thin Film Optical Filters* (Macmillan, New York, 1986).
- M. Mansuripur, *The Physical Principles of Magneto-optical Recording* (University Press, Cambridge, 1995).
- S. Chandrasekhar, *Liquid Crystals* (Cambridge University Press, Cambridge, 1992).
- A. Lakhtakia, *Beltrami Fields in Chiral Media* (World Scientific, Singapore, 1994).
- A. Lakhtakia, *Natural Optical Activity*, SPIE Milestone Series Vol. MS15.
- P.Y. Yu and M. Cardona, *Fundamentals of Semiconductors* (Springer, Berlin, 1995).
- G.E. Pikus and G.L. Bir, *Symmetry and Strain Induced Effects in Semiconductors* (Wiley, New York, 1974).
- S. Adachi, *Physical Properties of III-V Semiconductor Compounds* (Wiley-Interscience, New York, 1992).
- W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipies in C* (Cambridge University Press, Cambridge, 1988).
- W.A. McGahan, "Magneto-optical applications", in *Intermetallic Compounds*, Vol. 2, edited by J.H. Westbrook and R.L. Fleischer, p. 435–451, Ch. 19 (John Wiley&Sons Ltd, London, 1994).
- *International Tables for Crystallography*, edited by T.Hahn (Reidel, Dordrecht, 1983), Vol. A.
- E.D. Palik, *Handbook of Optical Constants of Solids* (Academic, New York, 1985), pp. 798–804.
- H. Goldstein, *Classical Mechanics* (Addison-Wesley, Reading, MA, 1965).
- P. Yeh, *Optical Waves inLlayered Media*, Wiley Series in Pure and Applied Optics (Wiley, New York, 1988).
- J.A. Kong, *Electromagnetic Wave Theory* (John Wiley & Sons., New York, 1990).

- L.D. Landau and E.M. Lifschitz, *Electrodynamics of Continuous Media* (Pergamon, New York, 1960).
- F. Agulló-Lopez, J.M. Cabrera, F. Agulló-Rueda, *Electrooptics* (Academic Press, San Diego, 1994).

9.8.2 Numbered References

1. R.M.A. Azzam and N.M. Bashara, "Generalized Ellipsometry for surfaces with directional preference: application to diffraction gratings", *J. Opt. Soc. Am.,* **62,** 1521–1523 (1972).
2. R.M.A. Azzam and N.M. Bashara, "Polarization transfer function of a biaxial system as a bilinear transformation", *J. Opt. Soc. Am. A,* **62,** 502–510 (1972).
3. P.S. Hauge, "Generalized rotating-compensator ellipsometry", *Surf. Sci.,* **56,** 148–160 (1976).
4. D.J. DeSmet, "Generalized ellipsometry and the 4 × 4 matrix formalism", *Surf. Sci.,* **56,** 293–306 (1976).
5. M. Elshazly-Zaghloul, R.M.A. Azzam and N.M. Bashara, "Explicit solutions for the optical properties of a uniaxial crystal in generalized ellipsometry", *Surf. Sci.,* **96,** 41–53 (1980).
6. H. Wöhler, M. Fritsch, G. Haas and D.A. Mlynski, "Characteristic matrix method for stratified anisotropic media: optical properties of special configurations", *J. Opt. Soc. Am. A,* **8,** 536–540 (1991).
7. K. Eidner, "Light propagation in stratified anisotropic media: orthogonality and symmetry properties of the 4 × 4 matrix formalism", *J. Opt. Soc. Am. A,* **6,** 1657–1660 (1989).
8. M. Schubert, B. Rheinländer, J.A. Woollam, B. Johs and C. Herzinger, "Extension of Rotating Analyzer Ellipsometry to Generalized Ellipsometry: Determination of the dielectric function tensor from uniaxial TiO_2", *J. Opt. Soc. Am. A,* **13,** 875–883 (1996).
9. G.E. Jellison, Jr. And F.A. Modine, "Two-modulator generalized ellipsometry: theory", *Appl. Optics,* **36,** 8184–8189 (1998); "Two-modulator generalized ellipsometry: experiment and calibration", *Appl. Optics,* **36,** 8190–8198 (1998).
10. D.W. Thompson, M.J. DeVries, T.E. Tiwald and J.A. Woollam, "Determination of optical anisotropy in calcite from ultraviolet to mid-infrared by generalized ellipsometry", *Thin Solid Films,* **313–314,** 341–346 (1998).
11. B. Lecourt, D. Blaudez and J.M. Turlet, "Specific approach of generalized ellipsometry for the determination of weak in-plane anisotropy: application to Langmuir-Blodgett ultrathin films", *J. Opt. Soc. Am. A,* **10,** 2769–2782.
12. P.I. Rovira, R.A. Yarussi, R.W. Collins, R. Messier, V.C. Venugopal, A. Lakhtakia, K. Robbi and M.J. Brett, "Transmission ellipsometry of a thin-film helicoidal bianisotropic medium", *Appl. Phys. Lett.,* **71,** 1180–1182 (1997).
13. A. En Naciri, L. Johann and R. Kleim, "Spectroscopic Generalized Ellipsometry based on Fourier analysis", *Appl. Opt.,* (1 Aug.), 4802–4811 (1999).
14. A. Berger and M.R. Pufall, "Generalized magneto-optical ellipsometry", *Appl. Phys. Lett.,* **71,** 965–967 (1997).

15. M. Schubert, B. Rheinländer, C. Cramer, H. Schmiedel, J.A. Woollam, B. Johs and C.M. Herzinger, "Generalized Transmission Ellipsometry for twisted biaxial dielectric media: Application to chiral liquid crystals", *J. Opt. Soc. Am. A,* **13,** 1930–1940 (1996).
16. J.-D. Hecht, A. Eifler, V. Riede, M. Schubert, G. Krauß and V. Krämer, "Birefringence and reflectivity of single-crystal $CdAl_2Se_4$ by generalized ellipsometry", *Phys. Rev. B,* **57,** 7037–7042 (1998).
17. M. Schubert, T. Hofmann, B. Rheinländer, I. Pietzonka, T. Saß, V. Gottschalch, J.A. Woollam, "Near-band-gap CuPt order – induced birefringence in $AL_{0.48}Ga_{0.52}InP_2$", *Phys. Rev. B,* **60,** 16618–16643 (1999).
18. S. Zangooie, R. Jansson and H. Arwin, "Ellipsometric characterization of anisotropic porous silicon Fabry-Pérot filters and investigation of temperature effects on capillary condensation efficiency", *J. Appl. Phys.,* **86,** 850–858 (1999).
19. J.F. Elman, J. Greener, C.M. Herzinger and B. Johs, "Characterization of biaxially-stretched plastic films by generalized ellipsometry", *Thin Solid Films,* **313–314,** 816–820 (1998).
20. C.H. Yan and H. Yao, "Anisotropic optical responses of sapphire (α-Al_2O_3) single crystals", *J. Appl. Phys.,* **85,** 6717–6722 (1999).
21. G.E. Jellison, Jr., F.A. Modine and L.A. Boatner, "The measurement of the optical functions of uniaxial materials using two-modulator generalized ellipsometry: rutile (TiO_2)", *Opt. Lett.,* **22,** 1808–1810 (1997).
22. G.E. Jellison, Jr. J.O. Ramey and L.A. Boatner, "Optical Functions of BiI_3 as measured by generalized ellipsometry", *Phys. Rev. B,* **59,** 9718–9721 (1999).
23. G.E. Jellison, Jr. And L.A. Boatner, "Optical Functions of uniaxial ZnO determined by generalized ellipsometry", *Phys. Rev. B,* **58,** 3586–3589 (1998).
24. *Guide to using WVASE* (J.A. Woollam, Lincoln, 1995)
25. J.L. Tsalamengas, "Interaction of electromagnetic waves with general bianisotropic slabs", *IEEE Transaction on Microwave Theory and Techniques,* **40,** 1870–1878 (1992).
26. Y. Wenyen and W. Wenbing, "The transmission properties of stratified chiroferrite media with obliquely incident plane waves", *Int. J. Infrared Millim. Waves,* **15,** 593–603 (1994).
27. J.S. Nefedov, "Microstrip slow-wave structures on the bianisotropic substrate", *Electromagnetics,* **17,** 343–360 (1997).
28. S.B. Borisov, N.N. Dadoenkova and I.L. Lyubchanskii, "Normal electromagnetic Waves in gyrotropic magnetooptic layered structures", *Opt. Spectrosc.,* **74,** 670–682 (1993).
29. A. Lakhtakia and W. Weiglhofer, "On light propagation in helicoidal bianisotropic mediums", *Proc. R. Soc. Lond. A,* **448,** 419–437 (1995).
30. A. Lakhtakia and W. Weiglhofer, "Further results on light propagation in helicoidal bianisotropic mediums: oblique propagation", *Proc. R. Soc. Lond. A,* **453,** 93–105 (1997).
31. W.S. Weiglhofer and A. Lakhatakia, "Wave propagation in a continuously twisted biaxial dielectric medium parallel to the helical axis", *Optik,* **96,** 179–183 (1994).

32. A. Lakhatakia and W.S. Weiglhofer, "Simple and exact analytic solution for oblique propagation in a cholesteric liquid crystal", *Microwave and Opt. Techn. Lett.,* **12,** 245–247 (1996).
33. C.R. Pidgeon, "Free carrier optical properties of semiconductors", *Handbook on Semiconductors,* Vol. 2, ed. By M.Balkanski (North-Holland, Amsterdam, 1980), pp. 223–328.
34. M. Schubert, T.E. Tiwald and J.A. Woollam, "Explicit solutions for the optical properties of arbitrary magneto-optic materials in generalized ellipsometry", *Appl. Opt.,* **33,** 177–187 (1999).
35. D.W. Berreman, "Optics in stratified and anisotropic media: 4 × 4 matrix formulation", *J. Opt. Soc. Am.,* **62,** 502–510 (1972).
36. H.G. Booker, "Oblique propagation of electromagnetic waves in a slowly-varying non-isotropic medium", *Proc. Roy. Soc. A,* **155,** 235–257 (1936).
37. H. Wöhler, M. Fritsch, G. Haas and D.A. Mlynski, "Faster 4 × 4 matrix method for uniaxial inhomogeneous media", *J. Opt. Soc. Am. A,* **5,** 1554–1557 (1988).
38. M. Schubert, "Polarization-dependent optical parameters of arbitrarily anisotropic homogeneous layered systems", *Phys. Rev. B,* **53,** 4265–4274 (1996).
39. See P.J. Lin-Chung and S. Teitler, "4 × 4 matrix formalism for optics in stratified anisotropic media", *J. Opt. Soc. Am. A,* **1,** 703–705 (1984), for comparison between the 4 × 4 matrix formalism developed respectively by Berreman (Ref. 35) and by Yeh (Ref. 41).
40. M. Schubert, "Generalized Ellipsometry and complex optical systems", *Thin Solid Films,* **313–314,** 323 (1998).
41. P. Yeh, "Optics of anisotropic layered media: A new 4 × 4 matrix algebra", *Surf. Sci.,* **96,** 41–53 (1980).
42. M. Mansuripur, "Analysis of multilayer thin film structures containing magneto-optic and anisotropic media at oblique incidence using 2 × 2 matrices", *J. Appl. Phys.,* **67,** 6466–6480 (1990).
43. W. Xu, L.T. Wood and T.D. Golding, "Optical degeneracies in anisotropic layered media: Treatment of singularities in a 4 × 4 matrix formalism", *Phys. Rev. B,* **61,** 1740–1743 (2000).
44. E. Toussaere and J. Zyss, "Ellipsometry and reflectance of inhomogeneous and anisotropic media: a new computationally efficient approach", *Thin Solid Films,* 234, 416–438 (1993).
45. S. Visnovsky, "Magneto-optical ellipsometry", *Czech. J. Phys. B,* **36,** 625–650 (1986).
46. S. Visnovsky, M. Nyvlt, V. Prosser, R. Lopusník, R. Urban, J. Ferré, G. Pénissard, D. Renard and R. Krishnan, "Polar magneto-optics in simple ultrathin-magnetic-film structures", *Phys. Rev. B,* **52,** 1090–1106 (1995).
47. C. Gu and P. Yeh, "Extended Jones matrix method II", *J. Opt. Soc. Am. A,* **10,** 966–973 (1993) and references therein.
48. D.W. Berreman and T.J. Scheffer, "Bragg reflection of light from single-domain cholesteric liquid-crystal films", *Phys. Rev. Lett.,* **25,** 577–581 (1970).
49. K. Eidner, G. Mayer, M. Schmidt and H. Schmiedel, "Optics in stratified media-The use of optical eigenmodes of uniaxial crystals in the 4 × 4-matrix formalism", *Mol. Cryst. Liq. Cryst.,* **172,** 191–200 (1989).
50. D.L. Jaggard and X. Sun, "Theory of chiral multilayers", *J. Opt. Soc. Am. A,* **9,** 804–813 (1992).

51. S. Bassani, C.H. Papas and N. Engheta, "Electromagnetic wave propagation through a dielectric-chiral interface and through a chiral slab", *J. Opt. Soc. Am. A,* **5,** 1450–1459 (1988).
52. A. Zunger and S. Mahajan, in *Handbook of Semiconductors,* 2nd. Ed., edited by S. Mahajan (Elsevier, Amsterdam, 1995) Vol. 3, p.1399.
53. A. Zunger, "Spontaneous atomic ordering in semiconductor alloys: causes, carriers, and consequencies", *MRS Bull.,* **22,** 20–26 (1997).
54. G.B. Stringfellow, "Order and surface processes in III-V semiconductor alloys", *MRS Bull.,* **22,** 27–32 (1997).
55. S.-H. Wei and A. Zunger, "Optical properties of zinc-blende semiconductor alloys: effects of epitaxial strain and atomic ordering" *Phys. Rev. B,* **49,** 14337 (1994).
56. D.E. Aspnes, "Approximate solution of ellipsometric equations for optically biaxial materials", *J. Opt. Soc. Am.,* **70,** 1275–1277.
57. B.O. Seraphin and N. Bottka, *Phys. Rev.,* **145,** 628 (1966).
58. T.E. Tiwald and M. Schubert, "Measurement of rutile TiO_2 dielecric tensor from 0.148 to 33 μ using generalized ellipsometry", *Proc. SPIE,* **Vol 4103,** 19 (2000).
59. M. Schubert, T.E. Tiwald and C.M. Herzinger, "Infrared dielectric anisotropy and phonon modes of sapphire", *Phys. Rev. B,* **61,** 8187–8201 (2000).
60. M. Schubert, B. Rheinländer, E. Franke, H. Neumann, J. Hahn, M. Röder and F. Richter, "Anisotropy of boron nitride thin-film reflectivity spectra by generalized ellipsometry", *Appl. Phys. Lett.,* **70,** 1819 (1997).
61. M. Schubert, B. Rheinländer, E. Franke, H. Neumann, T.E. Tiwald, J.A. Woollam, J. Hahn and F. Richter, "Infrared optical properties of mixed-phase thin films studied by spectroscopic ellipsometry and boron nitride as an example", *Phys. Rev. B,* **56,** 13306–13313 (1997).
62. A. Kasic, M. Schubert, S. Einfeldt, D. Hommel, T.E. Tiwald, "Free-carrier and phonon properties of n- and p-type hexagonal GaN films measured by infrared ellipsometry", *Phys. Rev. B,* **62,** 7365–7377 (2000).
63. M. Schubert, A. Kasic, T.E. Tiwald, J. Off, B. Kuhn, F. Scholz, "Optical phonons and free-carrier effects in MOVPE grown AlxGa1-xN measured by infrared ellipsometry", *MRS Internet J. Nitride Semicond. Res.,* **4,** 11 (1999).
64. A. En Naciri, L. Johann, R. Kleim, M. Sieskind and M. Amann, "Spectroscopic Ellipsometry of anisotropic materials: application to the optical constants of HgI_2", *Appl. Opt.* (1Febr.), 647–654 (1999).
65. A. Berger and M.R. Pufall, "Quantitative vector magnetometry using generalized magneto-optical ellipsometry", *J. Appl. Phys.,* **85,** 4583–4585 (1999).
66. T. Hofmann, V. Gottschalch, and M. Schubert, "Far-infrared dielectric anisotropy and phonon modes in spontaneously CuPt ordered $Ga_{0.52}In_{0.48}P$", unpublished.
67. See for example W.J. Tropf and M.E. Thomas, *Aluminum Oxide* (α-Al_2O_3) *revisited,* in E.D. Palik, *Handbook of Optical Constants* (Academic Press, New York, 1998) Vol. III, p. 653; E.D. Palik, *Optical Properties for the Materials in HOC I and HOC II,* ebenda, Vol. II, p. 313; F. Gervais, *Aluminum Oxide,* ebenda, Vol. II, p. 177, and references therein.
68. J.W. Orton and C.T. Foxon, "Group III nitride semiconductors for short wavelength light-emmiting devices", *Rep. Prog. Phys.,* **61,** 1–75 (1998).

69. A.K. Harman, S. Ninomiya and S. Adachi, "Optical constants of sapphire (α-Al_2O_3) single crystals", *J. Appl. Phys.,* **76,** 8032–8036 (1994).
70. A.C. DeFranzo and B.G. Pazol, "Index of refraction measurements on sapphire at low temperatures and visible wavelength", *Appl. Opt.,* **32,** 2224–2234 (1993).
71. S.-H. Wei, A. Franceschetti and A. Zunger, "E_1, E_2, and E_0' transitions and pressure dependence in ordered $Ga_{0.5}In_{0.5}P$", *Phys. Rev. B,* **51,** 13097–13102 (1995).
72. S.H. Wei, D.B. Laks and A. Zunger, "Dependence of the optical properties of semiconductor alloys on the degree of long-range order", *App. Phys. Lett.,* **62,** 1937–1939 (1993).
73. S.-H. Wei and A. Zunger, "Fingerprints of CuPt ordering in III-V semiconductor alloys: valence-band splitting, band-gap reduction, and x-ray structure factors", *Phys. Rev. B,* **57,** 8983–8988 (1998).
74. C.W. Higginbotham, M. Cardona and F.H. Pollak, "Intrinsic piezobirefringence of Ge, Si, and GaAs", *Phys. Rev.,* **184,** 821–829 (1969).
75. G. Heppke and F. Oestreicher, "Determination of the cholesteric screw sense", *Mol. Cryst. Liq. Cryst. (Letters),* **41,** 245–249 (1978).
76. R. Cano, "Optical rotary power of cholesteric liquid crystals", *Bull. Soc. Franc. Mineral.Crist.,* **90,** 333–351 (1967).
77. C. Cramer, H. Binder, M. Schubert, B. Rheinländer and H. Schmiedel, "Optical properties of microconfined liquid crystals", *Mol. Cryst. Liq. Cryst.,* **282,** 395–405 (1996).
78. I. Haller, "Thermodynamic and static properties of liquid crystals", *Prog. Sol. State Chem.,* **10,** 103–112 (1975); see also: St.Limmer, "Physical principles underlying the experimental methods for studying the orientational order of liquid crystals", *Fortschr. Phys.,* **37,** 879–931 (1989).
79. S.-T. Wu, "A semiempirical model for liquid-crystal refractive index dispersions", *J. App. Phys.,* **69,** 2080–2087 (1991).
80. W.A. McGahan, P. He and J.A. Woollam, "Optical and magneto-optical characterization of thin films", *Appl. Phys. Commun.,* **11,** 375–401 (1992).
81. W.A. McGahan and J.A. Woollam, "Magneto-optics of mutilayer systems", *Appl. Phys. Commun.,* **9,** 1–25 (1989).
82. K.W. Wierman, J.N. Hilfiker, R.F. Sabiryanov, S.S. Jaswal, R.D. Kirby and J.A. Woollam, "Optical and Magneto-optical constants of $MnPt_3$", *Phys. Rev. B,* **55,** 3093–3099 (1997).
83. J.M. Floraczek and E. Dan Dahlberg, "Detecting two magnetization components by the magneto-optical Kerr effect", *J. Appl. Phys.,* **67,** 7520–7525 (1990).
84. X. Gao, D.W. Glenn, S. Heckens, D. W. Thompson and J.A. Woollam, "Spectroscopic ellipsometry and magneto-optic Kerr effects in Co/Pt multilayers", *J. Appl. Phys.,* **82,** 4525–4531 (1997); X.Gao, D.W. Thompson and J.A. Woollam, "Determination of the interfacial magneto-optical effects in Co/Pt multilayer structures", *Appl. Phys. Lett.,* **70,** 3203–3205 (1997).
85. A. Mascarenhas, S. Kurtz, A. Kibbler and J.M. Olson, "Polarized band-edge photoluminescence and ordering in $Ga_{0.52}In_{0.48}P$", *Phys. Rev. Lett.,* **63,** 2108–2111 (1989).

86. M. Schubert, B. Rheinländer and V. Gottschalch, "Band-gap reduction and valence band splitting in spontaneously ordered $GaInP_2$", *Solid State Commun.,* **95,** 723–727 (1995).
87. M. Schubert, B. Rheinländer, E. Franke, I. Pietzonka, J. Škriniarová and V. Gottschalch, "Direct-gap reduction and valence band splitting of ordered indirect-gap $AlInP_2$ studied by dark-field spectroscopy", *Phys. Rev. B,* **54,** 17616 (1996).
88. M. Schubert, J.A. Woollam, G. Leibiger, B. Rheinländer, I. Pietzonka, T. Saß, and V. Gottschalch, "Isotropic optical constants of highly-disordered $Al_xGa_{1-x}InP_2$ ($0 \leq x \leq 1$)", *J. Appl. Phys.,* **86,** 2025–2035 (1999).
89. C.M. Herzinger, B. Johs, W.A. McGahan, J.A. Woollam, and W. Paulson, "Ellipsometric determination of optical constants for silicon and thermally grown silicon dioxide via a multi-sample, multi-wavelength, multi-angle investigation", *J. Appl. Phys.,* **83,** 3323–3336 (1998).

Emerging Areas in Ellipsometry

10 VUV Ellipsometry

James N. Hilfiker

J. A. Woollam Co., Inc.
Lincoln, NE

10.1 Introduction

Spectroscopic ellipsometry (SE) was established in the early 1970s after widespread acceptance of single-wavelength ellipsometry. The first SE systems offered limited spectral coverage from the ultraviolet to near infrared. With the exception of a few research instruments, conventional SE would remain limited to this spectral range until the 1990s. The vacuum ultraviolet (VUV), so termed because light transmission requires the evacuation of air, presented many challenges for ellipsometry. In addition to an absorbing ambient, optical materials become absorbing in the VUV. VUV ellipsometry was confined to a few research efforts, even though VUV spectroscopy had matured decades earlier.[1]

In the mid-1980s, a spectroscopic ellipsometer was constructed at the BESSY synchrotron in Berlin to measure in the VUV.[2] It has served as a center for VUV materials characterization ever since. Ellipsometry was also demonstrated in the extreme ultraviolet (EUV) at KEK-PF[3] in the 1990s. However, this instrumentation restricted general application.

The merits of VUV ellipsometry were not fully appreciated due to lack of available instruments. Then in 1999, commercial VUV ellipsometers were introduced with wavelengths down to 146 nm (8.5 eV). These instruments came in response to bulk and thin film characterization requirements for 157 nm lithography.

VUV ellipsometry was immediately applied to 157 nm lithography materials characterization. In addition, the sensitivity of short wavelengths is being explored for thin gate dielectric metrology. Access to extended photon energies is also beneficial for studying high energy electronic transitions in dielectrics and semiconductors.

This chapter is organized into the following sections:

- Historical review
- Current instrumentation
- Importance of VUV ellipsometry
- Survey of applications
- Future

10.2 Historical Review of Short Wavelength Ellipsometry

While VUV ellipsometry was briefly demonstrated as early as 1970,[4] it was not investigated further until the mid-1980s. At that time, researchers in Germany constructed a VUV ellipsometer to measure up to 35 eV (~35 nm).[5] In the mid 1990s, Japanese researchers demonstrated ellipsometry at 97 eV (~12.8 nm).[6] These remained the sole short-wavelength ellipsometry efforts until the end of the millennium.

10.2.1 BESSY Ellipsometer

In the mid 1980s, development began on a VUV ellipsometer at the BESSY storage ring in Berlin.[2] A photograph of this system is shown in Fig. 10.1. This ellipsometer was used to study material optical properties between 5 and 35 eV. Materials research has continued on the BESSY ellipsometer through the present day.[6,7–38] This instrument is briefly

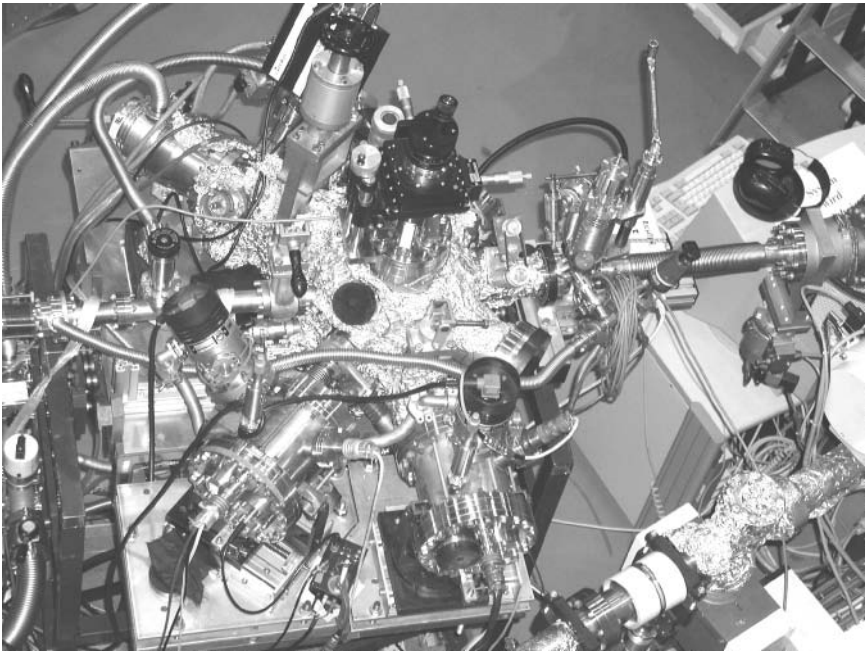


Figure 10.1 VUV ellipsometer at the BESSY 1 synchrotron in Berlin. Photograph from C. Cobet.

summarized for comparison to current technology, but further details can be found in the literature.[7–9,23]

The BESSY ellipsometer is designed for measurements from 5 eV to 35 eV (248 nm to 35 nm). This energy range is not covered adequately with one set of optics, so it is divided into two regions. Both sets of optics are available on the system, each at a separate angle of incidence: 45° and 67.5° for photon energies above and below 10 eV, respectively. An ultra-high vacuum (UHV) chamber houses the sample and the entire optical path to prevent ambient absorption. Thus, a load-lock is used to introduce samples for measurement.

The BESSY system components can be divided into the following general categories of any ellipsometer: 1) a light source, 2) wavelength selection, 3) polarization-state selection, 4) sample, 5) polarization-state analyzer, and 6) detector.

A very intense light source covering the spectrum from VUV to infrared (IR) is available from synchrotron radiation. The BESSY ellipsometer is installed on the 2m Seya-Namioka beamline to provide monochromatic light from 5 eV to 35 eV.[9] Unfortunately, this wide spectrum is not covered with a single set of optics. Below 10 eV, MgF_2 Rochon polarizers remain transparent and provide linear polarization with a high extinction ratio. A fixed polarizer before the sample selects the polarization state, while a rotating polarizer analyzes the reflected beam (RAE, see Chapter 5). Above 10 eV, MgF_2 is no longer transparent, so the BESSY ellipsometer relies on the source polarization. This requires tilting of the entire ellipsometer (including UHV chamber) to set the polarizer angle. Light from the monochromator can still contain unpolarized or elliptically polarized light on the level of a few percent.[8] For the analyzer above 10 eV, a rotating triple-reflection polarizer is implemented, as only reflective materials are available in this spectral range. The inset of Fig. 10.2 illustrates the operation of a triple-reflection polarizer. The three reflections polarize the beam without significant beam deviation, allowing polarizer rotation. However, this type of polarizer is non-ideal so the degree of polarization must be considered.

Photodiode detectors are preferred for signal detection due to their high linearity, lack of polarization dependence, uniform response, low expense, and ease of use. In particular, Si diodes were selected after earlier work with GaP Schottky diodes.[7,9] Fortunately system modularity allows it to evolve as new components become available.

Figure 10.2 illustrates the basic layout of the BESSY ellipsometer.[24] Light from the monochromator (M) enters the UHV chamber through a removable MgF_2 Rochon polarizer (P) and then through a rotary feedthrough (RF). The RF allows tilt of the entire setup along a tilt axis (T) to adjust the plane of incidence relative to the plane of polarization.

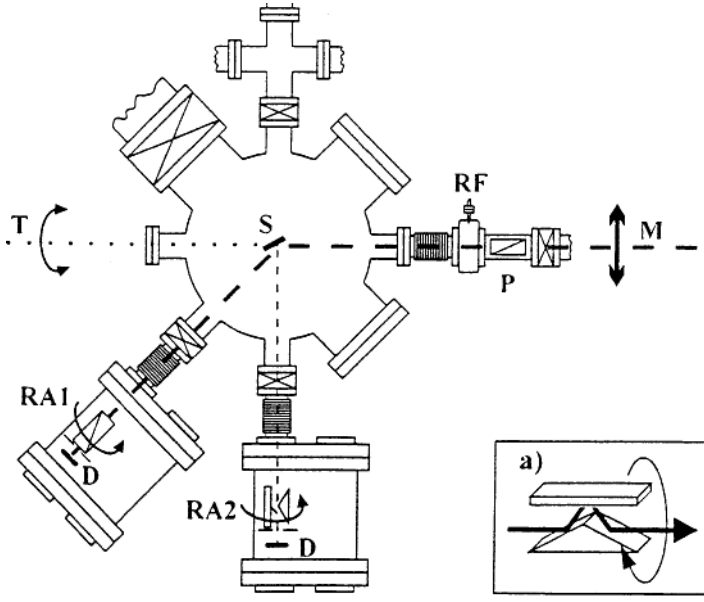


Figure 10.2 Layout of the BESSY ellipsometer, illustrating major components: light from monochromator (M), polarizer (P), rotating flange (RF), sample (S), tilt axis (T), rotating MgF_2 analyzer (RA1), rotating three-reflection analyzer (RA2), and detectors (D). The inset picture (a) shows details of the three-reflection rotating analyzer. Drawing from T. Wethkamp et al.[24]

The light then reflects from a sample (S) at either 45° or 67.5° angle of incidence. A rotating-MgF_2 analyzer (RA1) is used for photon energies up to 10 eV, while a rotating triple-reflection analyzer (RA2) is used above 10 eV. Silicon photodiode detectors (D) are used for signal collection. The inset of this figure shows the operation of the triple-reflection analyzer.

This system has been extensively used to develop both instrumentation and applications for VUV ellipsometry. The reader is directed to the many published references for more comprehensive review.[3,7–38]

10.2.2 EUV Ellipsometer

In the extreme ultraviolet (EUV), the optical constants of materials are typified by an index (n) near one and a low, but non-zero, extinction coefficient (k). Because there is little index difference (Δn) between a material and ambient, reflected intensities remain low. For example, the normal incidence reflectance for Au is about 1.5% at 60 eV.[39] The small

absorption in all materials prevents bulk transmissive optics such as lenses or windows. These complications make it difficult to control polarization in an EUV ellipsometer.

Multilayer structures, comprising two materials with small k and large Δn, can enhance reflection. Optimized reflectance near 45° can be utilized in a polarizer design,[40] as this corresponds to the "pseudo-Brewster" angle for materials with n near one. The Brewster condition will minimize p-reflected light while the multilayer structure enhances s-reflected light.

Improved performance has been demonstrated for transmission multilayer structures.[41] These are constructed by depositing the multilayer on a photoresist coated substrate. When the photoresist is dissolved, the free-standing multilayer structure is obtained.[42] Retardance can also be produced by operating at different angles of incidence.[41] Transmission-type multi-layers were found to have larger phase shifts than reflection-type.[43]

Multilayer polarizers and retarders permit the development of EUV ellipsometry. EUV ellipsometry should demonstrate increased thickness sensitivity, as the wavelength approaches an ultra-thin layer thickness (~10 nm). In addition, the high photon energy will interact with core electrons in a material.

An EUV ellipsometer was constructed using 97 eV (12.8 nm) synchrotron radiation at the Photon Factory, KEK in Tsukuba, Japan.[6,39] The

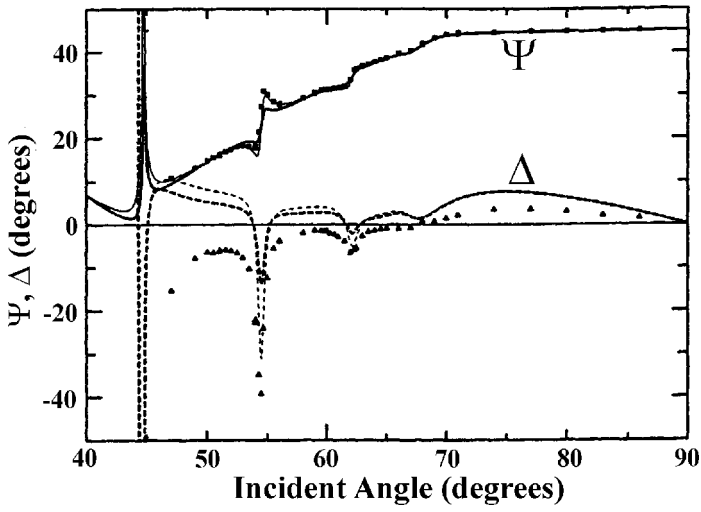


Figure 10.3 Ellipsometric result vs. angle of incidence of a Mo layer on Si measured at 97 eV. Simulation curves representing the best fit to ellipsometry and reflectance data are well resolved, although neither quantitatively matches the experimental data. Drawing from M. Yamamoto et al.[6]

synchrotron source produced linearly polarized light with 0.977 contrast factor. This was improved to 0.996 by adding a transmission polarizer before the sample. The ellipsometer is configured as a rotating analyzer ellipsometer (RAE, see Chapter 5) with a single reflection analyzer linked to the detector.[6]

This EUV ellipsometer was used to measure a Mo film on Si wafer, as this combination of materials is important in the soft x-ray region. Data were measured at 97 eV as a function of incident angle. The experimental data, represented as symbols in Fig. 10.3, exhibit resonant features at each angle where s-polarized reflectance goes through a minimum.[6] The strong resonance structure confirms penetration of the 40.9 nm thick Mo layer by the measurement beam. Thus, it is possible to study a buried interface with EUV ellipsometry, as even metals are semi-transparent.[44] Modeling attempts, shown in Fig. 10.3 as dashed and solid curves, reproduce the resonant structure but do not quantitatively match the experimental data.

10.3 VUV Ellipsometry Today

In 1999 the semiconductor industry became interested in thin films designed for 157 nm photolithography. Immediately, commercial ellipsometers were developed to meet the rapid requirements of the lithography roadmap. Now, there are many VUV ellipsometers within the world-wide research community. These instruments are predominately used by the semiconductor industry for lithography or thin gate dielectric applications. In addition, a few researchers are investigating material optical properties at high photon energies. Before the applications of VUV ellipsometry are discussed, current instrumentation is reviewed.

10.3.1 Current VUV Instrumentation

VUV ellipsometers were developed by three companies.[45–47] A general description of these instruments is available in the literature,[48–50] but full details are best obtained from the vendors. This technology is compared to the BESSY ellipsometer to demonstrate why it has become readily available.

Unlike the BESSY ellipsometer, the current instruments do not measure above 10 eV. This spectral limitation allows the use of transparent, birefringent polarizers and retarders. These same optical elements are used from the VUV to the near infrared (NIR), allowing broad spectral coverage from a single instrument (e.g., 0.73–9 eV, 138–1700 nm).

Ultraviolet light is obtained from commercial deuterium (D_2) lamps with a MgF_2 window in the lamp envelope. The D_2 lamp provides strong intensity from 115 to 400 nm (3.1 to 10.8 eV), thus it is often combined with other lamp sources (e.g., Xe lamp) for spectral coverage to longer wavelengths.

A monochromator provides wavelength selection and slit-width control. The slit width can be opened to increase signal or closed to insure adequate spectral resolution of optical features in the data. The monochromator may also contain order-sorting filters and a chopper wheel. Double-chamber monochromator designs are preferred for VUV ellipsometry to improve stray-light rejection. In addition, the monochromator is generally before the sample to avoid exposure of photosensitive materials. Fiber optics are not used because of their strong VUV absorption.

Transparent MgF_2 Rochon polarizers select and analyze the polarization state. This can be implemented in either a rotating analyzer (RAE) or rotating polarizer (RPE) ellipsometer design (see Chapter 5). The polarizers provide a high extinction ratio (10^5) over the design wavelengths of interest, so they can be treated as ideal. However, they become absorbing above 9 eV depending on material quality. The optical path in a typical Rochon polarizer is 25 to 60 mm, so even low levels of contamination can affect optical throughput. In addition to absorption, the birefringence of MgF_2 will cross zero at slightly higher photon energies – limiting its feasibility as a polarizer much beyond the current spectral range. Figure 10.4 compares the birefringence of magnesium fluoride with quartz and sapphire.[51]

Vacuum is not required for the spectral range up to 10 eV. Instead, the ambient is purged to avoid absorption from oxygen and water vapor prominent above 6.5 eV. Purge can be achieved with a noble gas like dry nitrogen or argon. Nitrogen gas exhibits absorption lines at higher photon energies, as illustrated in Fig. 10.5, suggesting a slight performance advantage for argon gas. However, nitrogen gas is generally lower cost and more readily available. To avoid ambient exposure of the entire instrument, a load-lock is typically incorporated.

The signal is detected with a photomultiplier tube,[49] a UV-enhanced Si photodiode,[50] or a combination of both.[48] InGaAs detectors can be used to extend spectral coverage into the NIR.

Current VUV ellipsometers offer many advantages. Although they are limited to photon energies below 10 eV, they can measure to much lower photon energies (e.g., to 0.73 eV). This allows data collection below and above the bandgap for many materials using the same instrument. In addition to wide spectral range, current instruments offer flexible angle of incidence. Angles are available from as low as 7°[49] to the straight-through (90°).

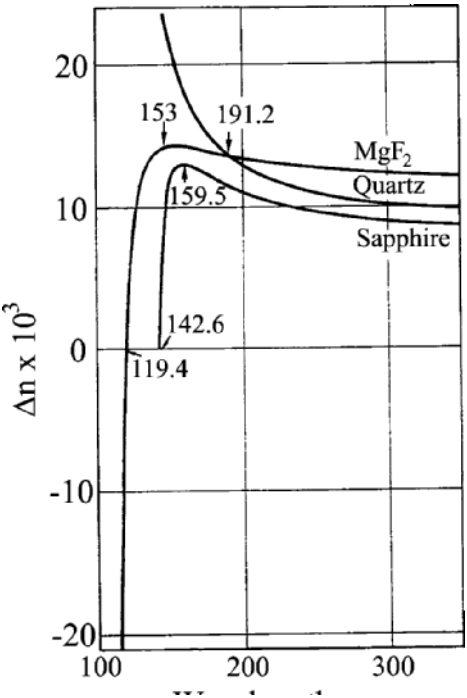


Figure 10.4 Birefringence of sapphire, MgF_2, and quartz. $\Delta n = -(n_o - n_e)$ for sapphire and $\Delta n = (n_o - n_e)$ for MgF_2 and Quartz. Drawing from Damany, Romand, and Vodar.[51]

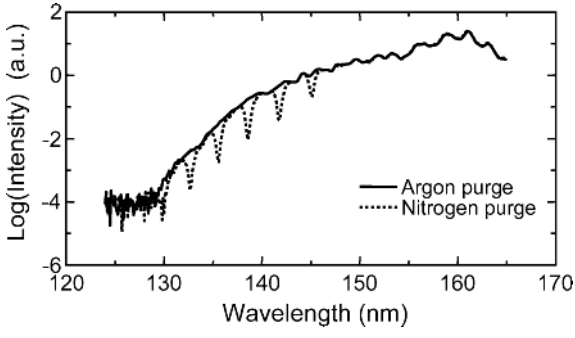


Figure 10.5 Log(Intensity) comparison between argon and nitrogen purge of the same optical path.

Angle of incidence control can be important when covering a broad spectral region, as optimum measurement conditions may change dramatically with wavelength. In addition, flexible control over both sample and input/output optics allow measurement of the reflected, transmitted, or scattered light beam.[49,52]

The spectral range limitation allows the use of MgF_2 for polarizers. In addition, MgF_2 retarders can be incorporated in the ellipsometer design. SE with an adjustable retarder eliminates the main limitations of RAE/RPE-type ellipsometers. Namely, an adjustable retarder can 1) increase accuracy of Δ when near 0° or 180°, 2) remove the handedness ambiguity in Δ, and 3) help distinguish between unpolarized and circularly polarized light for measurement of percent depolarization. In addition, a retarder allows measurement of Mueller-matrix elements.[53] The significance of these capabilities in the VUV spectral region is reviewed.

Transparent MgF_2 can be used in a Berek waveplate design[54] to adjust retardance to any value (0–90°) over a wide spectral range. With retarder-ellipsometry design, ψ and Δ can be accurately determined over their full range with no ambiguity (ψ: 0–90°, Δ: 0–360°). Verification of accuracy for this type of instrument was demonstrated from 190 nm to 1700 nm.[53] This same test is demonstrated for the VUV spectral range in Fig. 10.6. A measurement of "air" should return values for ψ and Δ of 45° and 0°, respectively. Any deviation from these ideal values indicates the level of measurement error present in a system. Figure 10.6(b) presents the same data converted to $\tan(\psi)$ and $\cos(\Delta)$, as these values are also reported in the literature. For this case, both values should be exactly one.

Handedness ambiguity in Δ is not typically a problem during data analysis, as it can often be determined from spectral dependence.[55] However, it can introduce an ambiguity for anisotropy measurements. For example, a generalized transmission ellipsometry measurement at normal incidence measures index difference, but cannot distinguish between the fast and slow axis. When the handedness ambiguity in Δ is removed, ambiguity is also removed from the fast-slow axis.

Measurement of depolarization can also benefit VUV applications. Sources of depolarization include film thickness nonuniformity, patterned film coverage, light bandwidth, and angular spread. These factors become important contributors to depolarization as wavelength decreases. Figure 10.7 shows three different depolarization measurements. In each case, the depolarization peaks increase at shorter wavelengths (higher photon energies). Depolarization is modeled and fit for each measurement. Figure 10.7(a) was modeled strictly with 2.3 nm bandwidth to match the monochromator slit width of this measurement. Figure 10.7(b) shows a nonuniform sample where 3% thickness nonuniformity was modeled in

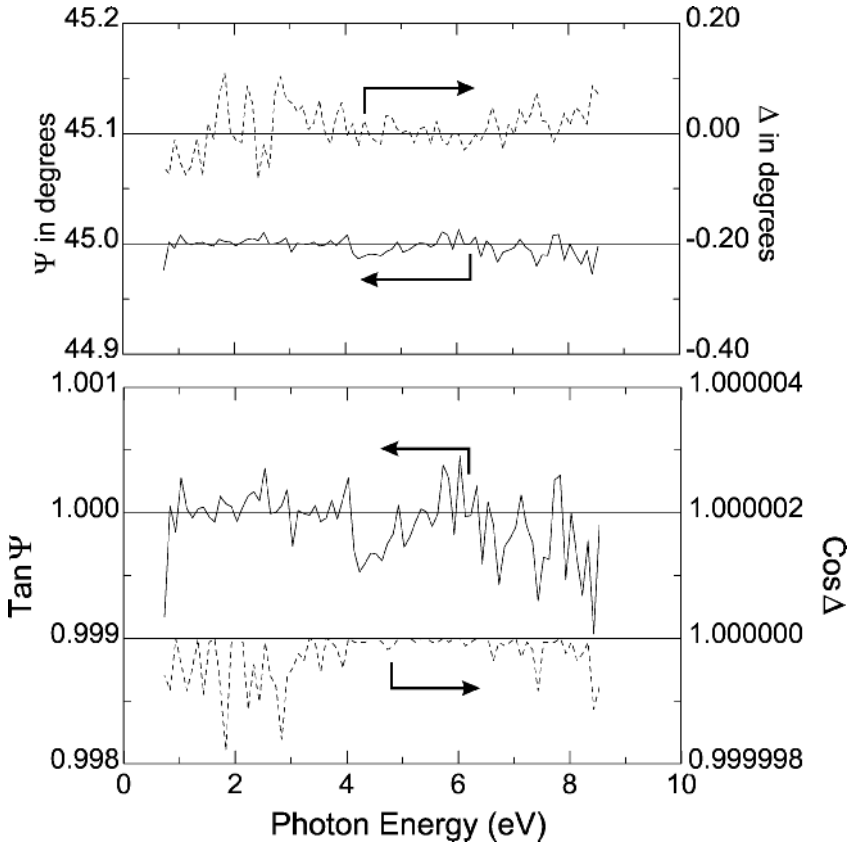


Figure 10.6 Straight-through measurements of 'air' using a VUV ellipsometer incorporating an adjustable retarder. Ideal measurement would result in (a) $\psi = 45°$, $\Delta = 0°$ and (b) $\tan(\psi) = 1$, $\cos(\Delta) = 1$.

addition to bandwidth. Figure 10.7(c) shows a more uniform film (0.5%) with much stronger depolarization peaks. This is attributed to the larger overall layer thickness. This film becomes absorbing above 7.5 eV which eliminates the source of depolarization.

Depolarization may also affect psi and delta. These effects are minimized when using a retarder, as the calculations can be based on Mueller-matrix elements rather than Jones-matrix elements.[53] If depolarization is measured and modeled correctly, the results (such as optical constants) should exhibit minimal disturbance from depolarizing non-idealities.

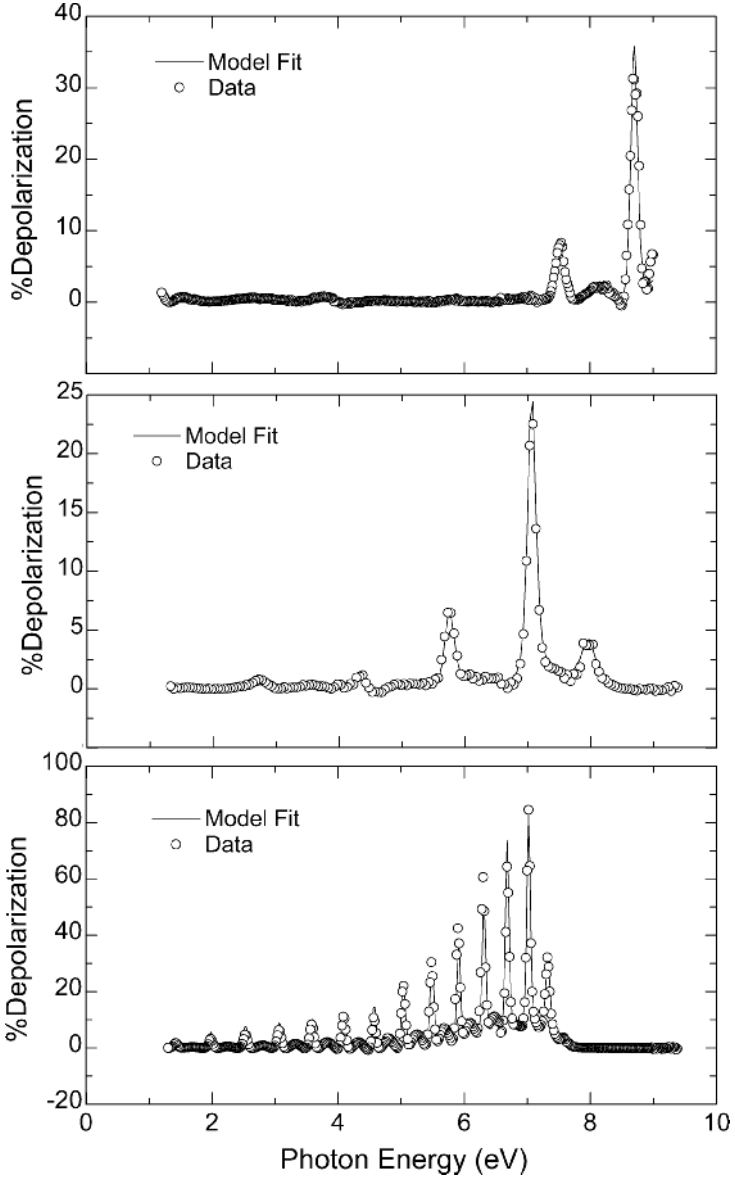


Figure 10.7 Measurements of percent depolarization using VUV ellipsometer. Data taken from (a) uniform oxide, with model of 2.3 nm instrument bandwidth; (b) nonuniform organic layer, with model of both bandwidth and 3% thickness nonuniformity; and (c) thicker organic layer with 0.5% nonuniformity added to bandwidth.

10.4 Importance of VUV Ellipsometry

VUV ellipsometry has extended the conventional range to higher photon energies. In this region, the optical response for most materials is dominated by band-to-band transitions. Light is absorbed by the material to excite an electron from the occupied valence band to the unoccupied conduction band. This process is illustrated in Fig. 10.8(a). Figure 10.8(b) shows a band diagram for GaAs. The maximum energy level of the filled valence band is considered zero for the energy scale. The energy required for any transition relates to the distance between bands. The smallest transition occurs from the top of the valence band to the bottom of the conduction band (fundamental gap). However, higher energy transitions may involve elevated conduction band states or deeper valence-electrons. The material optical properties are dependent on possible transitions and the density of states in each band. VUV ellipsometry helps extend access to larger electronic transitions in dielectrics and semiconductors. Detailed discussions of material optical properties and their relationship to electronic structure can be found in many references.[56–62]

Before VUV ellipsometry was available, the VUV optical response of materials was primarily studied using reflectance spectroscopy.[1,63] Ellipsometry provides many documented advantages over reflectance techniques[9,55,64,65] that become increasingly important in the VUV.

There are three main advantages of ellipsometry compared to reflectance:

1. Ellipsometry measures two parameters compared to the single parameter of reflectance.
2. Ellipsometry measures phase information not available from reflectance.
3. Ellipsometry measures a ratio of intensities, whereas reflectance measures absolute intensity.

Expanding on these three points, we find each to be significant in the vacuum ultraviolet. Measurement of two independent parameters allows ellipsometry data to be directly converted to "pseudo" optical properties for the measured material. This transformation is given as:[66]

$$\begin{aligned}\langle\varepsilon\rangle &= \langle\varepsilon_1\rangle + i\langle\varepsilon_2\rangle = \langle\tilde{n}\rangle^2 = (\langle n\rangle + i\langle k\rangle)^2 \\ &= \sin^2(\phi)\cdot\left[1 + \tan^2(\phi)\cdot\left(\frac{1-\rho}{1+\rho}\right)^2\right],\end{aligned} \tag{10.1}$$

where either complex dielectric function or complex index of refraction can be related to an ellipsometry measurement, ρ, at angle of incidence, ϕ.

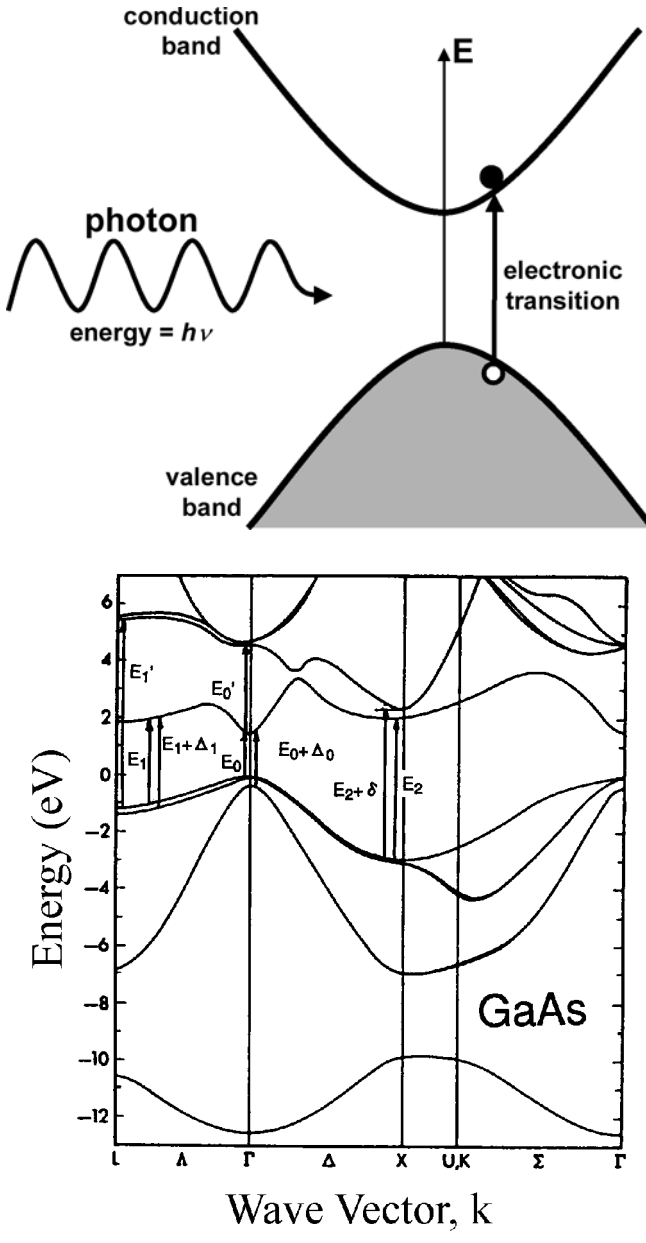


Figure 10.8 Optical absorption in the UV is primarily due to electronic transitions, depicted as (a) the excitation of an electron from the valence to conduction band upon interaction with a photon containing energy = hν (h – Planck's constant, ν – frequency). The band structure is complicated by its dependence on position in the lattice, as shown in (b) the band diagram for GaAs. Drawing (b) from Adachi.[56]

The ellipsometry measurement can be expressed using two terms (Ψ, Δ) that are related to ρ by:

$$\begin{aligned}\rho &= \mathrm{Re}(\rho) + i\,\mathrm{Im}(\rho) = \tan(\Psi)\cos(\Delta) + i\tan(\Psi)\sin(\Delta) \\ &= \tan(\Psi)e^{i\Delta} = \frac{|R_p|}{|R_s|}e^{i(\delta_p-\delta_s)} = \frac{\tilde{R}_p}{\tilde{R}_s}\end{aligned} \tag{10.2}$$

where the relationship to Fresnel reflection coefficients is also given.

Because reflectance measures a single parameter at each wavelength, there is not enough information to directly calculate optical constants. Instead, reflectance uses the Kramers-Kronig (K-K) relation, which links real and imaginary optical properties. However, this relationship requires integration of real or imaginary optical property over all photon energies to determine the opposite property. Thus, even broadband reflectance measurements will approximate the optical property outside the measured spectral range to complete the KK transformation. It is difficult to estimate possible errors in optical constants caused by this analysis. In the transparent region of a material, reflectance may be sufficient, but all materials become absorbing in the VUV. The two measured ellipsometry parameters instill more confidence in the final model for a layered structure. For reflectance, the model must match one set of measurements, whereas ellipsometry must match two curves (ψ and Δ) for each measurement angle.

The phase information contained in the ellipsometry parameter, Δ, has demonstrated high sensitivity to surface conditions and very thin films ($<$10 nm).[66] In addition, ellipsometry will gain sensitivity at shorter wavelengths. This can be demonstrated by considering a single layer on substrate. The ellipsometric values can be calculated from Fresnel reflection coefficients at each interface for angle of incidence, ϕ, as:[67]

$$\rho = \frac{r_{01p} + r_{12p}e^{-i2\beta}}{1 + r_{01p}r_{12p}e^{-i2\beta}} \times \frac{1 + r_{01s}r_{12s}e^{-i2\beta}}{r_{01s} + r_{12s}e^{-i2\beta}} \tag{10.3}$$

In this equation, the Fresnel reflection coefficients are determined from the complex refractive index of each material. The film phase thickness, β, is given as:[67]

$$\beta = 2\pi\left(\frac{t}{\lambda}\right)\sqrt{\tilde{n}_1^2 - \tilde{n}_0^2\sin^2\phi} \tag{10.4}$$

which is the link between ρ and wavelength, λ, or physical film thickness, t. It can be demonstrated that ρ will traverse a periodic polar curve as the thickness goes through a range of values equal to the thickness period, D_ϕ, where:[67]

$$D_\phi = \frac{\lambda}{2\sqrt{\tilde{n}_1^2 - \tilde{n}_0^2 \sin^2\phi}} \tag{10.5}$$

The ellipsometry values along this curve depend on angle of incidence and complex refractive index. Shorter wavelengths will traverse the entire period for thinner layers, as D_ϕ is directly proportionate to wavelength, λ. This is demonstrated in Fig. 10.9(a) for a film with $n = 2$ on substrate with $n = 4$ at angle of incidence, $\phi = 70°$. These optical constants are fixed at both wavelengths, $\lambda = 150$ nm and $\lambda = 1500$ nm, to keep all ellipsometry values along identical polar curves. Both curves start at the same point for $t = 0$ (solid symbols). As the thickness increases, the ellipsometry data for both wavelengths begins to traverse the cycle. At a thickness $t = 40$ nm, the data for $\lambda = 150$ nm (open triangles) has nearly completed an entire cycle. However, the data for $\lambda = 1500$ nm has barely moved along this path for $t = 40$ nm. In fact, it would require a thickness of 420 nm for the data at $\lambda = 1500$ nm to near a complete period.

In practice, optical constant dispersion and incident angle effects must be considered. Figure 10.9(b), shows similar polar plots simulated for the real case of SiO_2 film on silicon substrate. In this case, the two wavelengths (150 nm, 1500 nm) do not start at the same data point because of differences in refractive index. As the thickness increases from $t = 0$ to 40 nm, the curves cycle for both wavelengths, but along different paths. It is still quite evident that the shorter wavelength produces much larger changes in data for an equal thickness change. Overall, shorter wavelengths in the VUV should increase sensitivity for detection of extremely thin adsorbed films or roughness at an interface. However, it is more practical to compare VUV wavelengths to the shorter limit of conventional ellipsometry near $\lambda = 190$ nm.

Reflectance measurements are based on absolute intensity. This often requires an accurate reference sample and dual-beam configuration. In ellipsometry, one polarization component serves as a reference for the other. The absolute intensity is not measured. This can be very important in the VUV, as intensity variations can result from source or ambient-purge fluctuations. The sample can also affect overall intensity, if macroscopic features scatter a portion of the measurement beam. In this case, ellipsometry is not affected, as the measurement only collects the specular reflection and ignores the scattered component.

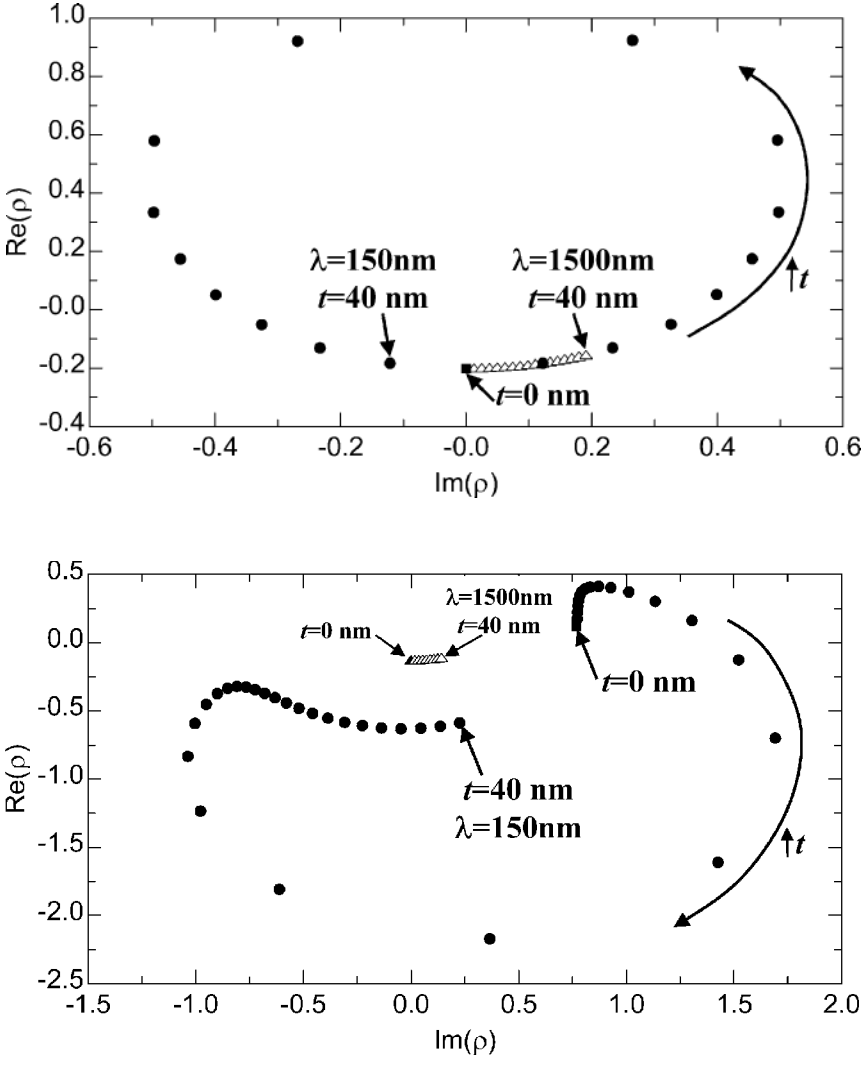


Figure 10.9 (a) Simulated spectra for wavelengths of $\lambda = 1500$ nm and $\lambda = 150$ nm at 70° angle of incidence. Identical optical constants are assumed at both wavelengths such that data traverse the same curve. (b) Simulated spectra for SiO_2 film on silicon at $\lambda = 1500$ nm and $\lambda = 150$ nm and 70° angle of incidence.

Although reflectance is more common, the advantages of VUV ellipsometry make it a desirable characterization alternative. It can be applied to bulk materials, thin films, and even multilayer structures. The next section discusses the applications of VUV ellipsometry.

10.5 Survey of Applications

The main applications of VUV ellipsometry can be grouped into three general areas: 157 nm lithography, gate dielectrics, and high-energy optical properties. Before these applications are addressed, it is important to characterize typical substrate materials in the VUV.

Silicon and thermal SiO_2 were characterized in 1999[53] based on earlier work up to 6.5 eV.[68] This became the first publication using current VUV ellipsometry and measurement was with a fixed-angle instrument. Further research on silicon has extended the optical constants to 9.5 eV.[69] The silicon optical constants over this range are shown in Fig. 10.10. The region near 7.8 eV is magnified in the inset to show a critical point attributed to a transition from $X_1^V \rightarrow X_1^C$.[69]

The silicon optical constants were described using the Gaussian-Broadened Polynomial Superposition (GBPS) parametric dispersion model of Herzinger and Johs.[53] This model is highly flexible and can describe all of the subtleties in the optical constant spectra while retaining Kramers-Kronig consistency. However, as this model is not based on the fundamental

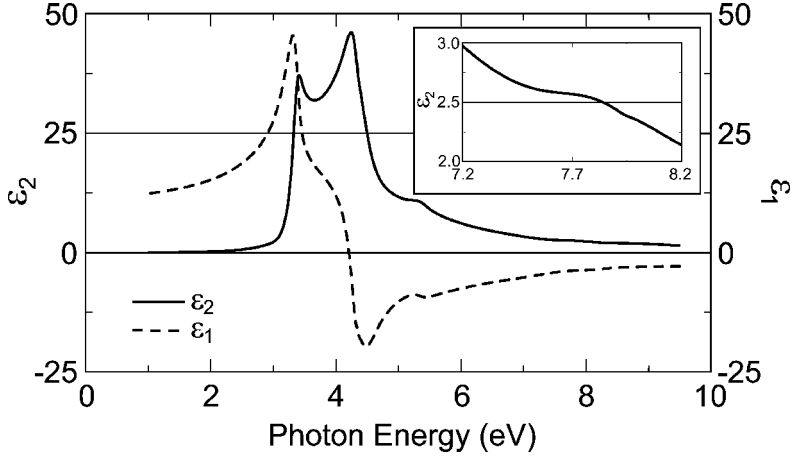


Figure 10.10 Silicon optical constants measured up to 9.5 eV (131 nm) with VUV ellipsometry. The inset figure magnifies the region around 7.8 eV (159 nm), where a critical point is attributed to the $X_1^V \rightarrow X_1^C$ transition. Data from Edwards.[69]

physics of material optical constants, the relevant model result is the determined optical spectra (not the individual parameter values).

Silicon dioxide films begin to absorb in the VUV. Figure 10.11(a,b) shows ellipsometry measurements from a 1000 nm thick thermally grown SiO_2 film on Si. A Sellmeier dispersion model is used at lower photon

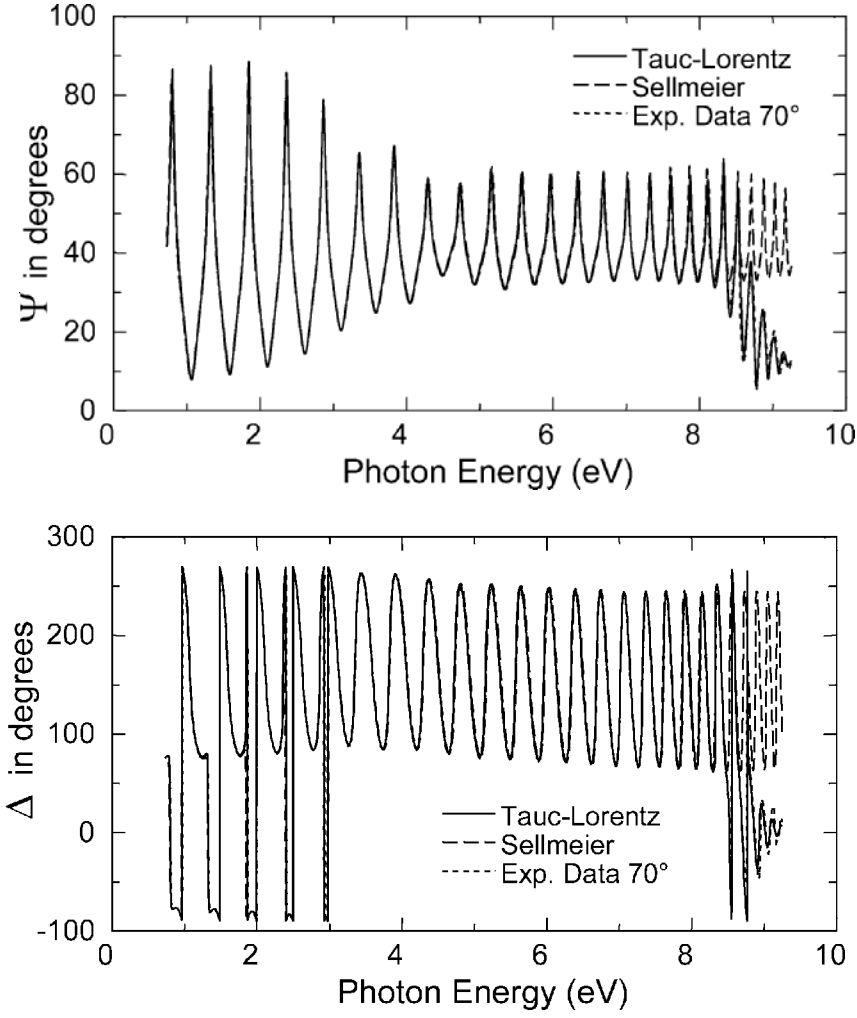


Figure 10.11 Experimental (a) psi and (b) delta data from a 1000 nm thick thermal SiO_2 film on Si are graphed along with the best-fit models using Sellmeier and Tauc-Lorentz dispersions. The resulting optical constants (c) modeled using a Tauc-Lorentz oscillator are shown up to 9.5 eV (131 nm).

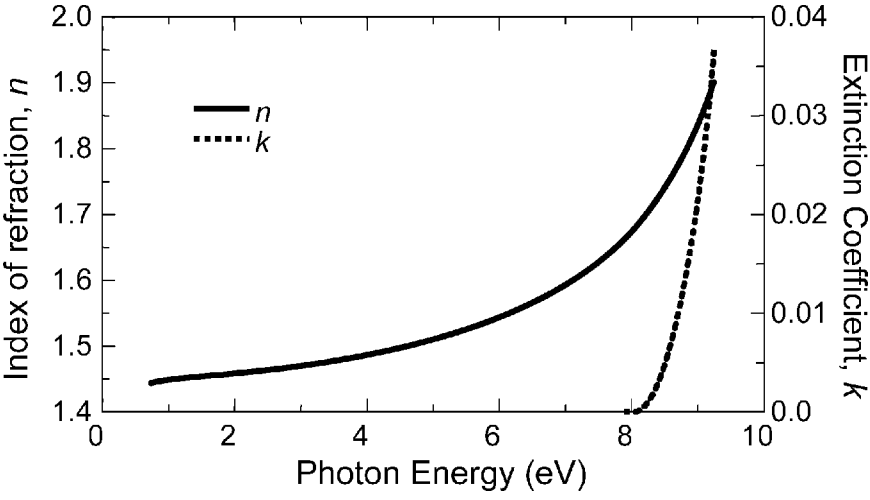


Figure 10.11 (*cont.*)

energies,[68] but is not adequate above 8 eV. Instead, a Tauc-Lorentz oscillator model[70] is used to fit this data. Both Sellmeier and Tauc-Lorentz fit results are graphed against the experimental data in Fig. 10.11(a,b). The Tauc-Lorentz fit follows the experimental curves for all wavelengths, while the Sellmeier model diverges from the experimental data above 8.25 eV. The resulting Tauc-Lorentz optical constants for SiO_2 are shown in Fig. 10.11(c).

The thick thermal SiO_2 demonstrates the importance of an appropriate optical model. In the VUV, absorption may become a consideration for dielectrics that are transparent at conventional SE wavelengths. Organosilicate glass (OSG) is used as an interlayer dielectric in the semiconductor industry due to a low dielectric constant. A transparent dispersion relation like the Cauchy or Sellmeier is often used for OSG. However, the film absorbs at shorter wavelengths (higher photon energies) and an oscillator model is needed. Figure 10.12 shows the ellipsometric data from an OSG film on Si with the corresponding fit.[71] The model used to describe optical constants up to 9.5 eV (131 nm) included a Tauc-Lorentz and two Gaussian oscillators. The spectral range is divided to show appropriate regions for reduced model types.[71]

Another common substrate for VUV characterization is CaF_2. This substrate was studied on the BESSY ellipsometer between 10 and 35 eV, covering the range of both valence and core electron transitions.[10] Below 10 eV, CaF_2 was thought to be transparent and isotropic. For this reason, it is the preferred substrate for 157 nm lithography optics. VUV

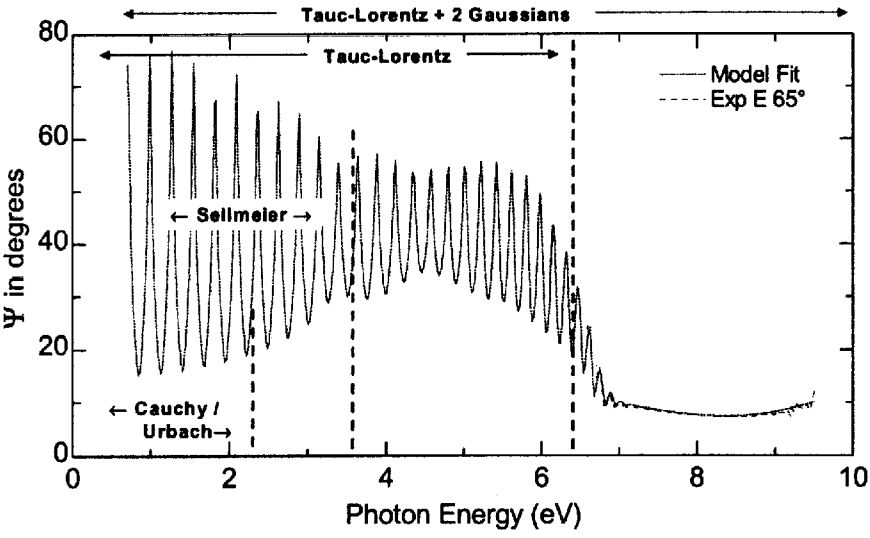


Figure 10.12 Ellipsometric angle Ψ data (dashed line) vs. photon energy for OSG on Si at 65° angle of incidence. The optical model (solid line) that best fit the data comprised of a Tauc-Lorentz and two Gaussian oscillators. The dashed vertical lines denote spectral range of viability for other optical models. From Edwards et al.[71]

ellipsometry measurements of bare CaF_2 provide a precise refractive index for this common optical substrate.[49,52] However, transmission measurements show inconsistencies due to contamination.[72] CaF_2 has received additional attention with the discovery of intrinsic birefringence.[73] This unexpected anisotropy along the **110** crystal direction interferes with the performance of lithography lens designs.[74] Generalized transmission ellipsometry can be used to characterize this effect.[75,76] Additional anisotropic substrates characterized with VUV ellipsometry include SiC[52,77–79] and TiO_2.[80] All three materials are discussed in further detail later in this chapter.

Many compound semiconductor substrates were characterized on the BESSY ellipsometer. These include GaAs,[9] InP,[9] GaP,[9] ZnS,[8,9] CdTe,[9] CdSe,[28] and SiC.[15,16,38] The spectral range of the BESSY instrument covers many higher energy electronic transitions including some core transitions.

10.5.1 Lithography

Lithographic exposure at 157 nm is targeted to succeed 193 nm for smaller device patterning. The optical constants of photolithography films

are critical to their overall performance. These films include photoresists, top and bottom anti-reflective (AR) coatings, hardmasks, photomasks, pellicles, and stepper optic coatings. Current VUV ellipsometers were developed to meet these characterization requirements.

10.5.1.1 Photoresists

For photoresist films, the optical constants are important because they govern the reflection, transmission, and absorption of light. Along with film thickness, the optical constants help define the functionality of a resist. Resists are photosensitive, so it is important that the measurement beam does not expose the film. A monochromator before the sample helps avoid this hazard. VUV ellipsometry has been used to characterize the properties of many common resist materials at 157 nm.[49,64,81–84]

Figure 10.13(a) shows the ellipsometric measurement from a typical photoresist layer. The oscillations at longer wavelengths are caused by interference between light that travels through the film and light reflecting from the film surface. At UV wavelengths, these oscillations cease because the light is absorbed before it penetrates the film.

There are two common modeling approaches for transparent films that absorb at shorter wavelengths: the wavelength-by-wavelength (λ-by-λ) approach and the oscillator model approach. Both may start by fitting the transparent region to determine film thickness. In the λ-by-λ approach, the resulting thickness is fixed to determine n and k directly from ψ and Δ. This works well if the sample structure is correctly described by the model. The data from Fig. 10.13(a) were modeled using the λ-by-λ approach to determine the photoresist optical constants shown in Fig. 10.13(b).[85]

The oscillator model approach describes the optical constants using a dispersion relation. The parameters of this relation are adjusted to match the experimental data. Both methods are acceptable and have inherent advantages. In many cases, these approaches are combined for further benefit.

The λ-by-λ approach is easy to perform on many different materials, provided the film is transparent over part of the measured spectral range. It can model complicated optical features that may be difficult to describe adequately with an oscillator model. However, the λ-by-λ approach does not insure "physical" optical constants and may be susceptible to incorrect results if the model structure is not correctly described.

Most oscillator models maintain Kramers-Kronig consistency, so the resulting optical constants maintain a "physical" shape. In addition, the number of "fit" parameters is greatly reduced, as the optical constants are

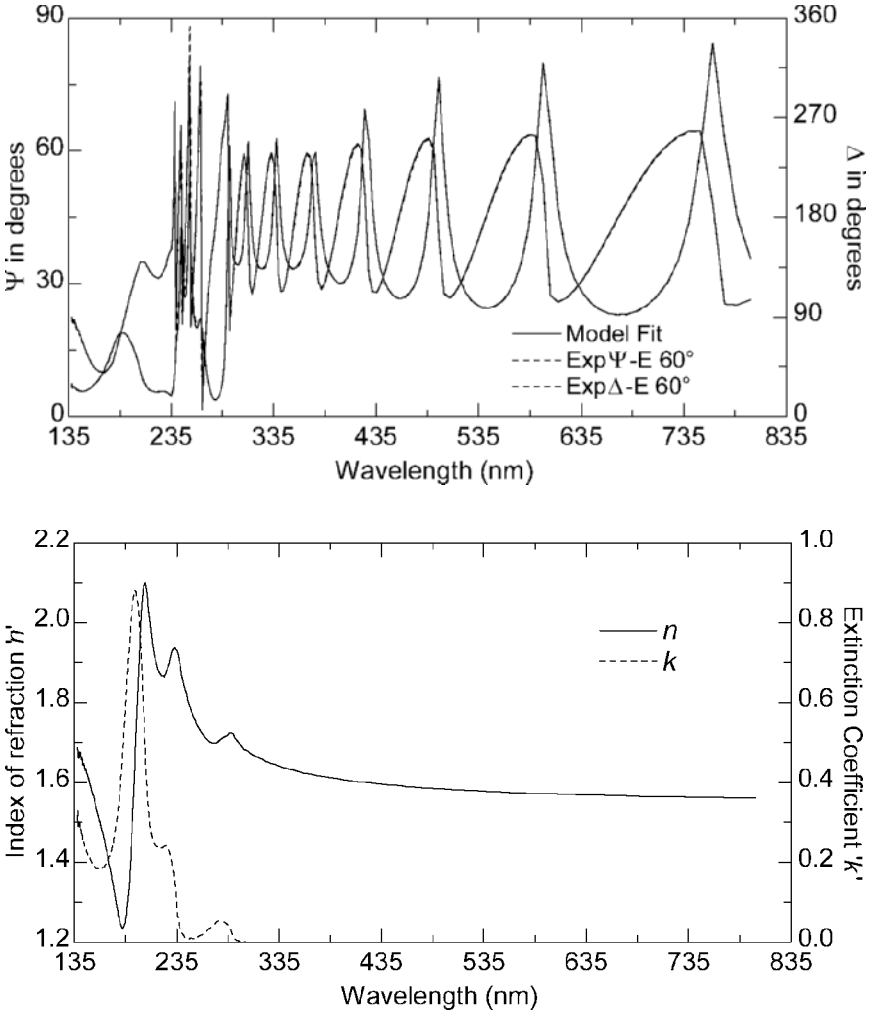


Figure 10.13 Resist data (a) and optical constants (b), from 85.

described by an equation over the entire spectral range. This may allow the thickness to be fit simultaneously with oscillator model parameters. Model improvements are possible when the transparent region alone is not sensitive to the correct sample structure – for instance, if there is a surface layer or index gradient in the film.

Photoresists are a predominant challenge for 157 nm lithography. The standard polyethylene backbone, used for all previous organic resists in semiconductor manufacturing, is too absorbing at 157 nm.[81] New

candidate materials, such as fluorocarbons and polysiloxanes, are under development.[81] Until a more transparent alternative is found, thinner resists and bi-layer designs are being incorporated.[84,86] For example, a thin resist layer can be deposited over a hardmask layer. The hardmask provides etch resistance because the thin resist cannot stand up to the etch process. Optimum design of this structure requires accurate optical constants for each material. Then, different resist and hardmask thickness combinations can be simulated. Many hardmask candidates have been investigated, including silicon nitrides, silicon oxynitrides, silicon carbide, and titanium nitride.[84,86,87] These sub-stoichiometric materials offer flexible optical constants as process conditions vary atomic composition.

10.5.1.2 Anti-Reflective Coatings

Anti-reflective (AR) coatings can be used with photoresists to suppress reflections from the substrate. A bottom anti-reflective coating (BARC) absorbs light to eliminate reflective notching of the sidewall and prevent variability in critical dimensions caused by standing waves. The optical properties of the BARC are designed to match the photoresist index to minimize reflections at the resist-BARC interface, while absorbing transmitted light before it reflects from the substrate. Optimized films enhance sidewall profiles.[86,87] Convenient AR coatings can be "tuned" to match optical constants for a given resist. Both organic and inorganic films have been used for this purpose. Silicon oxynitrides (SiON) and silicon-rich nitrides (SRN) have demonstrated significant variation in refractive index as the silicon, oxygen, and nitrogen compositions change. Figure 10.14 shows the real and imaginary refractive indices of SiON[48] and SRN[88] films processed for different compositions. In this manner, a wide range of optical properties are possible at the lithography exposure wavelength.

10.5.1.3 Photomasks

Photomasks have advanced beyond the single, opaque layer of a binary mask. Phaseshift masks have been developed to improve linewidth resolution. These masks operate by shifting the phase to create destructive interference between adjacent features.[89] This technique has successfully extended the performance of 248 nm lithography and is expected to provide similar benefits at future lithography wavelengths.

Unlike the single opaque chromium pattern of a binary mask, phase shift mask designs are more complicated. For instance, attenuated

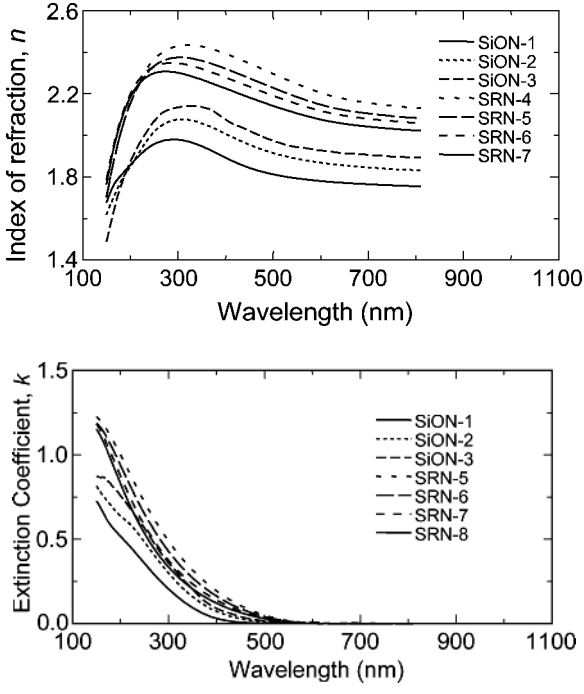


Figure 10.14 Variation in SiON [48] and SRN[88] under different process conditions.

phase-shifting photomasks use a semi-absorbing layer that must control the desired transmission and reflectivity and introduce 180° phase shift at the exposure wavelength. Often, a complex gradient is required to get all desired properties from a single film. Ellipsometry can characterize these films, but complicated optical gradients may require advanced methods. Variable angle SE measurements in both reflection and transmission geometries were combined with transmission intensity measurements to characterize the optical gradient profile in Cr-based films.[90,91] Similar films are expected to be important at 157 nm. However, new materials will need to be developed at these wavelengths, as the current materials are likely too absorbing. Preliminary VUV ellipsometry of Cr and CrO_x films help determine the potential of current materials for use at 157 nm.[92]

10.5.1.4 Pellicles

The pellicle is a protective barrier next to the mask that keeps particles away from the exposure focal point. Optical transparency is critical

for this application. Research is targeting discovery of an ultra-transparent pellicle material that can withstand 157 nm exposure.[72,93] Both VUV ellipsometry and transmission measurements are used to research new materials. VUV ellipsometry was found more consistent over a range of samples. Transmission-based absorbance measurements were fraught with problems: 1) variability in CaF_2 substrates due to contamination required measurement of each uncoated CaF_2 substrate to remain accurate, 2) optimum thickness range places absorbance between 3 and 90%, and 3) accurate thickness of material is important to get accurate absorbance.[93] In addition, transmission-based absorbance measurements do not determine the index of refraction, which is required for pellicle fringe optimization.[93]

Preliminary pellicle materials are studied by depositing thick layers on silicon.[72] The actual pellicle is a free-standing membrane near one micron thick. This thickness produces coherent interference, such that both thickness and optical constants can be determined from ellipsometry measurements. It is important to consider the depolarization effects in thick layers on both silicon and free-standing membranes. Retarder-based ellipsometry measurements were used to measure percent depolarization and quantify the thickness nonuniformity and measurement bandwidth effects.[93] After a pellicle is characterized, it is monitored to detect photochemical darkening. Thus, the lifetime for new materials can be investigated by measuring optical constant modifications.

10.5.1.5 Stepper Optic Coatings

Many optical elements are involved in lithography exposure tools. They can be enhanced with anti-reflection and high-reflection optical coatings. The coating performance will depend on optical constants and thickness. Fluorinated materials are candidates at 157 nm, as they remain transparent into the VUV. VUV ellipsometry has characterized an extensive range of optical coatings, including MgF_2,[75] LaF_3,[75,92,94] AlO,[78] AlF,[78] AlO-AlF,[78] TaO,[95,96] Ta_2O_5,[94] TiO_2,[94] Al_2O_3,[94] HfO_2,[94,97] ZrO_2,[97] and La_2O_3.[97] The optical properties of these films can now be used in coating design.

Optical coatings are often characterized on transparent substrates because they are applied to lenses and windows. This allows combination of SE and transmission measurements on the film. Transmission measurements help determine low-level absorption provided the substrate is optically characterized. Fortunately, current VUV ellipsometers can also acquire intensity transmission measurements. Figure 10.15 shows both SE and intensity transmission measurements for an Al_2O_3 film on CaF_2.[94] Both

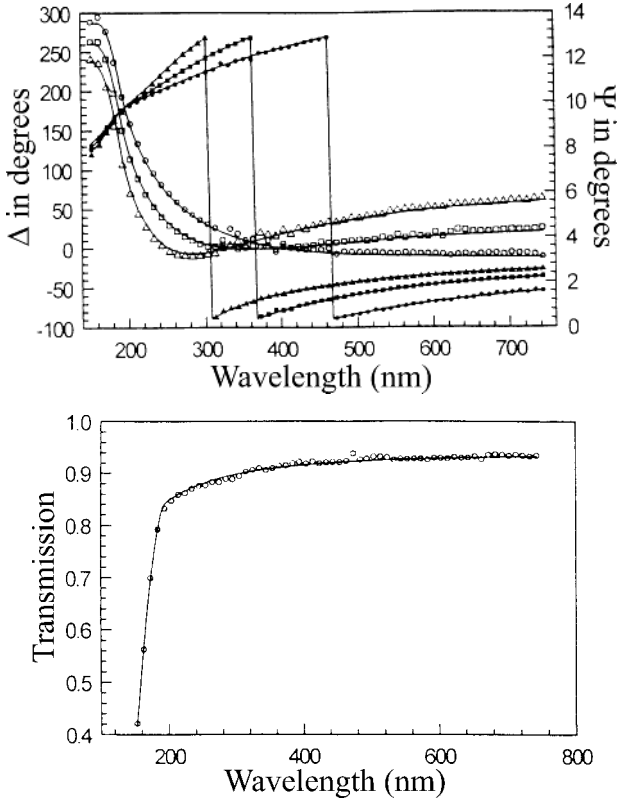


Figure 10.15 Ellipsometry and transmission measurements for an Al_2O_3 film on CaF_2. Both data sets are fit simultaneously to determine accurate optical constants and thickness.[94]

data sets are fit simultaneously to get a single, consistent set of optical constants. The model-fit curves are also shown on the figures.

10.5.1.6 Intrinsic Birefringence

Recently, the intrinsic birefringence of CaF_2 in the VUV raised concern regarding its suitability as an optical substrate at 157 nm.[73] Birefringent materials exhibit different refractive indices along the fast- and slow-axis. This causes a phase change for transmitted light with electric field components along each direction, as shown in Fig. 10.16.

Ellipsometry measures the phase change between the p- and s-polarization directions, as Δ. Thus, retardance can be directly measured if the ellipsometer

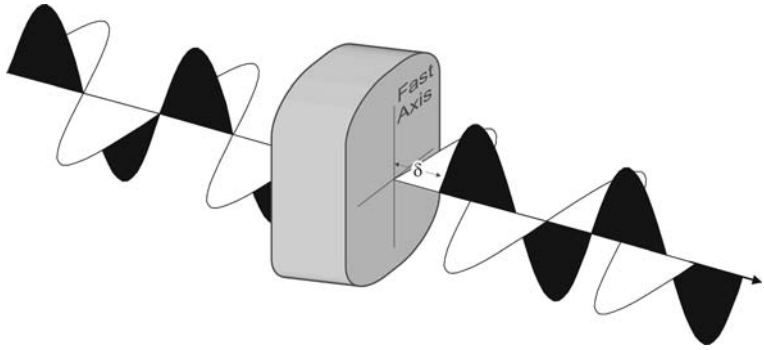


Figure 10.16 Retardance, δ, occurs when light travels through a birefringent material ($n_o \neq n_e$) where orthogonal components travel with different phase velocities.

(p- and s-) is aligned with the material (fast- and slow-optical axis). In this case, we can write the sample Jones matrix in terms of the retardance, δ, for an ideal retarder:

$$\begin{bmatrix} \tan(\psi) \cdot e^{i\delta} & 0 \\ 0 & 1 \end{bmatrix} \tag{10.6}$$

If the retarding sample is rotated relative to the ellipsometer coordinates, there can be cross coupling between polarization states (non-zero off-diagonal elements). A mathematical rotation can account for rotation of the coordinate system by angle, α:

$$\begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} \tan(\psi) \cdot e^{i\delta} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \tag{10.7}$$

Generalized ellipsometry measurements calculate both diagonal and off-diagonal matrix elements for a sample (see Chapter 9). If the generalized ellipsometry measurement is rotated to diagonalize the measurement, both orientation and retardance can be determined. Figure 10.17 shows the retardance from 20 mm thick CaF_2 as measured with generalized VUV ellipsometry.[75,76] The measured retardance in degrees is related to the optical anisotropy ($\Delta n = n_o - n_e$) and the retarder thickness, t:

$$\delta = \frac{360^\circ}{\lambda_0} t|\Delta n|, \text{ in degrees} \tag{10.8}$$

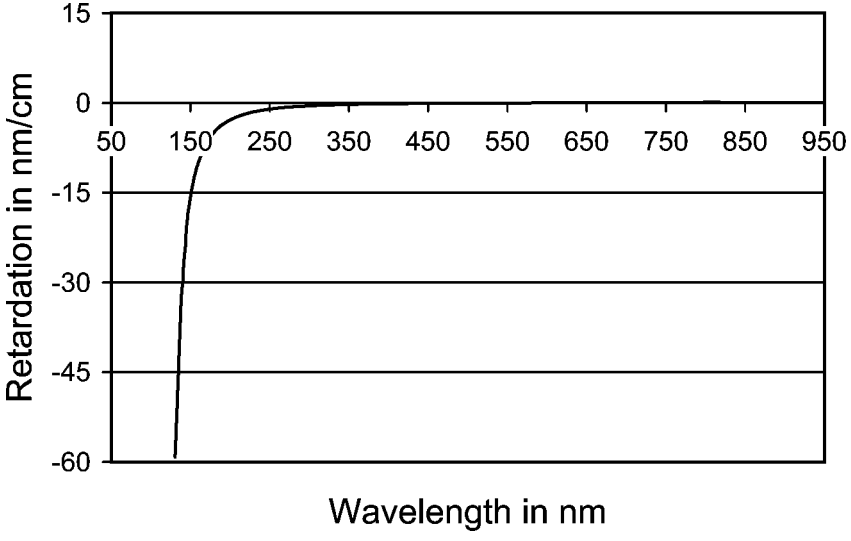


Figure 10.17 The retardance (in nm/cm) for CaF_2 along the [110] direction measured with generalized ellipsometry.

where λ_0 is the vacuum wavelength. The strong wavelength-dependence of CaF_2 retardance requires measurement at 157 nm, as extrapolation from the visible would be deceptively small.

10.5.2 Gate Dielectrics

As transistors shrink, the gate-dielectric thickness is reduced to maintain capacitance. Film thickness is currently approaching a level where the layer is unable to prevent tunneling current. One solution is to increase the dielectric constant of the gate-dielectric. This would allow a thicker film to have equivalent capacitance to a thin SiO_2 layer. Nitrided SiO_2 is already implemented for this purpose. It is critical to monitor thickness and nitrogen composition of these thin layers. VUV ellipsometry has been proposed as a possible solution with the hope that shorter wavelengths will 1) reduce correlation between thickness and index, and 2) provide composition related to the UV absorption of a film. Nitrogen content has been measured by VUV ellipsometry in thicker films.[34–37] Unfortunately, gate-dielectrics are too thin to provide accurate results. Researchers are also proposing combined techniques, such as VUV ellipsometry with grazing-angle x-ray reflectivity. The x-ray measurements provide thickness to

use in the VUV ellipsometry model, which calculates optical constants (related to nitrogen levels).[97–99]

A host of future gate-dielectric candidates are also being investigated. Oxides with perovskite crystal structure, like $SrTiO_3$, have been measured in the VUV to investigate microstructure and band structure.[100] High dielectric constants are also possible with metal oxides like HfO_2, ZrO_2, and La_2O_3. Figure 10.18 shows the index of four potential gate-dielectrics measured with VUV ellipsometry.[97]

10.5.3 High-energy Optical Constants

Optical properties in the VUV are dominated by high-photon energy electronic transitions from the valence to conduction band. The extended energy range of the BESSY ellipsometer also presents access to core-electron transitions. Major applications of the BESSY ellipsometer have focused on band-structure studies and optical response at high photon energies. This area has also carried over to recent VUV ellipsometry research.

Because conventional ellipsometry has access up to 6.5 eV, the main emphasis of VUV studies involves wide-gap transitions. Many dielectrics are studied as optical coatings (Section 10.5.1.5) or in microelectronics (Section 10.5.2). Other coatings, such as diamond-like carbon (DLC), are of interest due to physical properties including hardness and low-friction. DLC has been used for many applications, including protection from wear in the magnetic recording industry. Carbon can bond in both planar, sp^2, and tetrahedral, sp^3, configurations, which allow films to be either graphite-like or diamond-like. The VUV optical constants have been studied for graphite,[11] diamond,[12] and DLC[13] on the BESSY ellipsometer. Graphite-like films have a very low bandgap, whereas diamond exhibits its first direct gap near 7.1 eV. Thus, the bonding properties of DLC can be related to bandgap and optical response of the film. Studies of DLC with the BESSY ellipsometer showed a decrease in bandgap with annealing, suggesting increased sp^2 nature of the films.[13]

Another material that forms both sp^2 and sp^3 bonds is BN. This material is also of interest for its mechanical hardness, among other properties. VUV ellipsometry has been used to study both cubic (sp^3) and hexagonal (sp^2) BN films up to 9.5 eV.[14] The hexagonal structure exhibited two strong absorptions near 6.5 eV and 11 eV, whereas cubic BN has a single absorption near 15 eV.[14] The hexagonal crystal structure also introduces anisotropy in the optical response,[101] but this effect has not been considered in the VUV.

Anisotropy in other materials has recently been studied by generalized VUV ellipsometry (see Chapter 9). It is important that the measurement

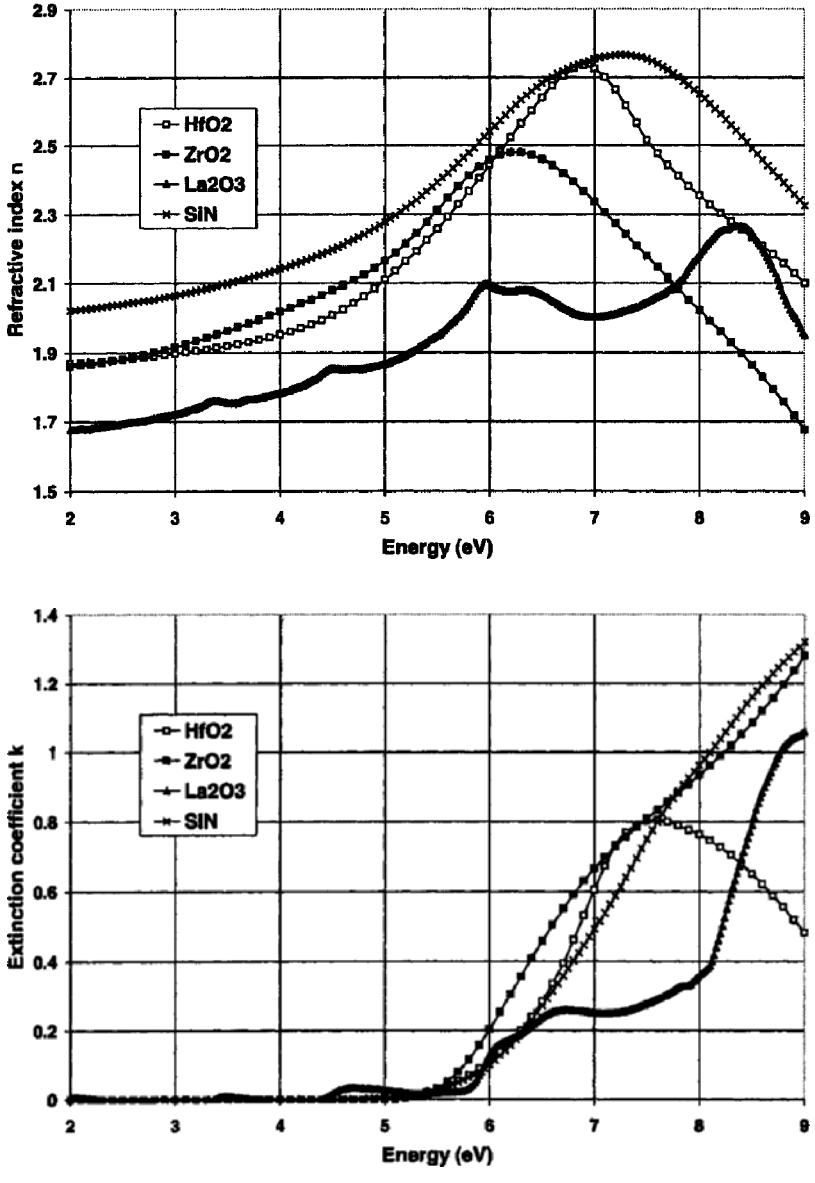


Figure 10.18 Optical indices extracted from VUV SE for potential gate dielectric films.[97]

electric field is able to excite a significant response along both the ordinary (n_o) and extraordinary (n_e) directions to determine both properties.

Convenient access to both n_o and n_e is obtained when the optic axis is parallel to the sample surface. If the material is transparent, generalized transmission ellipsometry can provide high sensitivity to index difference ($\Delta n = n_o - n_e$) as large path lengths can be traversed. For CaF_2, the measured anisotropy is very small, but can be accurately determined with common path lengths of 20 mm or longer. If the thickness is accurately known, Δn can be extracted from the measured retardance using Eq. (10.8). Sapphire exhibits much stronger anisotropy than CaF_2. The measurement in Fig. 10.19(a) is from a 0.515 mm thick sapphire substrate. The frequent oscillations in off-diagonal ellipsometry data occur as the slow axis retards light by many orders relative to the fast axis. The diagonal ellipsometry data for this normal incidence measurement is ideally equal to 45°. The perturbation near 6 eV is due to dichroism in the crystal ($k_o \neq k_e$). The off-diagonal data behavior near 157 nm (7.9 eV) is due to a "roll-over" of Δn. This behavior might not be correctly identified without measurements over a wide spectral range. The resulting Δn from this measurement is given in Fig. 10.19(b).

For rutile TiO_2 substrates, rotation of in-plane oriented substrates relative to the measurement plane of incidence allow excitement of both ordinary and extraordinary optical axes. Multiple sample orientations were measured and simultaneously regressed to extract both ordinary and extraordinary dielectric functions, as shown in Fig. 10.20 for the **110** oriented TiO_2 substrate.[80]

VUV ellipsometry has also characterized ordinary and extraordinary dielectric functions for 4H[61,72–74] and 6H[73,74] SiC polytypes. Anisotropy is related to the hexagonal crystal structure, which produces a different dielectric response for light traveling parallel or perpendicular to the c-axis of the crystal. To separate each crystal direction, a-plane samples with the optical axis lying in the sample surface are measured at multiple sample orientations relative to the plane of incidence. In addition, multiple angles of incidence are measured for each sample orientation. Figure 10.21 shows only the 70° data from a 4H-SiC substrate at three different sample orientations. The three measured curves would be identical for an isotropic material.

The λ-by-λ approach is able to obtain both ordinary and extraordinary dielectric constants as all three sample orientations are fit simultaneously. Each curve is a generalized ellipsometry measurement, even though only the diagonal measurement values are graphed. The off-diagonal components were included in the data analysis, as they help determine each measurement orientation. The resulting dielectric constants for both ordinary and extraordinary directions are shown in Fig. 10.22.[65]

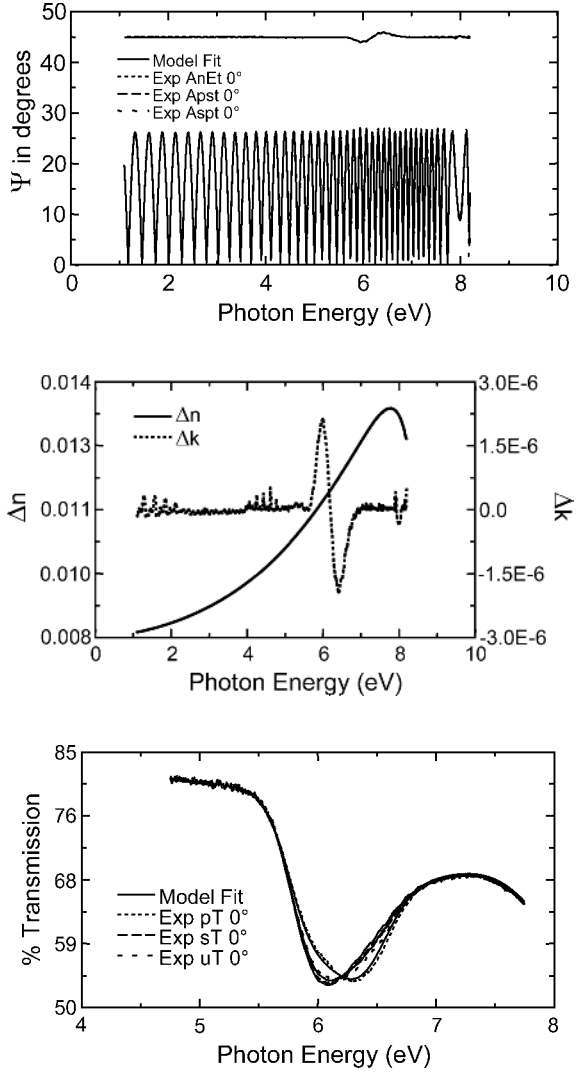


Figure 10.19 Generalized transmission ellipsometry measurements (a) taken through a sapphire substrate with optical axis parallel to the sample surface. The long path length (0.515 mm) through anisotropic material provides retardance by many wave orders. The anisotropy determined from this measurement is shown as (b) the index difference (Δn) and the extinction coefficient difference (Δk). The only dichroism ($\Delta k \neq 0$) measured above the noise-level occurs near 6 eV. This is verified by (c) a separation in the polarized transmission measurements over the same photon energy range.

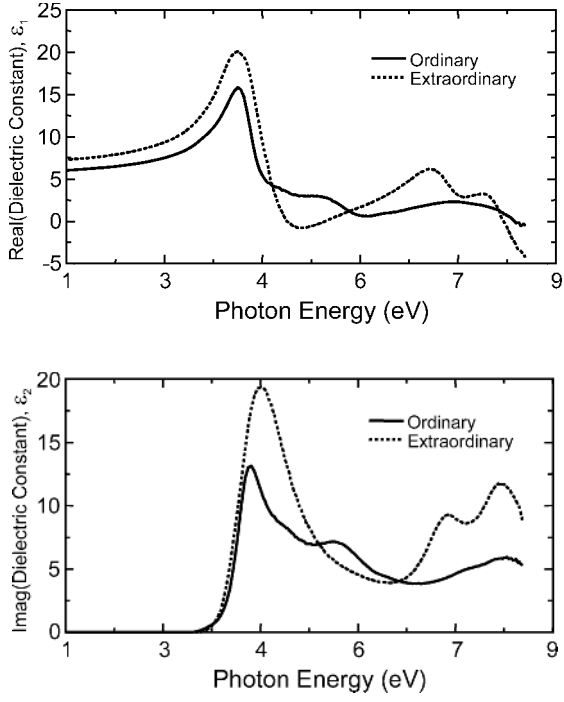


Figure 10.20 Real and imaginary parts of ε_o and ε_e versus photon energy for TiO_2 (110) orientation.[80]

The III-nitrides are promising for optoelectronic applications in the violet-ultraviolet due to a direct gap in this region that adjusts with composition. The E_1 and E_2 interband transitions of these materials are found at VUV photon energies. VUV ellipsometry has been used to study dielectric functions of GaN,[19–21,50] AlN,[22,23,102] AlGaN,[24,103,104] and AlInN.[105] The ternary compounds provide more flexibility to tune optical response. Figure 10.23 shows the pseudodielectric function for a range of GaN and $Al_xGa_{1-x}N$ samples. The E_1 and E_2 interband transitions are marked by arrows, but cannot be differentiated for $x \geq 0.016$.[103] As the Al concentration increases up to 0.12, the E_1 and E_2 peaks shift toward higher and lower energies, respectively. However, increasing free-electron concentration was also found to red-shift the E_1 transition, with no effect on E_2 position (as demonstrated on the range of GaN samples plotted). Thus, it is critical to understand both effects to quantify the resulting transition energies.[103]

BeZnSe has been considered for green and blue laser diodes, as bandgap can be tailored between that of BeSe (5.5 eV) and ZnSe (2.7 eV).

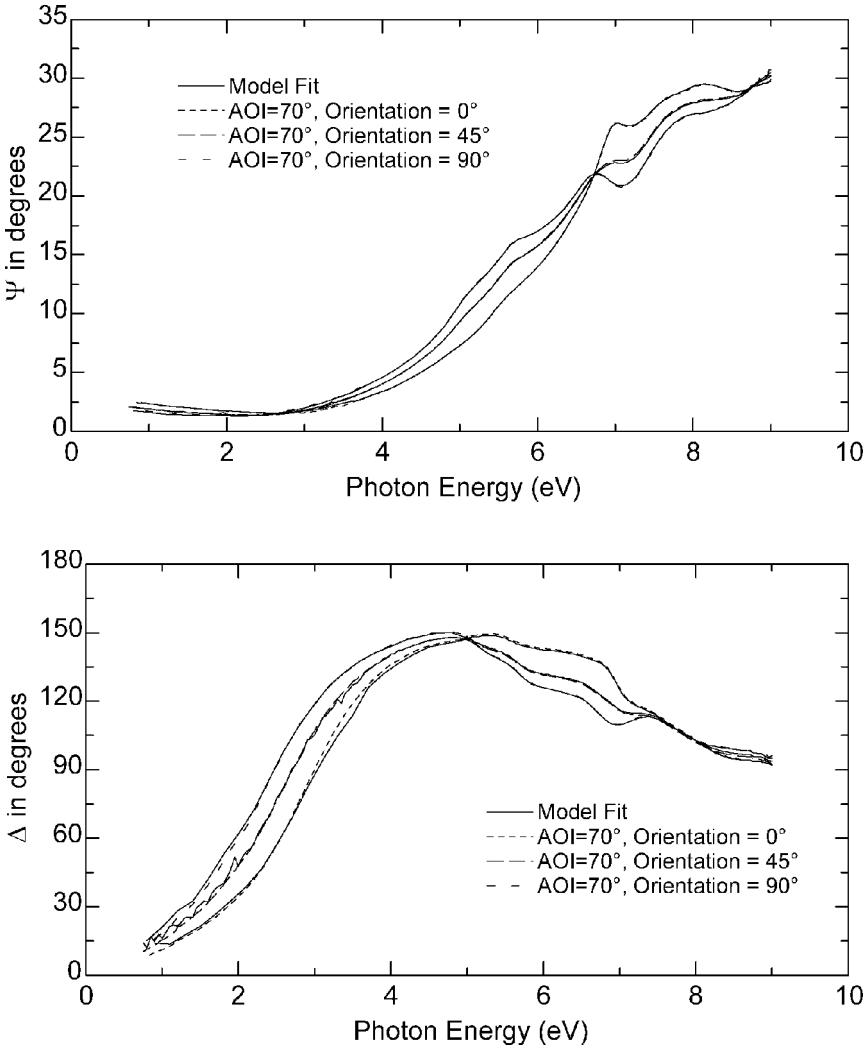


Figure 10.21 Generalized ellipsometry measurements of 4H-SiC at 70° angle of incidence and three different sample orientations rotated relative to the plane of incidence.[77]

The VUV optical properties of BeZnSe and BeTe were measured and then related to corresponding electronic transitions.[25–27] In addition, ZnSe has been studied as a potential infrared optical coating for use in low-earth orbit.[106] For this application, the film must be able to withstand a harsh environment, which was tested by oxygen plasma exposure. The film

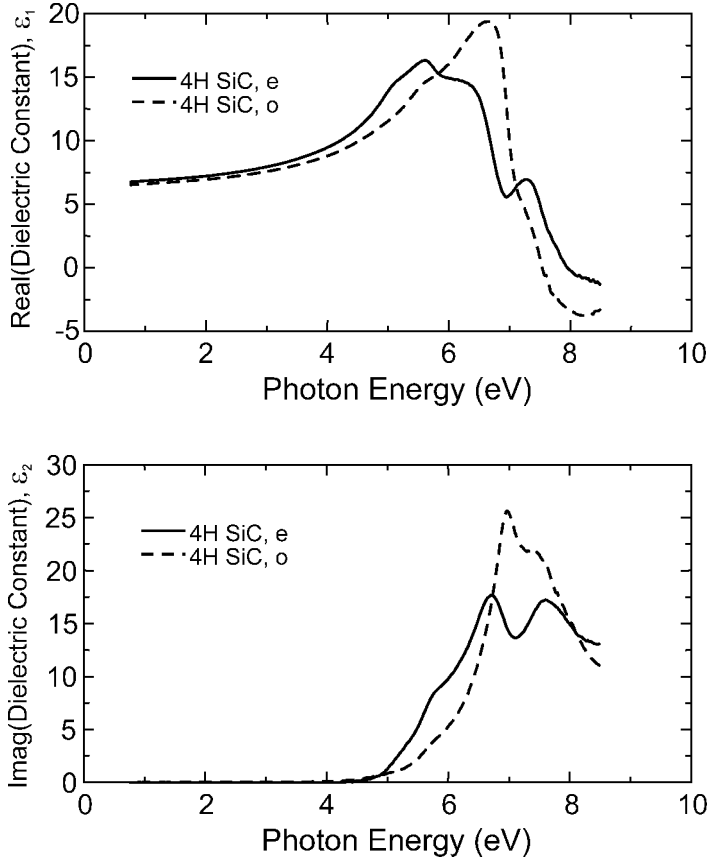


Figure 10.22 Real and imaginary parts of ε_o and ε_e versus photon energy for 4H-SiC.[75]

thickness decreased and VUV optical response dramatically changed, which was likely affected by significant surface roughening.[106] Silicone is another material of interest as a protective coating in low-earth orbit environment. The chemical modification and erosion of this type of material when exposed to atomic oxygen has been studied in laboratory exposures. VUV ellipsometry has been used to monitor the optical properties and surface modifications of CV-1144-O silicone films. The film modifications reduce the organic component and become more like SiO_2.[107]

Other semiconducting films, such as BGaAs,[108] GaNAs,[109] and GaNP[109,110] have recently been studied in the VUV. For all of these materials, the VUV region allows investigation of higher energy electronic transitions to provide insight into the optical behavior.

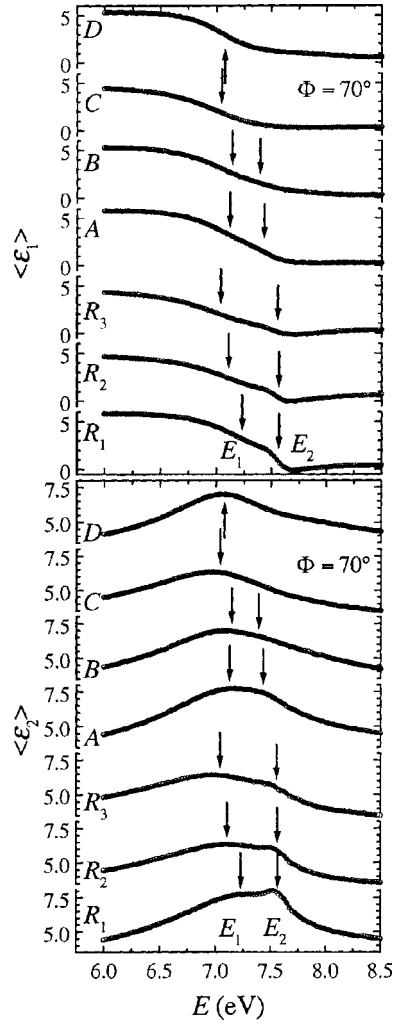


Figure 10.23 Pseudo-dielectric function measured for GaN and AlGaN to show effects of free-electron concentration (on the GaN samples) and alloy fraction (for AlGaN).[103]

Although most attention has been directed to dielectrics and semiconductors, VUV ellipsometry has also been used to study the semi-metal $TiSe_2$,[29] YBCO superconductors,[30] and a variety of metals.[31–33] Metal multilayers were studied to understand enhancement of the Magneto-Optic Kerr Effect (MOKE). Recent studies of extremely smooth Ir deposited layers have included VUV ellipsometry.[111]

10.6 Future of VUV Ellipsometry

Although VUV ellipsometry is applied to a wide variety of applications, a large majority of current use is directed toward immediate research requirements for 157 nm lithography. The lithography industry is expected to follow 157 nm with a Next Generation Lithography (NGL) technique. One possible solution for this next stage is EUV lithography.[112] Development of a commercial ellipsometer at these short wavelengths presents many new challenges, but may be important for EUV characterization. It is possible that optical lithography will be extended beyond 157 nm to a proposed laser-line around 126 nm. This would still present considerable challenges for commercial ellipsometry. Current systems are able to push measurements near 130 nm, but with greatly diminished signal-to-noise.

Use of current VUV ellipsometers may allow discovery of new applications and surpass the present dominance of VUV ellipsometry for 157 nm lithography. Strong absorption in the VUV by most materials should allow VUV ellipsometry to distinguish different materials that have similar refractive indices at longer wavelengths. In addition, the enhanced sensitivity of shorter wavelengths may lead to applications in thickness ranges requiring surface-sensitive investigations. It will be interesting to see the progression of this technique into new areas.

10.7 Acknowledgments

I am grateful for support and contributions from: John Bruning, Pierre Boher, Corey Bungay, Francis Celii, Christoph Cobet, N. V. Edwards, Craig Herzinger, Blaine Johs, Eric Joseph, Won Kim, Greg Pribil, Matthias Schubert, Ron Synowicki, Harland Tompkins, Tom Tiwald, John Woollam, M. Yamamoto, and Stefan Zollner.

10.8 References

1. J. A. R. Samson, *Techniques of Vacuum Ultraviolet Spectroscopy*, Wiley, New York, 2–3 (1967).
2. Berliner Elektronenspeicherring - Gesellschaft für Synchrotronstrahlung m. b. H., Albert-Einstein-Str. 15, 12489 Berlin, Germany.
3. High Energy Accelerator Research Organization (KEK), Photon Factory (PF), Oho 1-1, Tsukuba, Ibaraki, 305-0801, Japan.
4. M. Schledermann and M. Skibowski, *Appl. Opt*., **10,** 321 (1971).
5. J. Barth, R. L. Johnson, S. Logothetidis, and M. Cardona,, *SPIE Proceedings – Soft X-Ray Optics and Technology*, **733**, 265 (1986).

6. M. Yamamoto, K. Mayama, H. Kimura, Y. Goto, and M. Yanagihara, *Journal of Electron Spectroscopy and Related Phenomena*, **80**, 465 (1996).
7. R. L. Johnson, J. Barth, M. Cardona, D. Fuchs, and A. M. Bradshaw, *Rev. Sci. Instrum.*, **60**, 2209 (1989).
8. R. L. Johnson, J. Barth, M. Cardona, D. Fuchs, and A. M. Bradshaw, *Nuclear Instruments and Methods in Physics Research*, **A290**, 606 (1990).
9. J. Barth, R. L. Johnson, and M. Cardona, in *Handbook of Optical Constants of Solids II*, edited by E. D. Palik, Academic Press, San Diego, 213 (1991).
10. J. Barth, R. L. Johnson, M. Cardona, D. Fuchs, and A. M. Bradshaw, *Phys. Rev. B*, **41**, 3291 (1990).
11. D. Fuchs, R. Schlögl, A. M. Bradshaw, J. Barth, R. L. Johnson, and M. Cardona, *Synthetic Metals*, **34**, 417 (1989).
12. S. Logothetidis, J. Petalas, H. M. Polatoglou, and D. Fuchs, *Phys. Rev. B*, **46**, 4483 (1992).
13. S. Logothetidis, J. Petalas, and S. Ves, *J. Appl. Phys.*, **79**, 1039 (1996).
14. Y. Panayiotatos, S. Logothetidis, A. Laskarakis, A. Zervopoulou, and M. Gioti, *Diamond and Related Materials*, **11**, 1281 (2002).
15. S. Logothetidis, H. M. Polatoglou, J. Petalas, D. Fuchs, and R. L. Johnson, *Physica B*, **185**, 389 (1993).
16. S. Logothetidis and J. Petalas, *J. Appl. Phys.*, **80**, 1768 (1996).
17. C. Janowitz, J. Kalomiros, A. Ginoudi, E. C. Paloura, and R. L. Johnson, *Solid State Commun.*, **99**, 29 (1996).
18. O. Gunther, C. Janowitz, G. Jungk, B. Jenichen, R. Hey, L. Däweritz, and K. Ploog, *Phys. Rev. B*, **52**, 2599 (1995).
19. S. Logothetidis, J. Petalas, M. Cardona, and T. D. Moustakas, *Phys. Rev. B*, **50**, 18017 (1994).
20. S. Logothetidis, J. Petalas, M. Cardona, and T. D. Moustakas, *Materials Science and Engineering*, **B29**, 65 (1995).
21. J. Petalas, S. Logothetidis, S. Boultadakis, M. Alouani, and J. M. Wills, *Phys. Rev. B*, **52**, 8082 (1995).
22. T. Wethkamp, K. Wilmers, C. Cobet, N. Esser, W. Richter, O. Ambacher, M. Stutzmann, and M. Cardona, *Phys. Rev. B*, **59**, 1845 (1999).
23. L. X. Benedict, T. Wethkamp, K. Wilmers, C. Cobet, N. Esser, E. L. Shirley, W. Richter, and M. Cardona, *Solid State Commun.*, **112**, 129 (1999).
24. T. Wethkamp, K. Wilmers, N. Esser, W. Richter, O. Ambacher, H. Angerer, G. Jungk, R. L. Johnson, and M. Cardona, *Thin Solid Films*, **313–314**, 745 (1998).
25. K. Wilmers, T. Wethkamp, N. Esser, C. Cobet, W. Richter, M. Cardona, V. Wagner, H. Lugauer, F. Fischer, T. Gerhard, and M. Keim, *Phys. Rev. B*, **59**, 10071 (1999).
26. K. Wilmers, T. Wethkamp, N. Esser, C. Cobet, W. Richter, V. Wagner, A. Waag, H. Lugauer, F. Fischer, T. Gerhard, M. Keim, and M. Cardona, *Phys. Stat. Sol. B*, **215**, 15 (1999).
27. K. Wilmers, T. Wethkamp, N. Esser, C. Cobet, W. Richter, V. Wagner, H. Lugauer, F. Fischer, T. Gerhard, M. Keim, and M. Cardona, *Journal of Electronic Materials*, **28**, 670 (1999).
28. C. Janowitz, O. Gunther, G. Jungk, R. L. Johnson, P. V. Santos, M. Cardona, W. Faschinger, and H. Sitter, *Phys. Rev. B*, **50**, 2181 (1994).

29. T. Buslaps, R. L. Johnson, and G. Jungk, *Thin Solid Films*, **234**, 549 (1993).
30. J. Kircher, J. Humlicek, M. Garriga, M. Cardona, D. Fuchs, H. U. Habermeier, O. Jepsen, Sudha Gopalan, O. K. Anderson, Y. Fang, U. Welp, K. G. Vandervoort, and G. W. Crabtree, *Physica C*, **192**, 473 (1992).
31. S. Logothetidis, J. Petalas, N. K. Flevaris, and R. L. Johnson, *Thin Solid Films*, **234**, 538 (1993).
32. N. K. Flevaris, S. Logothetidis, J. Petalas, P. Kielar, M. Nyvlt, V. Parizek, S. Visnovsky, and R. Krishnan, *Journal of Magnetism and Magnetic Materials*, **121**, 479 (1993).
33. S. Logothetidis and N. K. Flevaris, *J. Appl. Phys.*, **75**, 7978 (1994).
34. J. Petalas, S. Logothetidis, A. Markwitz, E. C. Paloura, R. L. Johnson, and D. Fuchs, *Physica B*, **185**, 342 (1993).
35. S. Logothetidis, J. Petalas, A. Markwitz, and R. L. Johnson, *J. Appl. Phys.*, **73**, 8414 (1993).
36. J. Petalas and S. Logothetidis, *Phys. Rev. B*, **50**, 11801 (1994).
37. J. Petalas, S. Logothetidis, S. Boultadakis, and A. Markwitz, *J. Non-Cryst. Solids*, **187**, 291 (1995).
38. C. Cobet, K. Wilmers, T. Wethkamp, N. V. Edwards, N. Esser, and W. Richter, *Thin Solid Films*, **364**, 111 (2000).
39. M. Yamamoto and M. Furudate, *Thin Solid Films*, **313–314**, 751 (1998).
40. M. Yanagihara, T. Maehara, H. Nomura, M. Yamamoto, T. Namioka, and H. Kimura, *Rev. Sci. Instrum.*, **63**, 1516 (1992).
41. W. Hu, M. Yamamoto, and M. Watanabe, *Proc. SPIE* **2873**, 74 (1996).
42. H. Nomura, K. Mayama, T. Sasaki, M. Yamamoto, and M. Yanagihara, *Proc. SPIE*, **1720**, 395 (1992).
43. J. B. Kortright, H. Kimura, V. Nikitin, K. Mayama, M. Yamamoto, and M. Yanagihara., *Appl. Phys. Lett.*, **60**, 2963 (1992).
44. M. Yamamoto, K. Mayama, H. Kimura, M. Furudate, and M. Yanagihara, *Proc. SPIE*, **2873**, 70 (1996).
45. SOPRA-S.A., 26 Rue Pierre Joigneaux, 92270 Bois Colombes, France.
46. Sentech Instruments GmBH, Carl-Scheele-Str. 16, 12489 Berlin, Germany.
47. J. A. Woollam Co., Inc. 645 M Street, Suite 102, Lincoln, NE 68508 USA.
48. J. N. Hilfiker, B. Singh, R. A. Synowicki, and C. Bungay, *Proc. SPIE*, **3998**, 390 (2000).
49. P. Boher, J. P. Piel, P. Evrard, C. Defranoux, M. Espinosa, and J. L. Stehle, *Proc. SPIE*, **3998**, 379 (2000).
50. S. Peters, T. Schmidtling, T. Trepk, U. W. Pohl, J. T. Zettler, and W. Richter, *J. Appl. Phys.*, **88**, 4085 (2000).
51. *Some Aspects of Vacuum Ultraviolet Radiation Physics*, eds. N. Damany, J. Romand, and B. Vodar, Pergamon Press: Oxford, 42 (1974).
52. J. N. Hilfiker, C. Bungay, R. Synowicki, T. E. Tiwald, and J. A. Woollam, *Proc. 1st International Symp. on 157nm Lithography*, 787 (2000).
53. B. Johs, J. A. Woollam, C. M. Herzinger, J. N. Hilfiker, R. Synowicki, and C. Bungay, *Proc. SPIE*, **CR72**, 29 (1999).
54. US Patent # 5,757,494.
55. D. E. Aspnes, in *Handbook of Optical Constants of Solids*, edited by E. D. Palik, Academic Press, San Diego, 89–112 (1985).

56. S. Adachi, *Optical properties of crystalline and amorphous semiconductors: Materials and Fundamental Principles*, Kluwer Academic Publishers, Boston (1999).
57. P. M. Amirtharaj and D. G. Seiler, in *Handbook of Optics: Volume II Devices*, *Measurements*, *and Properties*, 2nd Ed. Edited by M. Bass, McGraw-Hill, New York (1995).
58. W. J. Tropf, M. E. Thomas, and T. J. Harris, in *Handbook of Optics: Volume II Devices*, *Measurements*, *and Properties*, 2nd Ed. Edited by M. Bass, McGraw-Hill, New York (1995).
59. *Handbook of Optical Constants of Solids*, Vol. 1–3, edited by E. D. Palik, Academic Press, San Diego, (1985, 1991).
60. P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors: Physics and Materials Properties*, 3rd Ed., Springer, Berlin, (2001).
61. C. F. Klingshirn, *Semiconductor Optics*, Springer-Verlag, Berlin (1995).
62. J. I. Pankove, *Optical Processes in Semiconductors*, Dover Publications, New York (1971).
63. W. R. Hunter, in *Handbook of Optical Constants of Solids*, edited by E. D. Palik, Academic Press, San Diego, 69 (1985).
64. J. N. Hilfiker, C. Bungay, R. Synowicki, T. E. Tiwald, and M. Liphardt, *Future Fab International*, **8**, 239 (2000).
65. T. Wagner, J. N. Hilfiker, T. E. Tiwald, C. Bungay, S. Zollner, *Phys. Stat. Sol. (a)*, **188**, 1553 (2001).
66. J. A. Woollam, B. Johs, C. M. Herzinger, J. N. Hilfiker, R. Synowicki, and C. Bungay, *Proc. SPIE*, **CR72**, 3 (1999).
67. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light*, Elsevier, Amsterdam, 283–293 (1987).
68. C. M. Herzinger, B. Johs, W. A. McGahan, J. A. Woollam, and W. Paulson, *J. Appl. Phys.*, **83**, 3323 (1998).
69. N. V. Edwards, private communication.
70. G. E. Jellison, Jr. and F. A. Modine, *Appl. Phys. Lett.*, **69**, 371 (1996), Erratum, *Appl. Phys. Lett.*, **69**, 2137 (1996).
71. N. V. Edwards, J. Vella, Q. Xie, S. Zollner, D. Werho, I. Adhihetty, R. Liu, T. E. Tiwald, C. Russell, J. Vires, and K. H. Junker. *Mat. Res. Soc. Symp. Proc.*, **697**, P4.7.1 (2002).
72. R. H. French, R. Wheland, D. J. Jones, J. N. Hilfiker, R. Synowicki, R. Zumsteg, J. Feldman, and A. Feiring, *Proc. SPIE*, **4000**, 1491 (2000).
73. J. H. Burnett, Z. H. Levine, and E. L. Shirley, *Phys. Rev. B*, **64**, 241102/1 (2001).
74. J. H. Burnett, Z. H. Levine, and E. L. Shirley, *Future Fab International*, **12**, 149 (2001).
75. J. N. Hilfiker and M. Suzuki, *Optical and Electro-Optical Engineering Contact*, **40**, 225 (2002).
76. J. N. Hilfiker, presented at *Proc. 157nm Technical Data Review* Dallas, Texas (2002).
77. J. A. Woollam, J. N. Hilfiker, T. E. Tiwald, C. Bungay, R. Synowicki, D. Meyer, C. M. Herzinger, G. Pfeiffer, G. Cooney, and S. Green, *Proc. SPIE*, **4099**, 197 (2000).

78. N. V. Edwards, O. P. A. Lindquist, L. D. Madsen, S. Zollner, K. Järrehdahl, C. Cobet, S. Peters, N. Esser, A. Konkar, and D. E. Aspnes, *Mat. Res. Soc. Symp. Proc.*, **693**, I8.3.1 (2002).
79. O. P. A. Lindquist, K. Järrehdahl, S. Peters, J. T. Zettler, C. Cobet, N. Esser, D. E. Aspnes, A. Henry, and N. V. Edwards, *Appl. Phys. Lett.*, **78**, 2715 (2001).
80. T. E. Tiwald and M. Schubert, *Proc. SPIE*, **4103**, 19 (2000).
81. C. Brodsky, J. Byers, W. Conley, R. Hung, S. Yamada, K. Patterson, M. Somervell, B. Trinque, H. V. Tran, S. Cho, T. Chiba, S.-H. Lin, A. Jamieson, H. Johnson, T. Vander Heyden, and C. G. Willson, *J. Vac. Sci. Technol. B*, **18**, 3396 (2000).
82. J. A. Woollam, J. N. Hilfiker, C. Bungay, R. Synowicki, T. E. Tiwald, and D. W. Thompson, *Proc. AIP - Characterization and Metrology for ULSI Technology*, **CP550**, 511 (2001).
83. J. P. Piel, P. Boher, P. Evrard, and J. L. Stehle, *Proc. AIP - Characterization and Metrology for ULSI Technology*, **CP550**, 543 (2001).
84. W. D. Kim, H. S. Kim, P. Gabella, J. Byers, D. A. Miller, M. Daniels, B. Birmingham, J. Tompkins, O. Aoki, J. N. Hilfiker, and T. E. Tiwald, *Proc. 1st International Symp. on 157nm Lithography*, 1121 (2000).
85. J. N. Hilfiker, C. Bungay, R. Synowicki, T. E. Tiwald, C. M. Herzinger, B. Johs, G. K. Pribil, and J. A. Woollam, *J. Vac. Sci. Technol. A*, **21**, 1103 (2003).
86. W. D. Kim, D. A. Miller, H. S. Kim, J. Byers, M. Daniels, B. Birmingham, and J. Tompkins, *Proc. SPIE*, **4226**, 93 (2000).
87. J. N. Hilfiker, F. Celii, W. Kim, E. A. Joseph, C. Gross, T. Y. Tsui, R. Willecke, J. Large, and D. Miller, *Semiconductor Fabtech*, **17**, 87 (2002).
88. E. Joseph, (in process).
89. P. F. Carcia, G. Hughes, R. H. French, G. Reynolds, C. C. Torardi, and L. Dieu, *Vacuum & ThinFilm*, **Sept**, 14 (1999).
90. F. D. Kalk, R. H. French, H. U. Alpay, and G. Hughes, *Proc. SPIE*, **2254**, 64 (1994).
91. B. Johs, R. H. French, F. D. Kalk, W. A. McGahan, and J. A. Woollam, *Proc. SPIE*, **2253**, 1098 (1994).
92. P. Boher, *Proc. 1st International Symp. on 157nm Lithography*, 759 (2000).
93. R. H. French, J. Gordon, D. J. Jones, M. F. Lemon, R. C. Wheland, E. Zhang, F. C. Zumsteg, K. G. Sharp, and W. Qiu, *Proc. SPIE*, **4346**, 89 (2001).
94. B. von Blanckenhagen, D. Tonova, and J. Ullman, *Appl. Opt.*, **41**, 3137 (2002).
95. E. Franke, M. Schubert, C. L. Trimble, M. J. DeVries, and J. A. Woollam, *Thin Solid Films*, **388**, 283 (2001).
96. E. Franke, C. L. Trimble, M. J. DeVries, J. A. Woollam, M. Schubert, and F. Frost, *J. Appl. Phys.*, **88**, 5166 (2000).
97. P. Boher, P. Evrard, J. P. Piel, and J. L. Stehle, *J. Non-Cryst. Solids*, **303**, 167 (2002).
98. P. Boher, P. Evrard, J. P. Piel, C. Defranoux, and J. L. Stehle, *Mat. Res. Soc. Symp.*, **636**, D9.2.1 (2001).

99. P. Boher, P. Evrard, J. P. Piel, and J. L. Stehle, *Proc. SPIE*, **4449**, 30 (2001).
100. S. Zollner, A. A. Demkov, R. Liu, P. L. Fejes, R. B. Gregory, J. A. Curless, Z. Yu, J. Ramdani, R. Droopad, T. E. Tiwald, J. N. Hilfiker, and J. A. Woollam, *J. Vac. Sci. Technol. B*, **18**, 2242 (2000).
101. M. Schubert, B. Rheinländer, E. Franke, H. Neumann, J. Hahn, M. Röder, and F. Richter, *Appl. Phys. Lett.*, **70** 1819 (1997).
102. M. P. Thompson, G. W. Auner, C. Huang, and J. N. Hilfiker, *Mat. Res. Soc. Symp. Proc.*, **622**, T6.5.1 (2001).
103. A. Kasic, M. Schubert, T. Frey, U. Köhler, D. J. As, C. M. Herzinger, *Phys. Rev. B*, **65**, 184302/1 (2002).
104. C. Cobet, N. Esser, J. T. Zettler, W. Richter, P. Waltereit, O. Brandt, K. H. Ploog, S. Peters, N. V. Edwards, O. P. A. Lindquist, and M. Cardona, *Phys. Rev. B*, **64**, 165203-1 (2001).
105. A. Kasic, M. Schubert, B. Rheinländer, J. Off, F. Scholz, and C. M. Herzinger, *Mat. Res. Soc. Symp.*, **639**, G6.13.1 (2001).
106. L. Yan, J. A. Woollam, and E. Franke, *J. Vac. Sci. Technol. A*, **20**, 693 (2002).
107. L. Yan, X. Gao, C. Bungay, and J. A. Woollam, *J. Vac. Sci. Technol. A*, **19**, 447 (2001).
108. G. Leibiger, V. Gottschalch, V. Riede, M. Schubert, J. N. Hilfiker, and T. E. Tiwald, *Phys. Rev. B*, (2002) submitted.
109. G. Leibiger, V. Gottschalch, R. Schwabe, G. Benndorf, and M. Schubert, *Phys. Stat. Sol. (b)*, **228**, 279 (2001).
110. G. Leibiger, V. Gottschalch, M. Schubert, G. Benndorf, R. Schwabe, *Phys. Rev. B*, **65**, 245207/1 (2002).
111. L. Yan and J. A. Woollam, *J. Appl. Phys.*, **92**, 4386 (2002).
112. C. Gwyn, *OEmagazine*, **June**, 22 (2002).

11 Spectroscopic Infrared Ellipsometry

Arnulf Röseler

Institut für Spektrochemie und Angewandte Spektroskopie
Berlin, Germany

11.1 Experimental Tools

Spectroscopic infrared ellipsometry is a method that combines a Fourier transform spectrometer with a photometric ellipsometer. In principle, there are two different approaches:

- the photometric ellipsometer utilizing a rotating analyzer and fixed polarizer or reversed[1].
- the photometric ellipsometer applying a photoelastic modulator (PEM) to introduce a time dependent phase difference.[2]

The first approach minimizes the influence of the imperfections of optical components on the quality of the measurement. In the course of this contribution we focus on the infrared region from approximately 5000 to 400 cm^{-1}, also called mid infrared. The wavenumber range is generally determined by the spectral response of the optical components, such as the beam splitter, the detector, the polarizers, and the retarder. The second kind of infrared ellipsometry has the advantage of fast measurements for relative changes, but is influenced by the real properties of the PEM and the amplifiers for the ground frequency and the second harmonic of the modulated retarder. Therefore, it is problematic whether an absolute accuracy can be obtained in determining the ellipsometric parameters Ψ and Δ.

11.1.1 Two Kinds of Instruments

11.1.1.1 Infrared Photometric Ellipsometry

Infrared spectroscopy is energy limited by the radiation energy of the glowbar, which has a temperature of about 1600 K. To overcome this problem Fourier transform spectrometers (FTS) have been used in recent

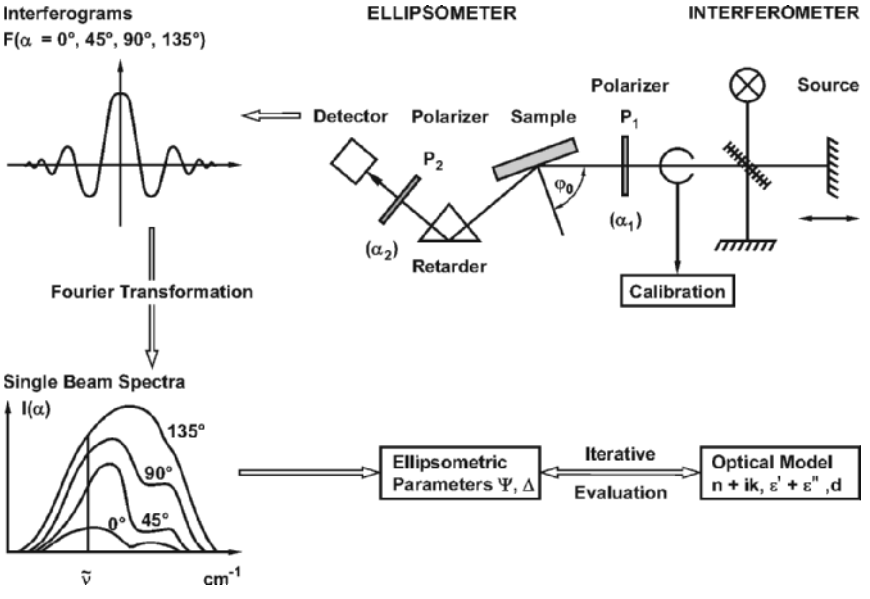


Figure 11.1 Optical layout and principle of infrared ellipsometry.

designs.[3, 4] They have the advantage[5, 6] of a high throughput combined with measuring of all spectral elements during a whole scan cycle. Infrared ellipsometry is only practicable in combination with an FTS. The sequence of the devices FTS-PE (PE – photometric ellipsometer) is advantageous, because for an opposite order the measurement would be disturbed by the thermal radiation of the sample.

The principle of the photometric measurement is shown in Fig. 11.1. The observed intensity at the detector gives the Fourier transform of the spectrum, depending on the optical path difference of the interferometer. From the Fourier transformation, we get the spectrum as function of the wavenumber. This intensity, written in terms of the Stokes parameters (s_0, s_1, s_2, s_3) and the polarizer azimuth (α_1, α_2), is[5]:

$$I = \frac{1}{2}(s_0 + s_1 \cos 2\alpha_2 + s_2 \sin 2\alpha_2). \tag{11.1}$$

This equation can be rewritten in terms of the ellipsometric parameters Ψ and Δ:

$$I = \frac{1}{2} s_0 (1 - \cos 2\Psi' \cos 2\alpha_2 + \sin 2\Psi' \sin 2\alpha_2) \tag{11.2a}$$

with

$$\tan\Psi' = \tan\Psi/\tan\alpha_1 \tag{11.2b}$$

and

$$\cos\Psi' = -s_1/s_0, \quad \sin 2\Psi'\cos\Delta = s_2/s_0. \tag{11.2c}$$

A rotating analyzer is generally used to determine the Stokes parameters in the ultraviolet-visible (UV-vis) range. Using FTS for measurements in the infrared requires constant radiation power during the scan. Therefore the analyzer must rotate in a scan and step manner. In other words, the analyzer is fixed at a position while the FTS scans the spectrum. The analyzer then moves to the next position and the FTS scans again. Another feature of the FTS that must be taken into consideration is that it generates partially polarized radiation. This is caused by the properties of the beam splitter in the interferometer and the mirrors of the instrument.

Four positions of the azimuth of the polarizer (0°, 45°, 90°, 135°) are sufficient to derive the Stokes parameters. For these four or even more azimuths (if used), calibration measurements are necessary with no sample in the ellipsometer to eliminate the influence of the polarizing properties of the FTS. The intensities measured with the sample in the optical path of the ellipsometer related to the calibration intensities are then used to calculate the Stokes parameters. With a fixed polarizer we have no influence of the FTS, but the detector and its optics can introduce artifacts in combination with the rotating analyzer.

In this arrangement, Δ is derived only from the cos function with the disadvantage of a reduced accuracy in the neighborhood of $|\cos(\Delta)| \approx 1$ and is unambiguous only in the range from 0° to 180°. This problem can be overcome using a retarder between the two polarizers of the ellipsometer.[7] Practically, retardation between ≈30° and ≈150° is sufficient, but a value of approximately 90° would be the best. In $\cos(\Delta)$, Δ must be replaced by $\Delta + \delta$ with the retardation δ of the retarder. Measurements with and without the retarder are necessary to get $\cos(\Delta)$ and $\sin(\Delta)$ independently. The retardation δ is measured with the retarder as the sample in the optical path. With known Δ and δ for $\sin(\Delta)$ we get:

$$\sin(\Delta) = \frac{\cos(\Delta)\cos(\delta) - \cos(\Delta + \delta)}{\sin(\delta)} \tag{11.3}$$

With the measured $\sin(\Delta)$ the fourth Stokes parameter can be calculated:

$$s_3 = \sin(2\Psi)\sin(\Delta) \tag{11.4}$$

11.1.1.2 Phase Modulated Ellipsometry with the Fourier Transform Spectrometer

For real time applications in the infrared ellipsometry a PEM is used, which introduces a periodic phase shift. This technique was developed by Drevillon and co-workers.[2, 8] The principal arrangement consists of a sequence FTS, polarizer, PEM, sample, analyzer and detector. An FTS has an interference modulation frequency, generated by the moving mirror, of some kHz. This frequency is much lower than the resonance frequency of the PEM, which is in the order of 40 kHz. Therefore the two signals are not overlapping. From the point of view of the PEM frequency a mercury cadmium telluride (MCT) detector must be used. This is nonlinear in its response and is in this application restricted to wavenumbers greater than approximately 700 cm^{-1}.

The arrangement is that of Fig. 11.2 with the notation P, M, and A for the azimuth angle of the polarizer, modulator, and analyzer, measuring the orientation clockwise, if looking from the detector to the source. Two measurement modes are usual with the plane of incidence as reference:

- mode A: $P\text{-}M = \pm 45°$, $M = 0°$, $(90°)$, $A = \pm 45°$
- mode B: $P\text{-}M = \pm 45°$, $M = \pm 45°$, $A = \pm 45°$.

The measured spectrum I in mode A, calculated by the Stokes-Mueller formalism for the settings: $P = 45°$, $M = 0°$ and $A = 45°$, is:

$$I = \frac{s_0}{2}\left(1 + \sin 2\Psi(\cos\Delta \cos\delta(t) + \sin\Delta \sin\delta(t))\right) \qquad (11.5)$$

with s_0 = first Stokes parameter, $\delta = a \sin \omega t$, a = amplitude, ω = frequency of the PEM.

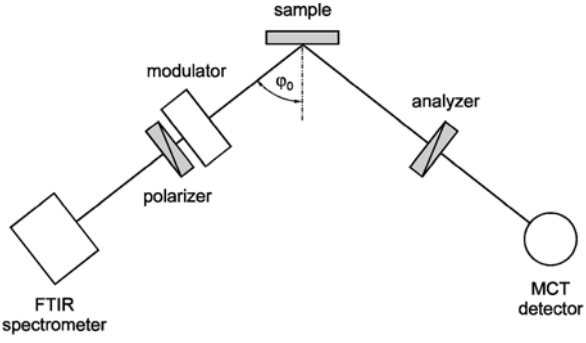


Figure 11.2 Outline of the phase modulated ellipsometer after Drevillon.[2]

The expansion of $\cos\delta$ and $\sin\delta$ in Fourier series results in:

$$\sin\delta = 2\sum_{m=0}^{\infty} J_{2m+1}(a)\sin((2m+1)\omega t) \tag{11.6a}$$

and

$$\cos\delta = J_0(a) + 2\sum_{m=1}^{\infty} J_{2m}(a)\cos(2m\omega t), \tag{11.6b}$$

where $J_k(a)$ are the Bessel functions of the argument a and the order k.

With Eq. (11.6a, b) we rewrite Eq. (11.5), neglecting contributions of higher order:

$$I = \frac{s_0}{2}(1 + J_0(a)\sin 2\Psi\cos\Delta + 2J_1(a)T_1\sin 2\Psi\sin\Delta\sin\omega t + 2J_2(a)T_2\sin 2\Psi\cos\Delta\cos 2\omega t)$$

and for the relation of the direct signal to the fundamental R_ω and second harmonic $R_{2\omega}$ we get:

$$R_\omega = \frac{2T_1J_1(a)\sin 2\Psi\sin\Delta}{1 + J_0(a)\sin 2\Psi\cos\Delta} \tag{11.7a}$$

$$R_{2\omega} = \frac{2T_2J_2(a)\sin 2\Psi\cos\Delta}{1 + J_0(a)\sin 2\Psi\cos\Delta} \tag{11.7b}$$

T_1, T_2 = response of the detection system for the frequency ω and 2ω, respectively.

In mode B with the settings $P = 0°$, $M = 45°$, $A = 45°$ the spectrum I is obtained by:

$$I = \frac{s_0}{2}(1 - \cos 2\Psi\cos\delta - \sin 2\Psi\sin\Delta\sin\delta) \tag{11.8}$$

Similar to Eq. (11.7) for the second harmonic the quotient $R_{2\omega}$ is:

$$R_\omega = \frac{-2T_1J_1(a)\sin 2\Psi\sin\Delta}{1 - J_0(a)\cos 2\Psi} \tag{11.9a}$$

$$R_{2\omega} = \frac{-2T_2J_2(a)\cos 2\Psi}{1 - J_0(a)\cos 2\Psi}. \tag{11.9b}$$

From Eq. (11.7a, b) and/or (11.9a, b) the ellipsometric angles Ψ and Δ can be easily obtained.

11.1.2 Optical Equipment for the Infrared-Ellipsometry

11.1.2.1 Polarizer of Grating Type

In the infrared only polarizers from the grating- and Brewster angle type are of interest. Polarizers on the base of birefringence as used in the visible and UV regions are not applicable in the infrared.

A metal grating on an isotropic substrate can be used for the generation of linearly polarized radiation. The first application was found by H. Hertz (1888) with a freestanding wire grid to demonstrate the transversal nature of the electromagnetic waves. In the infrared the polarizer consists of parallel metal strips evaporated on a transparent substrate. This arrangement is conducting parallel to the strips and acts as a mirror for radiation polarized parallel to the strips. In this direction, a minimum of light is transmitted. For a polarization perpendicular to the strips, we have a surface like an insulator and, therefore, the radiation is transmitted. The metal grid also acts as a diffraction grating. For a polarizer the zero diffraction order is observed and the whole intensity is found in this order for a suitable distance of the metal strips in relation to the wavelength. The plus/minus first order has a diffraction angle greater then 90° and is, therefore, imaginary. Sometimes these orders can be observed using visible laser light. The degree of polarization depends on the distance between the strips. As an orientation: a spacing of $\lambda/10$ or better for the metal strips of the shortest wavelength, which is to be measured, is recommended. For instance polarizers for the region from 5000 to 250 cm^{-1} are manufactured with approximately 3000 lines/mm on a KRS5 substrate and for the long wavelength range, lower than $\approx$700 cm^{-1}, polyethylene foils with 1200 lines/mm are used. Such a foil of $\approx$60 µm thickness produces interferences, which can disturb the measurement. By the calibration procedure of the ellipsometer these interferences can be cancelled, but if the cross section and the opening angle at the sample position are different for the calibration and measurement, disturbing residue interferences are observed. The polyethylene substrate has absorption bands at the CH infrared band positions in the neighborhood of 700, 1400, and 3000 cm^{-1}. It is not possible to investigate organic substances with CH bands in the spectrum using these polarizers.

The wire-grid polarizers are compact devices and easy to handle, but in the application an important feature must be regarded. The substrate can have birefringence by imperfections of the crystal or by stress of the substrate, caused by imperfections of the holder, which introduces an additional phase shift in the sample path. Therefore the polarizers must be arranged in such a manner that the metal grids are oriented in direction to

the sample. A disadvantage of the wire-grid polarizers is the wavenumber dependence on the degree of polarization. Under certain limitations a correction procedure for the calculation of the ellipsometric parameters is possible.[5, 9]

11.1.2.2 Polarizer of the Brewster-Angle Type

The reflection of electromagnetic radiation at the Brewster angle φ_B of a medium with the refractive index n is a well-known optical feature to generate linearly polarized light. At this angle the reflectivity for the parallel component of the polarized radiation is zero and is completely transmitted. The condition for transmission is:

$$\tan \varphi_B = n$$

The reflectivity of the perpendicular component is non-zero and given by Fresnel's equation. From this fact two arrangements are possible:

- one or more reflections for the perpendicular component
- a pile of transmitting sheets transmitting the parallel component.

The ratio of the intensities transmitted in the perpendicular (T_s) and parallel component (T_p) is[10]

$$\frac{T_s}{T_p} = \sin^2 2\varphi_B = \left(\frac{2n}{1 + n^2} \right)^2. \tag{11.10a}$$

We denote the degree of polarization as P (not to be confused with the azimuthal angle of the polarizer, also denoted by the same symbol). For a pile of m thin plates with two surfaces each, the degree of polarization, P, for the resulting linearly polarized radiation is given by

$$P = \frac{|\sin^{4m} 2\varphi_B - 1|}{\sin^{4m} 2\varphi_B + 1}. \tag{11.10b}$$

A stack of two plates of germanium results in $P = 0.995$ and five films of selenium results in $P = 0.997$. The Brewster angle polarizer is effective with a medium of high refractive index and, therefore, a great angle of incidence.[11] This results in great construction length and a low acceptance angle. The reflection can also be used, but the intensity is reduced by the reflectivity of the medium. A constant degree of polarization is the advantage.

11.1.2.3 Retarder (Prism Retarder)

The infrared covers a wide wavenumber range, which demands a retarder operating in the region of about 1:10 (i.e., 4000–400 cm^{-1}). Therefore, prisms and thin film retarders are used. Retarders consisting of birefringence material, as in the visible and UV, are in the infrared restricted to the near infrared range.

For the prism retarder different optical layouts are known, for example the Fresnel rhomb. They are used in the visible because the material (glass, quartz) has a relatively low refractive index and two reflections are necessary to reach approximately 90° phase difference. These retarders have the disadvantage of a great construction length and sometimes a beam displacement occurs. In the infrared materials with higher refractive indices are available (ZnSe, KRS5, CsI, Si, Ge) and the phase difference of ≈90° can be reached with only one reflection. The retardation of a prism depends on the refractive index and the angle of incidence on the prism base in the total reflection mode. The reflectivity for both polarization directions is equal to the unit, but intensity losses are introduced by the transmission of the prism surfaces. The angle of incidence must be greater than the critical angle (φ_c) according to $\sin\varphi_c = n_0/n_1$ where n_0 and n_1 are the refractive indices of the ambient and the prism material. The phase difference δ is then calculated by[10]

$$\tan\frac{\delta}{2} = -\frac{\cos\varphi_0}{\sin^2\varphi_0}\sqrt{\sin^2\varphi_0 - (n_0/n_1)^2} \tag{11.11a}$$

with $\varphi_0 \geq \varphi_c$ and $\delta = \delta_p - \delta_s$.

The maximum possible phase difference is:

$$\tan\frac{\delta_{\max}}{2} = \frac{n_1^2 - n_0^2}{2n_1 n_0} \tag{11.11b}$$

and in this case the angle of incidence $\varphi_{\max}$ is given by

$$\sin^2\varphi_{\max} = \frac{2n_0^2}{n_1^2 + n_0^2}. \tag{11.11c}$$

In the application of an infrared retarder some imperfection must be considered. Scattering inside the prism, which results in a decrease of the degree of polarization after passing the prism, and birefringence caused by the imperfections of the material or by stress introduced by the prism holder are observed. The scattering in the volume of the prism is mostly

constant, whereas the birefringence depends on the ray tracing inside the prism. These features are characteristic for each individual prism and must be determined by an ellipsometric measurement. As a consequence of these properties, the retarder prism, used in a photometric ellipsometer, must be measured in the same device and same measurement position to get the $\tan\Psi$ and Δ spectra of the retarder for evaluating the four Stokes parameters. KRS5 is a mixed crystal from TlI and TlBr, which results in a strong influence of the former properties. Here we find great differences in the quality from crystal to crystal. The great range of transmission from NIR to about 250 cm^{-1} makes this material very suitable for routine measurements. When using a silicon prism for FIR investigations, float-zone material is recommended to minimize the scattering. A good contact to the manufacturer in order to obtain the optimal material is essential. Figure 11.3a shows different degrees of phase polarization for two different KRS5 prisms. KBr is a suitable standard material to characterize other materials, because it has a low scattering and birefringence. A completely different way of characterizing the scattering of a prism is treated in connection with the reststrahlen bands of strong oscillators.

The refractive index spectrum calculated from the directly measured $\tan\Psi$ and Δ spectra of the prism normally differs from the well-known values in the literature, caused by the imperfections. The birefringence can also lead to values of $\tan\Psi \neq 1$. It is recommended that one use only measured data for calculations and correction procedures.

11.1.3 The Degree of Polarization

For the calculation of the ellipsometric parameters, totally polarized light is assumed. That means that the Stokes parameters fulfill the relation:

$$s_0^2 = s_1^2 + s_2^2 + s_3^2 \tag{11.12}$$

In many practical cases, s_0 contains a part of unpolarized radiation and is an addition of: $s_0 = s_{0u} + s_{0p}$, with s_{0u} for the unpolarized fraction and s_{0p} for the polarized part. Therefore, the total degree of polarization is given by

$$P = \frac{s_{0p}}{s_0} = \frac{\sqrt{s_1^2 + s_2^2 + s_3^2}}{s_0} \leq 1 \tag{11.13}$$

Equation (11.13) can be rewritten in terms of Ψ and Δ to get a feeling for the meaning of these values for the degree of polarization.

$$P = \sqrt{\cos^2 2\Psi + \sin^2 2\Psi(\cos^2\Delta + \sin^2\Delta)}. \tag{11.14}$$

In principle, $P = 1$, but in practice the measured Stokes parameters are averaged values as are the angle functions of Ψ and Δ with the possible consequence of $P \leq 1$.[5, 12, 13] The reasons for averaging are different.

Some possibilities are:

- averaging over the acceptance angle of the ellipsometer
- averaging over the spectral range for one sampling point (limit of resolution)
- averaging over the optical thickness of the film across the sample
- scattering inside the sample or at the surface
- reflection from rear surface of a transparent substrate (e.g., silicon wafer)
- other imperfections.

The term $(\cos^2\Delta + \sin^2\Delta)$ is of special interest. By the averaging procedure, it can deviate from unity and should be defined as the "degree of phase polarization"

$$P_{Ph} = \sqrt{\overline{\cos\Delta}^2 + \overline{\sin\Delta}^2}, \text{ with } 0 \leq P_{Ph} \leq 1. \tag{11.15}$$

Without any correlation between the phases in the parallel and perpendicular direction we get $P_{Ph} = 0$. For example, this is the case for the reflection of unpolarized light on a surface. From Eq. (11.14) follows

$$P = |\cos 2\Psi| = \frac{1 - \tan^2\Psi}{1 + \tan^2\Psi} = \frac{|I_p - I_s|}{I_p + I_s} \tag{11.16}$$

which is the same as the simple degree of polarization without regarding the phase contributions.

The averaging of the Stokes parameters with respect to the phase Δ and the angle Ψ is, if all existing differences Δ and Ψ have the same weight, determined by:

$$\overline{S_1} = \frac{-s_0}{(\Psi_2 - \Psi_1)} \int_{\Psi_1}^{\Psi_2} \cos 2\Psi d\Psi = -s_0 \cos 2\overline{\Psi} \operatorname{sinc} \partial\Psi = -s_0 \overline{\cos 2\Psi}, \tag{11.17a}$$

with $\operatorname{sinc} x = \dfrac{\sin x}{x}$

$$\overline{S_2} = \frac{s_0}{(\Psi_2 - \Psi_1)(\Delta_2 - \Delta_1)} \int_{\Psi_1}^{\Psi_2} \sin 2\Psi d\Psi \int_{\Delta_1}^{\Delta_2} \cos\Delta d\Delta$$

$$\overline{S_2} = s_0(\sin 2\overline{\Psi}\,\mathrm{sinc}\,\partial\Psi)(\cos\overline{\Delta}\,\mathrm{sinc}(\partial\Delta/2)) \tag{11.17b}$$

$$\overline{S_2} = s_0\,\overline{\sin 2\Psi} * \overline{\cos\Delta}$$

$$\overline{S_3} = s_0\,\overline{\sin 2\Psi} * \overline{\sin\Delta} \tag{11.17c}$$

with:

$$\overline{\Psi} = \frac{\Psi_2 + \Psi_1}{2};\ \overline{\Delta} = \frac{\Delta_2 + \Delta_1}{2};\ \partial\Psi = \Psi_2 - \Psi_1;\ \partial\Delta = \Delta_2 - \Delta_1$$

We can rewrite the degree of polarization P and P_{Ph} in terms of Ψ and Δ:

$$P = \left[\cos^2 2\overline{\Psi} + \sin^2 2\overline{\Psi}\,\mathrm{sinc}^2\left(\frac{\partial\Delta}{2}\right)\right]^{1/2}\mathrm{sinc}\,\partial\Psi \tag{11.18a}$$

and

$$P_{Ph} = \mathrm{sinc}\left(\frac{\partial\Delta}{2}\right) \tag{11.18b}$$

Since $\partial\Psi \leq 90°$, for Eq. (11.18a), sinc $\partial\Psi \geq 0.64$. For P_{Ph}, the situation is different because Δ can be much greater than 2π and the sincfunction approaches zero which results in $0 \leq P_{Ph} \leq 1$.

Using a retarder, Δ is determined in the range $0 \leq \Delta \leq 360°$ by the Stokes parameters s_2 and s_3 in Eq. (11.17a–c). In Eq. (11.17a) and Eq. (11.18a) sinc $\partial\Psi$ can be neglected because of $\partial\Psi \leq \overline{\Psi}$. The degree of the phase polarization can now be expressed in terms of the Stokes parameters:

$$P_{Ph} = \sqrt{\frac{\overline{s_2}^2 + \overline{s_3}^2}{s_0^2 - \overline{s_1}^2}}\,. \tag{11.19}$$

The calculation of Δ is not influenced by the sinc functions in Ψ and Δ Eq. (11.17b, c)

$$\tan\overline{\Delta} = \frac{\overline{s_3}}{\overline{s_2}} = \frac{\sin\overline{\Delta}}{\cos\overline{\Delta}}. \tag{11.20}$$

The polar angle of Δ is determined by the signs of $\sin\Delta$ and $\cos\Delta$.

P_{Ph} is identical with the phase of the coherence function after Born and Wolf[10]. The degree of the phase polarization is a very useful tool for checking the quality of measurement and the properties of the sample. Figure 11.3a–c shows typical P_{Ph}-spectra in context to their Ψ, Δ or n, k-spectra.

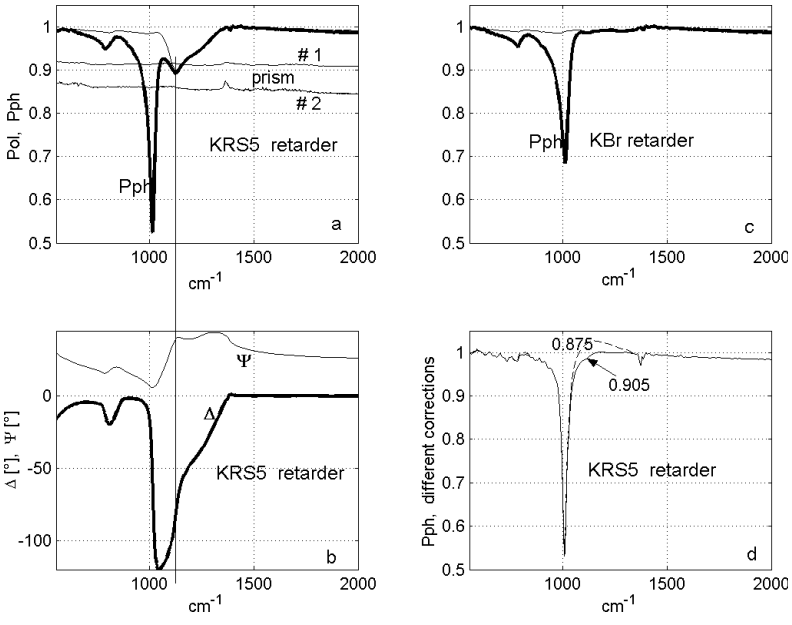


Figure 11.3 Artifacts in the ellipsometric measurements due to scattering inside the retarder prism shown for a quartz glass sample. The thin line in a) and c) is the total degree of the polarization. The thick line is the degree of the phase polarization. b) shows the Ψ and Δ spectra of quartz glass for comparison. d) shows the degree of the phase polarization for different scattering corrections.

In Fig. 11.3a, b, a typical example for the interpretation of P and P_{Ph}, for the measured polarization spectra of quartz glass is shown in combination with the corresponding Ψ, Δ spectra. The spectrum of P_{Ph} contains two minima in the range of the reststrahlen band. The first minimum near 1000 cm^{-1} is caused by a phase jump, which is connected with the Brewster angle for the angle of incidence $\varphi_0 = 70°$, corresponding to $n_B = 2.75$. The wavenumber position of the phase jump shifts, since the measurement was performed with an angle of acceptance of ±4° (compare the refractive index spectrum in Fig. 11.14d). The averaging across Δ results in a minimum in agreement with Eq. (11.18b). In the spectrum of the total degree of polarization, this minimum cannot be observed. Ψ is much smaller than 45° and therefore $\sin^2 2\Psi$ is approximately zero and we get from Eq. (11.18a) $P \approx 1$.

The second minimum near 1100 cm^{-1} is an artifact of the scattering inside of the KRS5 prism. The measured phase shift of the retarder prism is systematically changed by the portion of the unpolarized scattered radiation. In the neighborhood of $\Delta = 90°$ the effect of a systematic error in

Δ is very strong and results in a minimum of the phase polarization. The measured degree of phase polarization of two different KRS5 retarders is shown in Fig. 11.3a for comparison.

The same measurement of quartz glass, performed with a KBr retarder has only the first, but not the second minimum, which is connected with the scattering (Fig. 11.3c).

The second minimum in Fig. 11.3a can be cancelled by a correction procedure, regarding the measured degree of phase polarization (Fig. 11.3d). This example shows the very sensitive effect of slightly different correction values and is therefore a useful independent control of the measured P_{Ph} value of the retarder prism.

11.1.4 Linearity of the Detection System

The linearity of the detection system of the Fourier transform spectrometer is very important. The radiation modulated by the moving mirror in the interferometer by a frequency, depends on the wavenumber. Nonlinearity in the detection generates second harmonics, which, after Fourier transformation, appear as false intensities at twice the wavenumber. In ellipsometry, ratios of different polarization azimuths are measured with sometime very different intensities. The ellipsometric parameters calculated by means of these ratios are then affected by systematic errors. The nonlinearity depends on the type of detector used in the measurement. Detectors based on semiconductors are nonlinear, but the standard pyroelectric detector DTGS has good linearity and can be used as a reference. No linearity is very well seen in the degree of the phase polarization, if this P_{Ph} is compared with the corresponding spectrum measured with a DTGS. The MCT detector used in the high performance FTS is nonlinear. This effect can be reduced by using the MCT in the photovoltaic mode or by software correction.[14] Both procedures are connected with a residue of nonlinearity, which cannot be cancelled.

In ellipsometry, we try to find a compromise, keeping in mind the aim of the measurement. For high absolute determination of Ψ and Δ, a linear detector system is necessary, whereas relative measurements of small changes in the ellipsometric parameters can be performed with high sensitivity using a nonlinear detector. The art is to find a suitable compromise.

11.1.5 Infrared Synchrotron Radiation

In the last decade,[15, 16] synchrotron radiation from storage rings was developed as a powerful source of infrared radiation. One advantage of

this unique source is a low opening angle combined with a high radiation density per unit area described by the quantity "brilliance". In the mid-infrared, the source size is diffraction limited and allows measurements with a very small throughput. Ellipsometric measurements benefit from these properties and it is possible to reduce the sample area and the opening angle of the incidence beam. The latter reduces variation of the incidence angle and, therefore, a higher angle resolution can be achieved. Many papers applying infrared synchrotron radiation as a radiation source for ellipsometry have been published (e.g.,[17, 18]). Synchrotron radiation is polarized. At the Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung (BESSY) the distribution of the radiation field of the infrared synchrotron radiation was measured ellipsometrically and characterized for the first time by the four Stokes parameters using a retarder.[19]

11.2 Applications

11.2.1 Optics of Absorbing Media

The optical behavior of absorbing media depends on the complex refractive index $\hat{n}$ or the complex dielectric function ε. Between these quantities the relation exists:

$$\hat{n} = \sqrt{\varepsilon} \text{ or in more detail } \varepsilon = \varepsilon' + i\varepsilon'' = (n^2 - k^2) + i(2nk).$$

The absorption of radiation in an isotropic layer of the thickness d is described by Lambert-Beer's law, which results in the transmission T through the material:

$$T = \frac{I}{I_0} = e^{-\alpha d} = 10^{-\varepsilon c d} \tag{11.21a}$$

I, I_0 = transmitted and incident intensity
α = absorption coefficient
ε = molar absorption coefficient [m^2/Mol], not to be confused with the dielectric function
c = concentration [Mol/m^3].

The absorption index k, the imaginary part of the complex refractive index, is related to quantities in Eq. (11.21a) by

$$\alpha = 4\pi\tilde{\nu}k \quad \text{or} \quad k = \frac{\varepsilon c}{4\pi\tilde{\nu}\ln(10)} \tag{11.21b}$$

$\tilde{\nu}$ = wavenumber [cm^{-1}]

An electromagnetic wave, with an angle φ_0 incident into an absorbing medium, is transferred to an inhomogeneous wave. Planes of constant amplitude have a different direction than the planes of constant phase. The electric field E in the medium is given by the expression

$$E = E_0 e^{-i\omega\left(t - \frac{\vec{r}\cdot\vec{s}}{V}\right)}, \quad \omega = 2\pi\nu,$$

where E_0 is the field strength of the incident wave of the frequency ν, $\vec{r}$ the position vector, $\vec{s}$ the unit vector for the direction of propagation inside of the medium, and V the velocity in the medium. The position dependency of E, written in components,[5, 10] (see Fig. 11.4) is

$$\omega\left(\frac{\vec{r}\cdot\vec{s}}{V}\right) = \frac{\omega}{V}[x\sin\varphi_1 + z\cos\varphi_1)$$

$$= 2\pi\tilde{\nu}[x(n_0\sin\varphi_0) + z\mathrm{Re}(\hat{n}_1\cos\varphi_1) + iz\mathrm{Im}(\hat{n}_1\cos\varphi_1)]$$

n_0, φ_0 = refractive index and angle of incidence for the first medium
n_1, φ_1 = refractive index and angle of refraction for the second medium.

The y-direction (perpendicular to the plane of incidence) can be omitted, since the radiation only propagates in the plane of incidence. For the

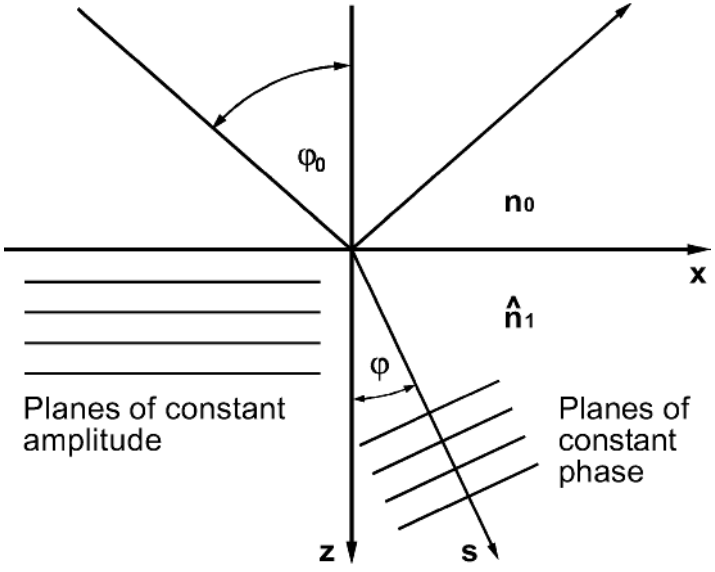


Figure 11.4 Scheme of the inhomogeneous wave inside an absorbing medium.

time independent part of the field we get

$$E = E_0 \exp[-2\pi\tilde{\nu}\,\mathrm{Im}(\hat{n}_1\cos\varphi_1)z]\exp[2\pi i\tilde{\nu}(x(n_0\sin\varphi_0) + z\mathrm{Re}(\hat{n}_1\cos\varphi_1)] \tag{11.22}$$

From this equation we see that the amplitude only depends on the z-direction. The planes of constant amplitude are always parallel to the surface. The direction of propagation φ, which is the normal of the planes of constant phases, deviates from Snellius' law (see Fig. 11.4) and can be derived from Eq. (11.22):

$$\tan\varphi = \frac{n_0\sin\varphi_0}{\mathrm{Re}(\hat{n}_1\cos\varphi_1)} = \frac{n_0\sin\varphi_0}{\mathrm{Re}\sqrt{(\hat{n}_1{}^2 - (n_0\sin\varphi_0)^2)}} \tag{11.23}$$

Generally we have the fact that $\varphi \neq \varphi_1$ (*Snellius*) for $k \neq 0$.

From Eq. (11.22) the intensity in the medium is calculated to:

$$I = I_0\exp[-4\pi\tilde{\nu}\,\mathrm{Im}\sqrt{\hat{n}_1{}^2 - (n_0\sin\varphi_0)^2}z] \tag{11.24a}$$

and for the general form of α we get:

$$\alpha = 4\pi\tilde{\nu}\,\mathrm{Im}\sqrt{\hat{n}_1{}^2 - (n_0\sin\varphi_0)^2} \tag{11.24b}$$

From Eq. (11.24b) for $\varphi_0 = 0$, we get $\alpha = 4\pi\tilde{\nu}k$.

11.2.2 Vibration Modes – the Concept of Weak and Strong Oscillators

11.2.2.1 The Oscillator Model

Ellipsometry in the infrared range is dominated by the absorption bands of molecular vibrations and the resonances of solid states. Thin film interference is also an important feature, but in many cases the films are not thick enough to produce one or more orders of the interference pattern.

The dielectric function, ε, of a selective vibrational band for an isotropic sample can be described by the model of the Lorentz oscillator.[5, 20] An ensemble of oscillators is given by the sum of individual oscillators.

$$\varepsilon = \varepsilon' + i\varepsilon'' = \varepsilon_\infty + \frac{4\pi\dfrac{e^2}{m}Nf}{\omega_0^2 - \omega^2 - i\gamma\omega} \quad \text{for one oscillator.}$$

For several oscillators

$$\varepsilon' = \varepsilon_\infty + \sum_l \frac{F_l(\tilde{\nu}_{0,l}^{\,2} - \tilde{\nu}^2)}{(\tilde{\nu}_{0,l}^{\,2} - \tilde{\nu}^2)^2 + (\Gamma_l\tilde{\nu})^2}, \qquad \varepsilon'' = \sum_l \frac{F_l(\Gamma_l\nu)}{(\tilde{\nu}_{0,l}^{\,2} - \tilde{\nu}^2)^2 + (\Gamma_l\tilde{\nu})^2} \tag{11.25}$$

with $F = (\varepsilon_0 - \varepsilon_\infty)\tilde{\nu}_{0,l}^{\,2}$, $(\varepsilon_0 - \varepsilon_\infty) = \dfrac{Ne^2}{\pi m \nu_{0,l}^{\,2}} f$ for the l^{th} oscillator

$$F = \frac{4\pi N e^2 f}{(2\pi c)^2 m} = 8.96{*}10^{-16}\text{Nf } [\text{cm}^{-2}];\ \Gamma = \frac{\gamma}{2\pi c}$$

F_l = parameter of the l^{th} oscillator strength [cm^{-2}]
Γ_l = damping constant of the l^{th} oscillator [cm^{-1}]
γ = damping constant [sec^{-1}]
l = index of a single oscillator
ε_0, ε_∞ = real part of the dielectric function for zero and very high frequency, respectively
$\tilde{\nu}$, $\tilde{\nu}_{0,l}$ = wavenumber in the spectrum and of the oscillator position [cm^{-1}]
N = number of oscillators per unit volume [cm^{-3}]
e, m = effective electron charge and mass respectively
$\nu_{0,l}$ = time frequency of the l^{th} oscillator [s^{-1}]
f = oscillator strength
c = light velocity

11.2.2.2 The Weak Oscillator

The ellipsometric properties of an oscillator are different and depend on the values of the parameters F and Γ. In principle, two kinds of oscillators can be distinguished: the strong and the weak. Figures 11.5a, b, c, and d show typical examples for both cases.

The weak oscillator is modelled with $\tilde{\nu}_0 = 1000\ \text{cm}^{-1}$, $F = 2000\ \text{cm}^{-2}$, $\Gamma = 8\ \text{cm}^{-1}$, and $\varepsilon_\infty = 1.96$. For this oscillator ε' and n are greater then the unit in all parts of the spectrum. This is typically for the weak oscillator. Figure 11.5a shows n, k and in combination with the corresponding ellipsometric spectra of Ψ and Δ for isotropic bulk material (Fig. 11.5b). Spectra of the weak oscillator are frequently observed in connection with organic matter.

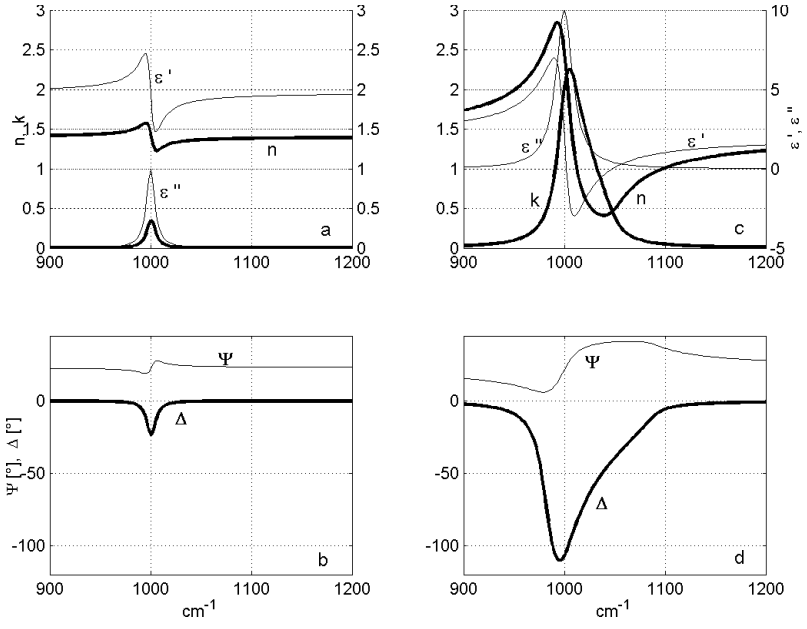


Figure 11.5 The optical constants, n and k, and dielectric function, ε, for a weak and strong oscillator are shown in a) and c), respectively. b) and d) show the corresponding ellipsometric spectra of Ψ and Δ.

11.2.2.3 The Strong Oscillator

The strong oscillator is observed in solids and inorganic molecules with optical constants, expressed in ε or $\hat{n}$, very different from those of the weak oscillator. The corresponding ellipsometric spectra have pronounced features, which are additionally modified if a film with a strong oscillator is applied.

The strong oscillator can be simulated with the data:

$$\tilde{\nu}_0 = 1000\ \text{cm}^{-1},\ F = 2{*}10^5\ \text{cm}^{-2},\ \Gamma = 15\ \text{cm}^{-1} \text{ and } \varepsilon_\infty = 1.96.$$

Figure 11.5c shows the optical constants of bulk, combined with the corresponding ellipsometric spectra (Fig. 11.5d). The refractive index spectrum covers a great range of n values and crosses the Brewster angle twice for a typical angle of incidence (70°), but the maximum of the phase shift in Δ is lower than 180° by the effect of the absorption index k. For wavenumbers higher than the oscillator frequency, the refractive index is below one for a limited range, as a typical feature. Whereas the spectrum

of ε'' is always symmetric, for a strong oscillator the absorption index k is asymmetric with a wavenumber shift of the maximum, caused by the complex square root of ε.

For perspective, we consider it a weak oscillator if the maximum of k is lower than about 0.5. A strong oscillator requires a refractive index lower than unity on the high frequency part of the n-spectrum and k_{Max} in the range $k \geq 1$. This limit is fuzzy.

11.2.3 Inversion of Infrared Ellipsometric Measurements

The calculation of the optical constants n and k or ε from measured $\tan\Psi$, Δ spectra is not straightforward, except for the case of an isotropic bulk surface without any film. In all other cases, iteration procedures must be applied. In this section an example is demonstrated for both cases.

11.2.3.1 Calculation of the Dielectric Function and the Optical Constants from Bulk Material

The ellipsometric (complex) amplitude ratio ρ, is described by[10]

$$\rho = \frac{r_p}{r_s} = \tan\Psi \exp(i\Delta), \quad \tan\Psi = \frac{|r_p|}{|r_s|}; \quad \Delta = \delta_p - \delta_s,$$

which can be rewritten applying the Fresnel equations for r_s and r_p, $a = n_0 \sin\varphi_0 \tan\varphi_0$; $w = \sqrt{\varepsilon - (n_0 \sin\varphi_0)^2}$ and using the abbreviations a and w

$$\rho = \frac{a - w}{a + w}.$$

By solving this equation for ε we get the well-known relations of metal optics

$$\varepsilon' = n^2 - k^2 = (n_0 \sin\varphi_0 \tan\varphi_0)^2 \left[1 + \tan^2\varphi_0 \frac{\cos^2 2\Psi - \sin^2 2\Psi \sin^2\Delta}{(1 + \sin 2\Psi \cos\Delta)^2}\right] \tag{11.26a}$$

$$\varepsilon'' = 2nk = -(n_0 \sin\varphi_0 \tan\varphi_0)^2 \frac{\sin 4\Psi \sin\Delta}{(1 + \sin 2\Psi \cos\Delta)^2} \tag{11.26b}$$

$$n = \sqrt{\frac{\sqrt{\varepsilon'^2 + \varepsilon''^2} + \varepsilon'}{2}} \tag{11.27a}$$

$$k = \sqrt{\frac{\sqrt{\varepsilon'^2 + \varepsilon''^2} - \varepsilon'}{2}}. \tag{11.27b}$$

Equations (11.26) and (11.27) are also valid for the attenuated total reflection (ATR) with $n_0 > 1$. Infrared spectroscopic ATR measurements are a very powerful tool for investigating weak bands or samples which are normally not easy to handle (e.g., soft or liquid).[21–23]

In practice, measurements of the absorption index are restricted to values of approximately $k > 0.01$ by the limited accuracy of Δ. The refractive index can be measured without this restriction. For n in this case, a reduced form of the Eqs. (11.26) and (11.27) can be derived.

$$n = n_0 \sin\varphi_0 \sqrt{1 + \left[\frac{\tan\varphi_0 \cos 2\Psi}{1 \pm \sin 2\Psi}\right]^2} \tag{11.28}$$

As an example for the evaluation of isotropic bulk material in Figs. 11.6a through d the n, k spectra of silicone rubber (b, d) are shown, with the $\tan\Psi$, Δ spectra (a, c) for comparison.

11.2.3.2 Inversion of the Thin Film-Bulk System

The inversion of thin film equations is not possible in an algebraic form. According to the specific situation and the model used, different iteration procedures can be found in the literature. In the following, an interesting example of infrared ellipsometric investigations is demonstrated, applying the model of isotropic film/substrate system. This iteration procedure from Reinberg[24] in the original version is only available for a non-absorbing film. However, an extension for an absorbing film is possible and will be given later.

Starting from the thin film equation for the s and p polarization (perpendicular and parallel directions), we can rewrite the inversion problem in a square equation using the ellipsometric ratio $\rho = r_p/r_s$.

$$r_\nu = \frac{r_{\nu 1} + r_{\nu 2} X}{1 + r_{\nu 1} r_{\nu 2} X}; \quad \nu \text{ stands for the s or p polarization}$$

$$X = \exp(4\pi i d_1 \tilde{\nu}(n_1 \cos\varphi_1)) = \exp(4\pi i d_1 \tilde{\nu} \sqrt{n_1^2 - (n_0 \sin\varphi_0)^2}). \tag{11.29}$$

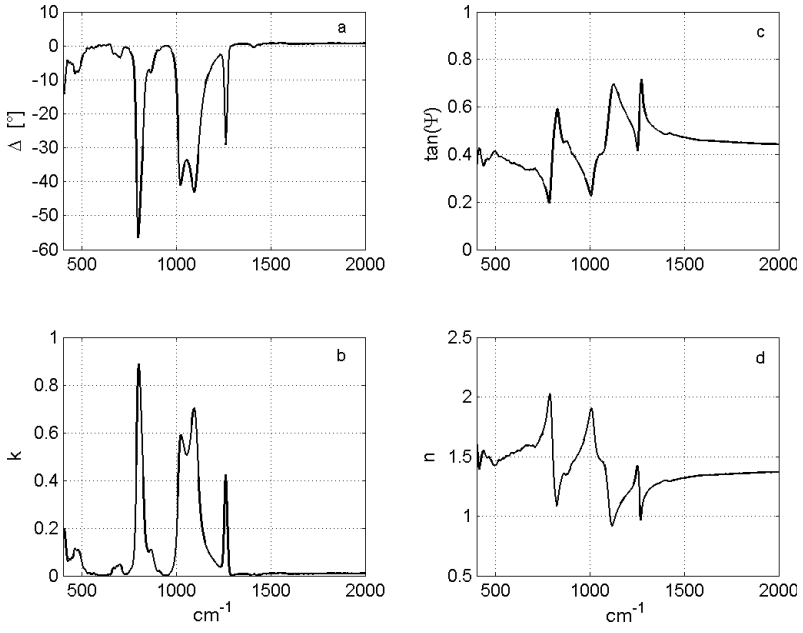


Figure 11.6 Ellipsometric spectra a) and c) show Ψ and Δ, respectively, and the optical constants b) and d) show *n* and *k*, respectively of a silicone rubber, applying the model of an isotropic surface.

n_1, d_1 = refractive index and thickness of the film
φ_1 = angle of propagation inside of the film

$$(\rho F - C)X^2 + (\rho E - B)X + (\rho D - A) = 0, \tag{11.30}$$

where

$A = r_{p1}$,
$B = r_{p2} + r_{p1}r_{s1}r_{s2}$,
$C = r_{p2}r_{s1}r_{s2}$,
$D = r_{s1}$,
$E = r_{s2} + r_{p1}r_{p2}r_{s1}$,
$F = r_{p1}r_{p2}r_{s2}$

Now a solution of Eq. (11.30) must be found by the variation of n_1 with the condition $|X| = 1$. With the knowledge of n_1 we determine the

thickness d_1 by the argument of X:

$$d_1 = \frac{\arg(X)}{4\pi\tilde{\nu}\sqrt{n_1^2 - (n_0 \sin\varphi_0)^2}}. \tag{11.31}$$

For a film of suitable thickness the interference order must be considered. This algorithm is applied for each wavenumber of the ellipsometric spectrum.

Infrared spectra of organic media always have regions without absorption. These regions can be used to apply the Reinberg algorithm, to determine the refractive index outside of the vibrational resonance's (n_∞) and thickness (d) are determined. Using these values, an oscillator fit is possible to determine the corresponding n and k spectra.

The Reinberg algorithm is not applicable for a free standing film or a film with the same refractive indices of the medium in front and behind the film. Deriving the optical constants of a thin layer on a metallic substrate is a second problem, because the algorithm reacts sensitive on $\tan\Psi$. In this specific case $\tan\Psi$ is near unity and difficult to measure with the necessary high accuracy.

An example is given in Fig. 11.7a, b for a fluorinated fuel tank.[25] The fuel tank, consisting of polymer material, is covered inside with Teflon-like thin film to prevent the penetration of the hydrocarbons. The spectra of the optical constants of the substrate are calculated by Eqs. (11.26 and 11.27) from the spectra $\tan\Psi$ and Δ, measured at the outer surface of the tank. These spectra show the C-H bands at 700, 1470 and 3000 cm^{-1}. The spectra of the film of the inner surface have a slope in $\tan\Psi$ and Δ, caused by the Teflon-like film (Fig. 11.7a, b). From the region free of absorption between 2600 and 1600 cm^{-1} we get n_∞ and the thickness d of the film by using the Reinberg algorithm (Fig. 11.7d). These values are necessary for the oscillator fit of the C-F stretching band close to 1100 cm^{-1} (Fig. 11.7c). The result of the fit is shown in Fig. 11.7a, b in comparison with the measured ellipsometric spectra.

11.2.3.3 Modification of the Reinberg Algorithm

In the former version, the Reinberg algorithm demands that the absorption index be zero. However, this procedure can be extended for $k \neq 0$, if the condition $|X| = 1$ is changed. For solving Eq. (11.30), the refractive index of the film is now assumed to be complex and X is calculated from a field of n_j and k_j values: $X_j = X(n_j, k_j)$. The target $|X| = 1$ is now replaced by the expression

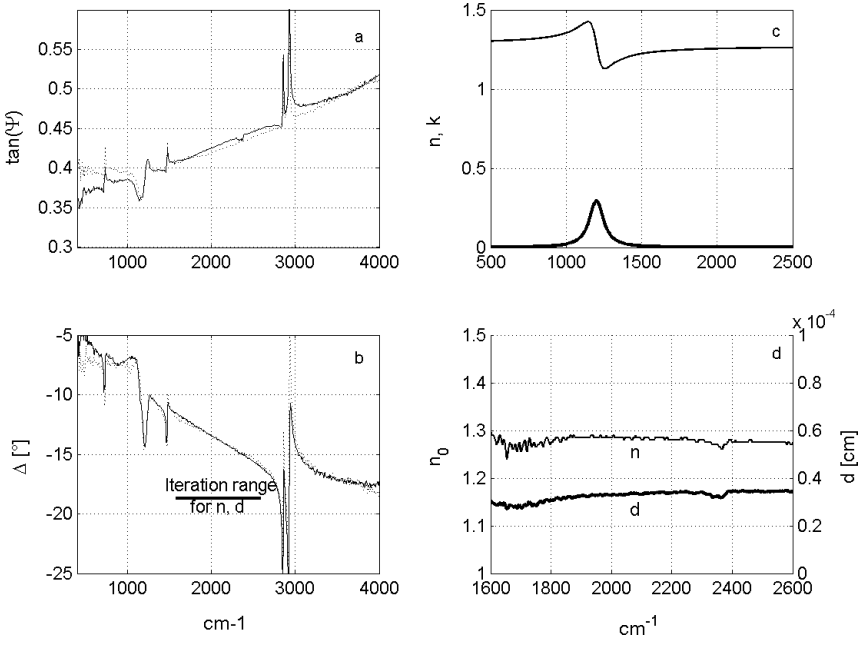


Figure 11.7 Measurement and modeling of a Teflon-like film inside of a fluorinated fuel tank. Parts a) and b) are the spectra of tanΨ and Δ of the film covered tank together with the fitted result (broken line). Part c) contains the fitted C-F band (Teflon). Part d) shows the film thickness and the refractive index outside of the absorption bands.

$$X_{j,m} = X_j(d_m) = 4\pi i\tilde{\nu}d_m\sqrt{\hat{n}_{1,j}^2 - (n_0 \sin\varphi_0)^2} \qquad (11.32)$$

for a series of thickness values d_m.

Searching for the minimum of Y_m from $Y_m = \mathrm{Min}(X_j - X_{j,m})$ in Eq. (11.32) results in a threefold number of the values n, k, d for the best agreement with the experimental ratio ρ for each wavenumber. A test for the quality of the calculation and for the model used is the independence of the thickness d from the wavenumber. An example of the application of the extended Reinberg algorithm, the determination of the optical constants, and the thickness of a gold island film, is given in Fig. 11.8.

In a first step of the calculation, n, k and d are changed in suitable steps, which also gives an idea of the "calculation noise" in their spectra (thin line in Fig. 11.8). With the mean value of the constant thickness, taken over a great wavenumber range, we can again determine n and k, now with much higher precision (thick line in Fig. 11.8). In this case, we can calculate the anisotropic island film (uniaxial) with an isotropic algorithm,

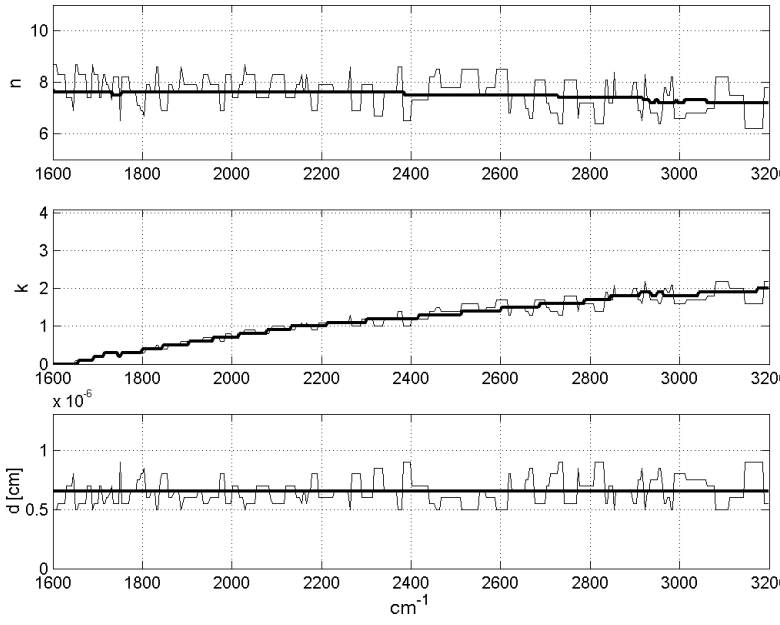


Figure 11.8 The optical constants, *n* and *k*, and the thickness *d* of a gold island film, calculated by the modified Reinberg algorithm.

because the ordinary ($n_o \sim 8$) and extraordinary ($n_e \sim 1.5$) refractive indices are very different.

The optical properties of metallic island films are completely different from the properties of the bulk.[26] The refractive index *n* and absorption index *k* are approximately constant in a wide wavenumber spectral range. *n* is greater than that of the bulk for parts of the infrared spectrum. *k* is much lower in the whole spectral range. Such films, covered with a monolayer of organic molecules, show a surface enhanced infrared absorption in the absorption bands. This effect, called SEIRA (surface enhanced infrared absorption), is used to enhance the detection limit of molecules in the infrared.

11.2.4 Anisotropy Features in the Infrared Ellipsometric Spectra

The ellipsometric spectra of the infrared absorption bands are determined by changing the dipole moment of the vibrations in the molecule. The direction of the bonds is fixed in space in many cases, for instance if

Table 11-1. Characteristics and Examples of Weak and Strong Oscillators.

Matter / Oscillator	Weak oscillator	Strong oscillator
Bulk, isotropic or anisotropic	Organic matter	Inorganic matter
Example	Silicone rubber	Quartz glass
Film on dielectric substrate	Organic films	Berreman effect (z-comp.), Response at oscillator frequency
Example	Petrol tank (CF_4/PET)	SiO_2/ Si, SiH/ Si
Film on metal substrate	inorganic films; only z-component	Berreman effect (z-comp.), d $<$ 200 nm Surface wave, d $>$ 1000 nm
Example	LB-film / Au	Al_2O_3 / Al, SiO_2 / Al

molecules are deposited on surfaces or appear in the crystalline matter. This situation can be simulated using different oscillator parameters in Eq. (11.25) for the x, y, and z components of the vibration. Generally, many special cases must be distinguished between the weak and strong oscillator, bulk and film, whereas for the film, the substrate, metallic or dielectric, is important. Table 11-1 gives an overview and summarizes the different kinds of response. The following sections discuss this in more detail.

11.2.4.1 Ellipsometric Response of an Anisotropic Medium

Typical features of the ellipsometric spectra for the anisotropic case are calculated using the equations of anisotropic matter and for a film on an isotropic or anisotropic substrate,[26–29] using the oscillator model of Eq. (11.25) for the determination of the n, k spectra of the model oscillator:

$$r_{01p} = \frac{n_{1x}n_{1z}\cos\varphi_0 - n_0(n_{1z}^2 - n_0^2\sin^2\varphi_0)^{1/2}}{n_{1x}n_{1z}\cos\varphi_0 + n_0(n_{1z}^2 - n_0^2\sin^2\varphi_0)^{1/2}} \tag{11.33a}$$

$$r_{12p} = \frac{n_{2x}n_{2z}(n_{1z}^2 - n_0^2\sin^2\varphi_0)^{1/2} - n_{1x}n_{1z}(n_{2z}^2 - n_0^2\sin^2\varphi_0)^{1/2}}{n_{2x}n_{2z}(n_{1z}^2 - n_0^2\sin^2\varphi_0)^{1/2} + n_{1x}n_{1z}(n_{2z}^2 - n_0^2\sin^2\varphi_0)^{1/2}} \tag{11.33b}$$

$$r_{01s} = \frac{n_0\cos\varphi_0 - (n_{1y}^2 - n_0^2\sin^2\varphi_0)^{1/2}}{n_0\cos\varphi_0 + (n_{1y}^2 - n_0^2\sin^2\varphi_0)^{1/2}} \tag{11.33c}$$

$$r_{12s} = \frac{(n_{1y}^2 - n_0^2 \sin^2\varphi_0)^{1/2} - (n_{2y}^2 - n_0^2 \sin^2\varphi_0)^{1/2}}{(n_{1y}^2 - n_0^2 \sin^2\varphi_0)^{1/2} + (n_{2y}^2 - n_0^2 \sin^2\varphi_0)^{1/2}} \tag{11.33d}$$

$$n_0 \sin\varphi_0 = n_2 \sin\varphi_2 \qquad \text{(isotropic substrate)} \tag{11.34}$$

$$r_l = \frac{r_{01l} + r_{12l} e^{i\delta_l}}{1 + r_{01l} r_{12l} e^{i\delta_l}}, \quad l = s \text{ or } p \tag{11.35}$$

$$\delta_p = 4\pi\tilde{\nu} d \left(\frac{n_{1x}}{n_{1z}}\right)(n_{1z}^2 - n_0^2 \sin^2\varphi_0)^{1/2}; \quad \delta_s = 4\pi\tilde{\nu} d (n_{1y}^2 - n_0^2 \sin^2\varphi_0)^{1/2} \tag{11.36}$$

r_{01p}, r_{01s}, r_{12p}, r_{12s} = reflection at the air-film and film-substrate boundary for the p and s components, respectively
φ_0 = angle of incidence
n_0 = refractive index of the ambient medium
n_2 = refractive index of an isotropic substrate
n_{1x}, n_{1y}, n_{1z} = complex refractive indices of the film in the corresponding directions
n_{2x}, n_{2y}, n_{2z} = complex refractive indices of the substrate in the corresponding directions
r_p, r_s = reflection of the film for the p and s components, respectively
φ_2 = angle into the substrate
δ_p, δ_s = phase angles for the p and s components, respectively
$\tilde{\nu}$ = wavenumber
d = thickness of the film

The above equations describe the general case of a biaxial film on a biaxial substrate. For the special case of a uniaxial film with the optical axis perpendicular to the surface, we have the special case where

$$n_{1x} = n_{1y} = n_{1o}, \; n_{1z} = n_{1e}$$

with n_{1o} and n_{1e} for the ordinary and extraordinary refractive indexes, respectively, we get:

$$r_{01p} = \frac{n_{1o} n_{1e} \cos\varphi_0 - n_0 n_{1e} \cos\varphi_{1e}}{n_{1o} n_{1e} \cos\varphi_0 + n_0 n_{1e} \cos\varphi_{1e}} = \frac{n_{1o} n_{1e} \cos\varphi_0 - n_0 (n_{1e}^2 - (n_0 \sin\varphi_0)^2)^{1/2}}{n_{1o} n_{1e} \cos\varphi_0 + n_0 (n_{1e}^2 - (n_0 \sin\varphi_0)^2)^{1/2}}$$

$$r_{12p} = \frac{n_{2o} n_{2e} (n_{1e}^2 - (n_0 \sin\varphi_0)^2)^{1/2} - n_{1o} n_{1e} (n_{2e}^2 - (n_0 \sin\varphi_0)^2)^{1/2}}{n_{2o} n_{2e} (n_{1e}^2 - (n_0 \sin\varphi_0)^2)^{1/2} + n_{1o} n_{1e} (n_{2e}^2 - (n_0 \sin\varphi_0)^2)^{1/2}} \tag{11.37}$$

The perpendicular component is given by the suitable application of Eq. (11.33c). With the equations (11.33a, c) and (11.35 through 11.37)

and the special case of $n_{1x} = n_{1y} = n_{1o}$, $n_{1z} = n_{1e}$, the anisotropic film-substrate response can be calculated.

For simplifying the pool of different possibilities of the refractive indices for the different directions, we restrict the orientations of the vibrations of the oscillator to the three cases shown in Table 11-2, which are important in practice.

11.2.4.2 Matter with Weak Oscillators (Column 2 in Table 11-1)

Bulk

In the category of matter with weak oscillators, mostly organic compounds are found. For the bulk surface, these three special cases for the vibration components listed in Table 11-2 result in different features for the various components, as shown in Fig. 11.9. The minima in Δ for the cases 1 (isotropic, thick line) and 2 (xy-plane, thin line) have different values and are connected with corresponding dispersion like features in $\tan\Psi$. The spectra of $\tan\Psi$ and Δ for the z component (case 3, dotted line) are very different from the spectra of cases 1 and 2. In the spectra of Δ we find a maximum and in the spectra of $\tan\Psi$ an opposite gradient with respect to the cases 1 and 2. The wavenumber position of the maximum of Δ is shifted to higher wavenumbers. For comparison in Fig. 11.9a–c the n, k and ε', ε'' spectra (Fig. 11.9i) of the model oscillator are shown. The parameters are the same for the selected orientations. In an oscillator fit for determining the corresponding parameters, it is possible to derive the orientation of the dipole moment from the features of $\tan\Psi$ and Δ in Fig. 11.9a, b. An example is the spectrum of silicone rubber in Fig. 11.6, measured at an angle of incidence $\varphi_0 = 70°$.

Film on dielectric substrate

For a film on an isotropic substrate the properties in $\tan\Psi$ and Δ are exchanged with respect to the former discussed bulk spectra, as seen in

Table 11-2. Three cases of the vibrational components of the dipole moment.

Case	Used directions of the dipole moment of the vibration
1	the isotropic case: the x, y and z component of the dipole moment are equal (random distribution)
2	the x and y component of the dipole moments have equal values (random distribution in the surface) and are different to the z direction
3	only the z component of the vibrations is present (the dipole moment of the vibration is oriented perpendicular to the surface)

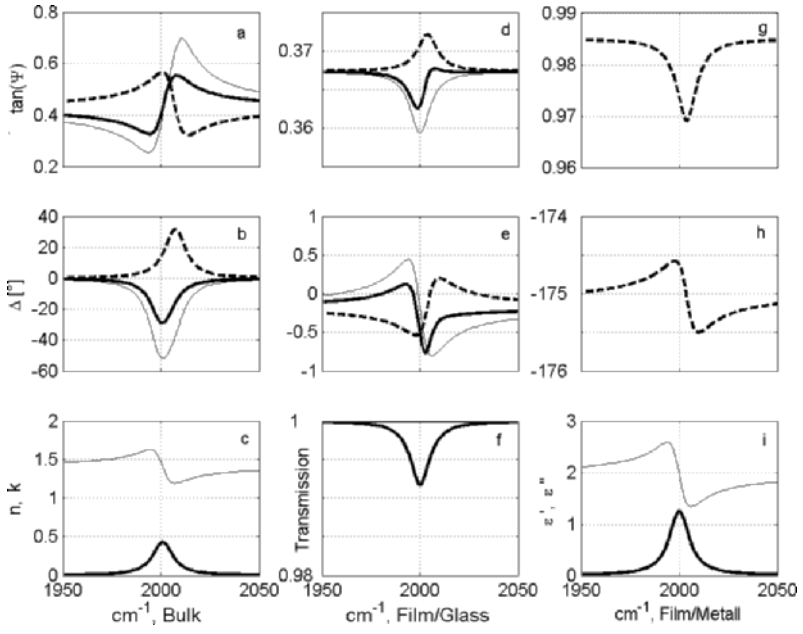


Figure 11.9 Features of the ellipsometric spectra for an anisotropic weak oscillator for bulk and on different substrates (details are described in the text). Lines in a, b, d, e, g, h: thick, isotropic; thin, xy-component; dotted, z-component.

Fig. 11.9d, e. Maxima and minima now are found in $\tan\Psi$ and a dispersion-like behavior in Δ. These ellipsometric spectra around the oscillator allow distinguishing between a bulk sample and a film-covered substrate. Additionally, Fig. 11.9f shows the corresponding transmission spectrum for the same thickness and an angle of incidence equal to zero is taken from a sample cell with KBr windows ($n = 1.52$). In this case only the x and y components of the vibration contribute to the observed response. An example is the characterization of a fluorinated fuel tank, as demonstrated in connection with the Reinberg algorithm in Section 11.2.3.2.

Film on metal substrate

The same film discussed above results on a metallic substrate in a response only for the z component of the dipole moment, because the electric field on the surface of the metal vanishes (Fig. 11.9g, h). The wavenumber position of the band is also shifted, as for the z component of the former cases. Thicker films with thicknesses in the micrometer range show a contribution from the x and y components, because the air-film boundary now has a greater distance to the metal surface. For

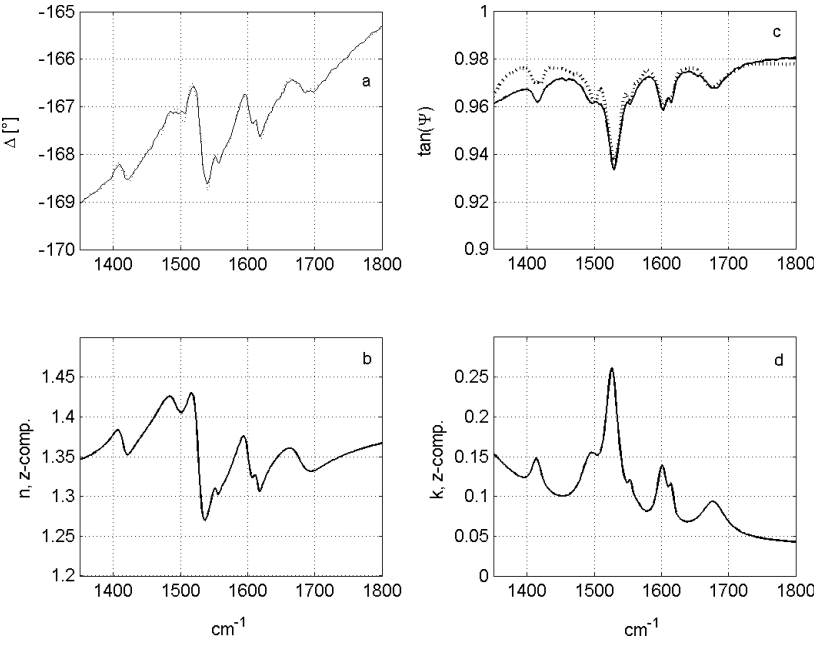


Figure 11.10 Ellipsometric spectrum and oscillator fit of a Langmuir-Blodgett film built up by six double layers of oxadiazol (60 nm thickness). Solid line, measurement; dotted line, fit results.

comparison, Fig. 11.9c, i show the optical constants (n, k and ε', ε'' of the bulk and film matter with the same model oscillator.

A Langmuir-Blodgett (LB) film of 60 nm thickness, consisting of six double layer oxadiazol, on a gold substrate is an example for this arrangement.[34] The successful oscillator fit delivers only the z-component of the dipole moment, caused by the metallic substrate (Fig. 11.10a–d).

11.2.4.3 Matter with Strong Oscillators (Column 3 in Table 11-1)

Bulk

In accordance with the three special cases of the orientation of the oscillator components (Table 11-2) in the former section, now the ellipsometric spectra for the strong oscillator will be evaluated[35]. New features are observed by the Berreman effect and the excitation of surface waves on metal substrate for a greater angle of incidence ($\varphi_0 > 65°$).

The optical constants of a strong oscillator show an asymmetric slope in n and k, caused by Maxwell's square relation between the complex

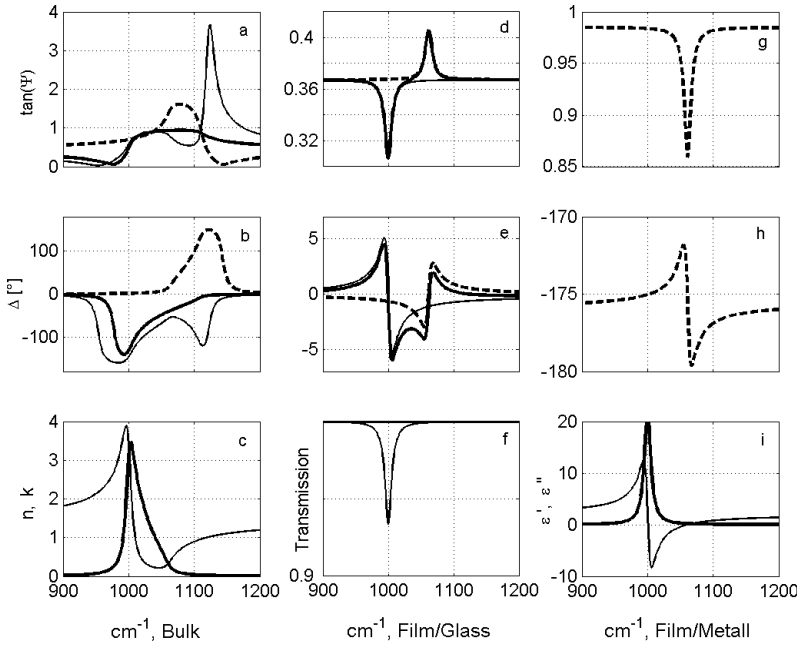


Figure 11.11 Features in the ellipsometric spectra for an anisotropic strong oscillator for bulk and on different substrates (details are described in the text). Lines in a, b, d, e, g, h: thick, isotropic; thin, xy-component; dotted, z-component.

dielectric function ($\varepsilon = \varepsilon' + i\varepsilon''$) and the complex refractive index ($\hat{n} = n + ik$). A region of the high frequency part of the oscillator position with $n < 1$ and $\varepsilon' < 0$ is a typical situation, which results in some additional features of the ellipsometric spectra.

For bulk, the wavenumber range of the spectra is more extended. For the isotropic case $\tan\Psi$ never exceeds unity, which is possible for the anisotropic spectra in Fig. 11.11a–b. An anisotropic dipole moment, lying parallel to the surface, generates a feature approximately at the position of $n \cong 1$ with a strong value of $\tan\Psi$ and an additional dispersion-like feature of Δ, even if there is no additional oscillator. Also, the properties of the z component are shifted in the wavenumber. $\tan\Psi$ values that exceed unity are typical for deviations from the isotropic case. In Fig. 11.11c, i the n, k and ε', ε'' − spectra, are shown for comparison. The wavenumber position of the corresponding oscillator can only be found by an oscillator fit and is not directly derivable from the wavenumber position of the feature in the ellipsometric spectra. The bands are shifted to higher wavenumbers.

Film on dielectric substrate

A film on an isotropic substrate ($n_s = 1.5$, e.g., glass) shows a minimum in the $\tan\Psi$ spectrum at the oscillator position for $\varphi_0 = 70°$ (greater

then the Brewster angle) and a maximum, which is an additional feature near the position of $\varepsilon' = 0$ or $n = k$ ($n < 1$), Fig. 11.11d, e. This response is caused by the Berreman effect,[30, 31] observed for thin films ($d < 200$ nm) with a strong oscillator (details are discussed later). In the Δ-spectrum corresponding dispersion-like features are observed. The spectra of the x, y, and z components strongly differ, whereas the isotropic spectrum is a mixture of both. Figure 11.11f shows the corresponding transmission spectrum, as also shown in Fig. 11.9f.

Using an angle of incidence lower than the Brewster angle of the substrate (e.g., $\varphi_0 = 65°$ for silicon with n(substrate) = 3.42), the properties of the ellipsometric spectra in the extreme values in $\tan\Psi$ and the gradients in Δ are completely changed, as demonstrated in Fig. 11.12a, b in comparison to the spectra of the film on glass in Fig. 11.11d, e.

A comparison of the theoretical and experimental film spectra on a silicon substrate is shown in Fig. 11.12a, b, d, e. The first column shows the calculated ellipsometric spectra for a film thickness of 8 nm and an angle

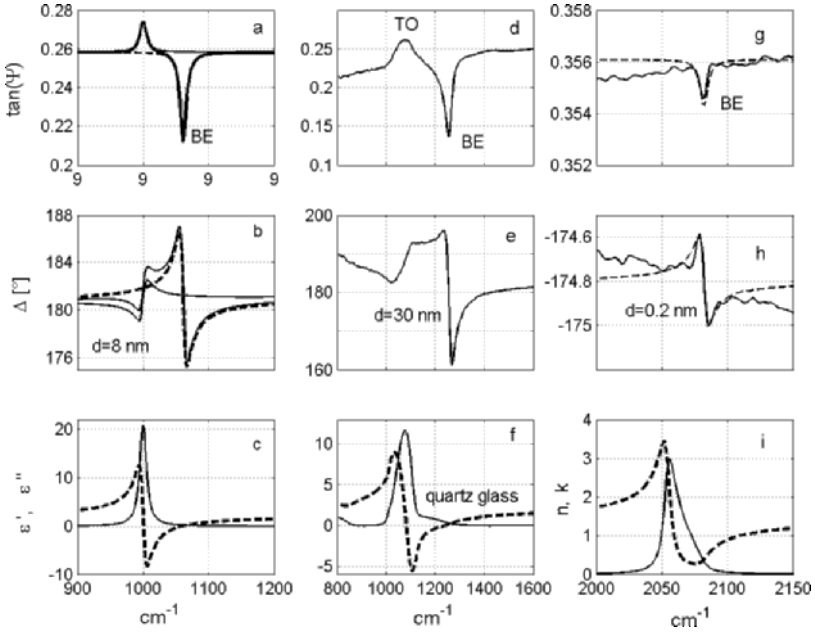


Figure 11.12 Thin films on a silicon substrate. a), b), and c) show the results of modeling the ellipsometric spectra. d), e), and f) show the experimental results for a SiO_2 film (30 nm) together with the quartz glass spectrum for comparison. The parts g), h), and i) show the Berreman effect of a measured mono-layer of SiH together with the derived oscillator spectrum. Lines in a, b, d, e, g, h: thick, isotropic; thin, *xy*-component; dotted, *z*-component.

of incidence of $\varphi_0 = 65°$, simultaneously, together with the ε', ε'' spectra applied. The second column shows the spectra of a 30 nm thick SiO_2 film measured. The response of the transversal optical mode (TO) and the Berreman effect is also seen. For comparison, the same column contains the ε', ε'' in Fig. 11.12f spectra of quartz glass.

A monolayer of SiH on Si (111)-surface with a thickness of 0.2 nm results in features in the $\tan\Psi$ and Δ spectra (shown in Fig. 11.12g, h),[36] which are only formed by the Berreman effect, because here only a z-component is effective. The corresponding spectra of the optical constants, calculated by an oscillator fit on the base of the film algorithm (assuming these relations are valid for such a thin film), are shown in Fig. 11.12i.

Film on metal substrate

- The Berreman effect and excitation of surface waves -

The ellipsometric response of a thin film ($d < 200$ nm), which contains a strong oscillator, on a metallic or dielectric substrate results in a new spectral feature with a shifted wavenumber position referring to the vibrational frequency. This Berreman effect is only observed for the z component of the reflection and is caused by launching a leaky waveguide within the film. The excitation is coupled with a phase condition, which is realized in the neighborhood of $\varepsilon' = 0$ and for $\varepsilon'' \ll 1$.[5, 31, 32] The Berreman effect can be observed at dielectrics and metals in dependence on the film thickness with a maximum response at the so-called Berreman thickness[32], which is in the range of approximately 50 to 100 nm. The response depends also on the reflectivity of the film boundaries, which is optimal for a metal substrate and great angle of incidence.

Figure 11.13 shows the $\tan\Psi$ spectrum of the Berreman effect in dependence on the film thickness in a logarithmic scale. The start of the response of the oscillator frequency is about two orders later than for the Berreman effect, because here the leaky waveguide results in a much larger interaction length in the film material along the surface and therefore in a higher sensitivity in reference to a transmission measurement. As seen, the Berreman effect drops down with increasing thickness and a new feature arises at higher frequencies, due to the excitation of a surface wave near the wavenumber position in the refractive spectrum of the strong oscillator, where $n = 1$.

Such a surface wave can only be excited in special situations like Otto[37] arrangement or for a grating geometry, because the wave vector of the surface wave is greater than the corresponding vector of the electromagnetic radiation in the vacuum. Therefore, the evanescent wave, connected with the total reflection, must be combined with the reflection at the surface, using an air gap of suitable distance. This arrangement is well

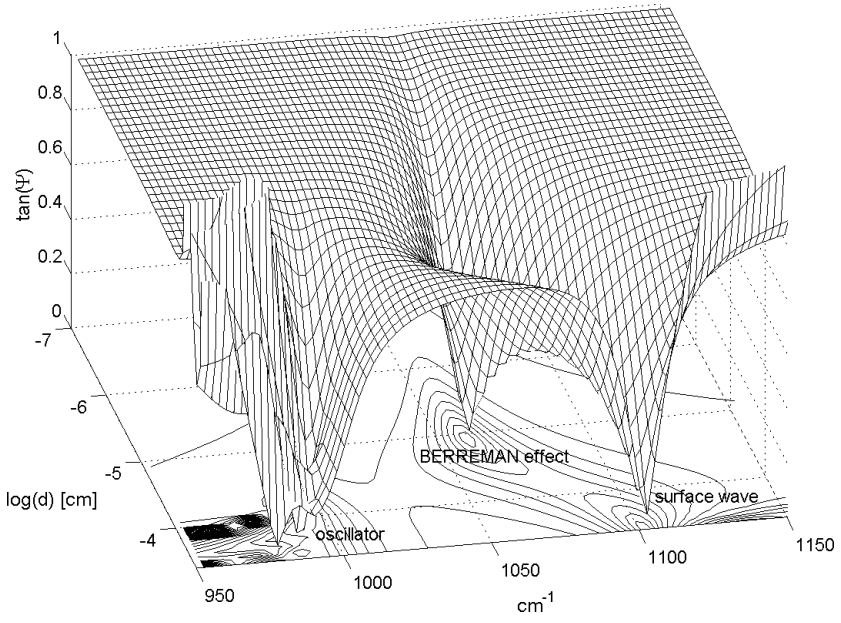


Figure 11.13 Modeling of the Berreman effect and of the excitation of a surface wave.

described by the algorithm of thin film optics. The dispersion relation of the surface wave without damping is given by[38, 39]:

$$k_x^2 = (2\pi\tilde{\nu} n_0 \sin\varphi_0)^2 = (2\pi\tilde{\nu})^2 \frac{\varepsilon_a \varepsilon_b}{\varepsilon_a + \varepsilon_b}, \quad \text{with the condition } \varepsilon_a \varepsilon_b < 0 \tag{11.38}$$

$\varepsilon_{a,b}$ = real part of the media a and b.

One of the joined materials must have a negative ε', which means that the surface wave is only observed in combination with a metallic substrate or for dielectrics in the neighborhood of strong oscillators (reststrahlen bands). In the later case the film has a wavenumber region where $n < 1$. In combination with the ambient air, which acts like an air prism, we have the situation of the Otto arrangement. The thickness of the film in the range of some micrometers corresponds to the air gap in the Otto arrangement. The wavenumber position of the surface wave is given by Eq. (11.38) in connection with the spectrum of the strong oscillator. From the condition in Eq. (11.38), it is not possible to observe a surface wave at the semiconductor surface ($\varepsilon' > 0$) in this arrangement.

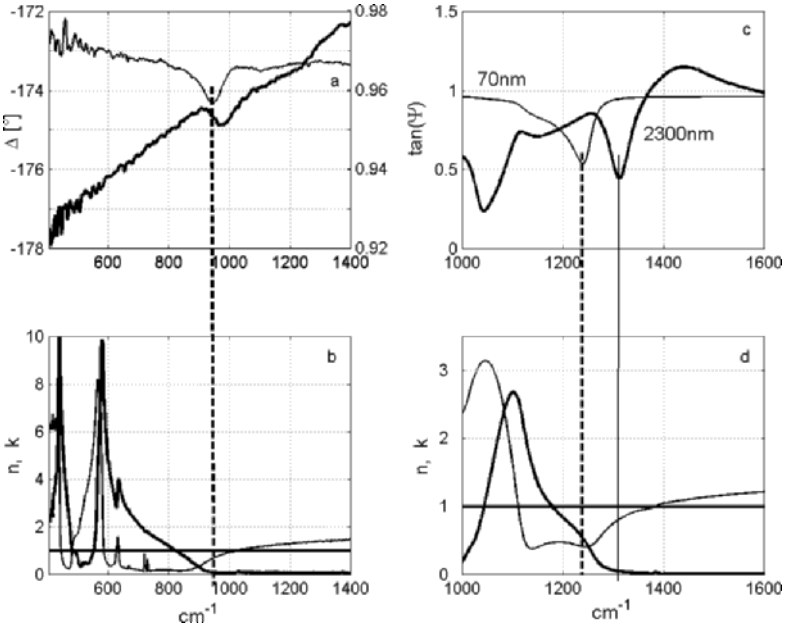


Figure 11.14 Measured Berreman effect and surface wave on an aluminum substrate. The native aluminum oxide film on an aluminum mirror together with the spectra of the optical constant of an Al_2O_3-ceramic are shown in a) and b). Spectra of two SiO_2 films of different thickness (70 and 2300 nm) on an aluminum mirror as an example for the Berreman effect and for the surface wave excitation are shown in c) and d) together with the quartz glass spectrum for comparison.

An example of the Berreman effect is shown in Fig. 11.14a, b for the native oxide film of an aluminum mirror together with the n, k-spectrum of an aluminum oxide ceramic. In this case the Berreman shift is large as 300 cm^{-1}. Figure 11.14c shows the response of two evaporated SiO_2 films with different thickness on an aluminum substrate and the n, k-spectrum of quartz glass for comparison. The film of 70 nm thickness exhibits the Berreman effect near the $n = k$ wavenumber position of the quartz glass spectrum (Fig. 11.14d). The response of the film with 2300 nm thickness is connected with the excitation of a surface wave in the neighborhood where $n = 1$ in the reference spectrum of quartz glass. It seems that the optical properties of the evaporated film and the quartz glass are not identical, since they result in different theoretical wavenumber positions.

Some special cases of the anisotropic film substrate combination have been evaluated. In general, the ellipsometric response depends on the angle of incidence in relation to the Brewster angle and on the film and

the substrate refractive indices. The film thickness and the wavenumber position of the oscillator in relation to the $\lambda/4$ and $\lambda/2$ positions of the interferences in the spectrum lead to changed ellipsometric features[40]. Therefore, it is recommended that each ellipsometric arrangement be proved by a separate calculation.

11.3 References

1. J. R. Beattie, *Philos. Mag*., **46**, 235 (1955).
2. B. Drevillon, J. Perrin, R. Marbot, A. Violet, and J. L. Dalby, *Rev. Sci. Instrum*., 53, 969 (1982).
3. M. J. Dignam and M. D. Baker, *Appl. Spectrosc*., **35**, 186 (1981).
4. A. Röseler, *Infrared Phys.,* **21**, 349 (1981).
5. A. Röseler, Infrared Spectroscopic Ellipsometry, Akademie Verlag, Berlin (1990).
6. A. Röseler, *Thin Solid Films,* **234**, 3007–313 (1993).
7. A. Röseler and W. Molgedey, *Infrared Phys*., **24**, 1 (1984).
8. B. Drevillon, *Thin Solid Films,* **313**, 625–630(1998).
9. E. Wold and J. Bremer, *Applied Optics,* **33**, 5982–5993 (1994).
10. M. Born and E. Wolf, *Principles of Optics,* Pergamon Press, Oxford (1980).
11. M. Luttmann, J. L. Stehle, C. Defranoux, J. P. Piel, *Thin Solid Films,* **313**, 631–641 (1998).
12. A. Röseler, *J. Opt. Soc. Am*., **A9**, 1124 (1992).
13. K. Forcht, K. A. Gombert, R. Joerger, M. Köhl, *Thin Solid Films,* **302**, 43 (1997).
14. Firmware, BRUKER Optik, Ettlingen, Germany.
15. E. Schweizer, J. Nagel, W. Braun, E. Lippert, and A. M. Bradshaw, *Nucl. Instrum. Methods Phys. Res. A,* **239**, 630 (1985).
16. G. P. Williams, *Nucl. Instrum. Methods Phys. Res. A,* **291**, 8 (1990).
17. C. Bernhard, R. Henn, A. Wittlin, M. Klaser, T. Wolf, G. Muller-Vogt, C. T. Lin, M. Cardona, *Physical Review Letters,* **80**, 1762–1765 (1998).
18. J. Kircher, R. Henn, M. Cardona, P. L. Richards, *J. Opt. Soc. Am B,* **14**, 705–712 (1997).
19. U. Schade, A. Röseler, E. H. Korte, M. Scheer, and W. B. Peatman, *Nucl. Instrum. Methods Phys. Res. A,* **455**, 476 (2000).
20. E. E. Bell, *Handbuch der Physik, Optical Constants and their Measurement,* Ed. Flügge, Bd 25a, Springer Verlag, Berlin/Heidelberg/New Yor (1967).
21. A. Röseler, E. H. Korte, *J. Mol. Struct.,* **349**, 321. (1995).
22. T. E. Tiwald, D. W. Thompson, J. A. Woollam, S. V. Pepper, *Thin Solid Films,* **313**, 718–721 (1998).
23. A. Röseler, E. H. Korte,. In *Handbook of Vibrational Spectroscopy,* J. Chalmers, P. R. Griffiths, Eds., Wiley: Chichester (2001).
24. A. R. Reinberg, *Appl. Opt.,* **11**, 1273 (1972).
25. A. Röseler, E. H. Korte,. *Thin Solid Films,* **313–314**, 708 (1998).
26. A. Röseler and E. H. Korte, *Thin Solid Film*s, **313–31**4, 732 (1998).

Next Page

27. D. den Engelsen, *J. Opt. Soc. Am.*, **61**, 1460 (1971).
28. R. M. A. Azzam and N. M. Bashara, *Ellipsometry and Polarized Light,* North-Holland, Amsterdam (1977).
29. M. Schubert, *Phys. Rev. B,* **53**, 4265–4274 (1996).
30. D. W. Berreman, *J. Opt. Soc. Am.*, **62**, 502 (1972).
31. D. W. Berreman, *Phys. Rev.,* **130**, 2193, (1963).
32. D. W. Berreman, In *Proceedings of the International Conference on Lattice Dynamics, Copenhagen,* Wallis, R. F., Ed., Pergamon Press: Oxford, 397–402 (1965).
33. B. Harbecke, B. Heinz, P. Grosse, *Appl. Phys.*, **A38**, 263 (1985).
34. D. Tsankov, K. Hinrichs, A. Röseler, E. H. Korte, *Phys. Stat. Sol. (a)*, **188**, 1319–1329 (2001).
35. M. Schubert, T. E. Tiwald, C. M. Herzinger, *Physical Review B*, **61**, 8187–8201 (2000).
36. H. Angermann, W. Henrion, A. Röseler, M. Rebien, Wet-Chemical Passivation of Si(111)- and Si(100)-Substrates, E-MRS, Strasbourg, France, 1.-4.6.1999, *Materials Science and Engineering* **B73**, 178 (2000).
37. A. Otto, *Z. Phys.*, **216**, 398 (1968).
38. A. A. Maradudin, Surface Waves, Advances in Solid State Physics, Ed. Treusch, Vieweg, Braunschweig 1981.
39. G. N. Zhizhin, E. A. Vinogradov, M. A. Moskalova, V. A. Yakolev, *Appl. Spectr. Reviews,* **18**(2), 171 (1982/83).
40. R. Ossikovski, B. Drevillon, M. Firon *J. Opt. Soc. Am A,* **12**, 1797–1804 (1995).

12 Ellipsometry in Life Sciences

Hans Arwin

Laboratory of Applied Optics
Department of Physics and Measurement Technology
Linköping University, Sweden

It's never too late

He used to be the saddest ray
Excitement wasn't in his way
Variation none
Polarizer gone
and thus he missed his yesterday

But then one day, the quarter wave plate
Decided to take part in his fate
He just let him in
Taught him to spin
and said: my ray, it's never too late!

Anja Wikström

Dedication:
To Roger Jansson († November 1, 2002)
for his invaluable contributions to our
ellipsometry research during
20 years

12.1 Introduction

Ellipsometry is widely used for analysis of materials and processes on surfaces and in thin films. The majority of the materials studied are metals, semiconductors, polymers, etc. and the scientific questions addressed are normally not biologically related. In this chapter we will address applications of ellipsometry in the area of life sciences. Also here numerous applications are found especially for adsorption studies of biomolecules at interfaces, and the number of investigations in which ellipsometry is used as a key tool to address biorelated issues is very large – probably in the order of thousands or more. The most common application is to monitor and quantify the amount of organic material adsorbed at a solid/liquid or at a solid/air interface. This is very simple from a measurement point of view, but the interpretation is not always straightforward.

With this brief background, we formulate the overall objectives of this chapter as: (1) to consolidate that ellipsometry really can provide key information in many areas within life sciences, and (2) to bring forward that the potential is far from being fully utilized considering the recent instrumental and methodological development. The life science area is very large and a limitation is made here to studies of biomolecules at interfaces with a few other types of applications only briefly mentioned. A main reason is that adsorption at optically reflecting interfaces is one of the major strengths of ellipsometry due to its thin film sensitivity and *in situ* measurement possibilities.

The ellipsometer studies reported within life sciences address a variety of phenomena, which can be sorted into type of biomolecule(s), surface, ambient, or type of application. In a systematic approach a large matrix would be needed to fully map out the full picture. Table 12-1 is an attempt to classify different types of experiments involving surfaces and biomolecules according to type of exposure to biomolecules and the corresponding phenomena addressed. Later in this chapter a few of these applications will be exemplified in some detail. A large number of different surfaces are being used depending on the area of study. Among model surfaces, oxidized silicon is the most common, but metals like gold, chromium, and titanium and its oxide, etc. are also frequently used. Often the surface energy and surface chemistry of these surfaces are chemically modified. Also their microstructure is altered, e.g. by making them textured or porous, but very few ellipsometric studies on real biosurfaces are found. In sensor applications, surfaces are often made selective and/or more sensitive by chemical or biological functionalization. The various types of experiments can be carried out in air or *in situ* in a liquid. A measurement *in situ* offers several advantages and should always be considered especially if the adsorption kinetics is required. Finally it is possible to distinguish

Table 12-1. Classification of ellipsometer applications after type of exposure to biomolecules.

Type of surface exposure	Phenomena addressed
One type of biomolecule	Quantification of surface mass Adsorption kinetics Basic interaction mechanisms Layer microstructure Adsorption modeling
Two or more types of biomolecules simultaneously	Competitive adsorption Replacement (Vroman effect)
Two or more types of biomolecules in sequence	Multilayer formation Biointeractions - protein-protein - antigen-antibody Enzymatic degradation Replacement (Vroman effect) Cleaning
Complex solutions	Biological response - coagulation - complement

between different types of applications. Most of the investigations can be described as aiming towards a basic understanding of interaction mechanisms between biomolecules and surfaces. Recently sensor applications have been demonstrated and their commercialization is in progress. Also contributions to the emerging area of nanobiotechnology are seen and ellipsometry is expected to be invaluable in future development in this area.

The presentation here is far from complete but the outline is to first discuss some special issues related to surfaces that interact with biological molecules and to present the unique features of ellipsometry in this context. This will hopefully serve as a motivation for employing ellipsometry. The methodology is then described using a handbook approach, followed by a brief review of the major application area, that is, monitoring of bioadsorption by discussion of a handful of representative examples and by showing how these types of experiments can be varied. A more detailed discussion about some selected and more advanced applications addressing the determination of optical functions and microstructures of biolayers is also included. Finally some emerging application areas utilizing new and/or

advanced techniques and methodology like spectroscopy, imaging, infrared, and generalized ellipsometry as well as advanced modeling will be discussed.

12.2 Historical Background

Tronsted, et al. mentioned in 1932 and later reported ellipsometric measurements on monolayer films of long-chain fatty acids on mercury.[1] In fact he used the independent knowledge of the length of the fatty acid chains to test the validity of the Drude formulations for thin transparent films on metal surfaces. Among the pioneering works are the immunological studies by Rothen[2] and studies of kinetics of protein adsorption by Vroman, et al.[3] In a review from 1972, Poste and Moss[4] describe some of the early works including studies of cell coat materials. Another review of ellipsometry on organic films was presented by Debe.[5] At the seven conferences on ellipsometry, from the 1963 Nebraska meeting up to and including the second international conference on spectroscopic ellipsometry (ICSE-2) in Charleston, there have been only a few contributions that address biomolecule/surface interactions and this field has been dominated by applications and developments related to semiconductors and metallic materials. The developments and challenges of ellipsometry in life sciences have been summarized by the author at ICSE-2 and also at other conferences.[6–8] However, even though ellipsometry appears to find applications in life sciences over the years, it has only been used by a small number of research groups and mainly in very simple types of applications. An increasing interest in ellipsometry for more advanced applications is seen in physical chemistry,[9] but it may be fair to cite Poste and Moss from 1972 and admit that "It is perhaps reasonable to suggest that ellipsometry has yet to play its most significant role in the study of biological surfaces".[4] We are now moving into the era of nanotechnology and it seems that nanobiotechnology will provide important contributions to this development. Perhaps now is the opportunity for a real breakthrough for "bioellipsometry" due to its possibilities for true quantitative measurement with nm sensitivity and *in situ* advantage to kinetically following events at the solid/liquid interface.

12.3 The Interfaces Under Study

For the interfaces under study, we must be aware of potential pitfalls and take special precautions to obtain high quality experimental data, firstly because we frequently monitor layers of nm thickness which are very sensitive to contamination and secondly because ellipsometry does

not provide any chemical information[†] and we can therefore seldom determine what substance we actually measure on.

The simplest interface from a measurement point of view is the **solid/air interface.** The ellipsometric measurement is then similar to applications in traditional studies of inorganic samples, e.g., semiconductors and metals, and all experimental possibilities like spectroscopy, variable angle of incidence, etc. are available. However, the awareness of potential problems is crucial. A representative example of a study is the investigation of adsorption of protein molecules onto oxidized silicon or onto a metal surface. Of interest may be how the adsorption depends on protein concentration, pH in solution, surface treatment or functionalization. In most cases a layer with thickness less than a few nm is adsorbed and the surface must be well characterized before the adsorption takes place. In accurate studies this must be done for each sample individually. Among the problems to handle is then to avoid *contamination* due to transfer of the surface through the air/liquid interface. It requires extreme care to avoid the deposition of small molecules and denatured proteins that are present at the surface of the liquid. A transfer of the sample into the liquid is, however, necessary because the protein adsorption occurs from the liquid phase. Much of the contamination will desorb due to rinsing and drying of the sample but an uncertainty is introduced. Another uncertainty due to the rinsing is that it is hard to know whether a *steady state* is probed or not in the subsequent measurement. Some molecules may also desorb during the rinsing phase. Another uncertainty is the drying process. Often the samples become blown with dry air or nitrogen but it is not easy to know how much *bound water* still remains in the adsorbed biolayer. Finally errors may be introduced due to small differences in sample alignment and difficulties in finding exactly the same spot on the surface before and after adsorption of biomolecules. All the above-mentioned practical difficulties may be considered as small but recall that a resolution well below a nm is aimed. In practice one often repeats the experiments many times and bases the conclusions on statistics.

For measurements at the **solid/liquid interface** it is necessary to have a cell that contains the surface and the surrounding liquid. This will be discussed in more detail later in this chapter. Due to the cell, measurements on surfaces in a liquid are subject to more limitations. The cell windows and the liquid itself must be transparent at the wavelength(s) used and the angle of incidence is in most cases fixed. We have to consider that adsorption of molecules always implies *replacement* of other molecules. If the solvent is water, the adsorbing molecules replace water and eventually also adsorbed ions and/or other constituents of the solvent. The interface is also an electrochemical interface and effects like contact adsorption of ions,

[†]In the emerging area of IR-ellipsometry on biolayers this may change.

charge depletion layers in the solid, etc. may be potential factors that influence the optics of the interface. An adsorbed layer on such an interface may be dense and rigid or extended and very hydrated or something in between. However, one of the major advantages of measuring *in situ* is that a baseline can be determined prior to adsorption. Several systematic errors will then affect the final results only to a second order. It is also possible to measure on the same spot on the surface during the whole experiment.

The **ambient** must also be considered. In most ellipsometric measurements the ambient is air or vacuum and the ambient index is unity. For measurements in a liquid one has to know or determine the refractive index spectrum of the liquid and the effects on the refractive index of molecules that are dissolved in the liquid must be checked. In addition, as mentioned above, the optics of the ambient phase close to the interface may deviate from that of the bulk. If additional liquid is added to the cell, mixing may introduce refractive index variations resulting in noise. Also small particles in the liquid may be a source of noise.

12.4 From Optics to Biology

12.4.1 The Unique Possibilities

It seems from the previous section that a lot of experimental difficulties are introduced by applying ellipsometry into the life sciences. However, unique possibilities are opened up. First of all, ellipsometry has a thickness resolution of the order of parts of a nm which implies that it is possible to address issues related to layers with molecular dimensions. Ellipsometry is also a photons-in photons-out technique and is thus non-destructive and allows *in situ* studies at the solid/liquid interface. This advantage should not be underestimated in life science applications because it (1) allows layers to be studied in model systems very closely mimicking their normal environment in biological systems, (2) the dynamics of layers can be studied directly, and (3) no molecule labellings are required. These three advantages allow direct studies of key phenomena such as competitive adsorption of macromolecules, molecular interaction on surfaces, and exchange reactions and more as mentioned in Table 12-1. With special cell designs, the effects of flow rate on macromolecular adsorption can be investigated to mention additional possibilities. Taken all together ellipsometry provides a convenient tool to study surface dynamics in biological systems. This is a major advantage since biological systems rarely rest – a living system is by definition on the move and is based on chemical reactions, transport, synthesis, and degradation.

12.4.2 Verification of Ellipsometric Results

Ellipsometry is an indirect technique that measures polarization changes upon reflection. It is therefore fully understandable that a layman is skeptical about ellipsometric data, especially when presented as the ellipsometric angles Δ and ψ. However, independent experimental verifications are not simple because very few methods provide information with ellipsometric resolution from the solid/liquid interface. As we deal with soft matter, comparisons using electron microscopy, like those for semiconductor samples, are virtually impossible. However, for one of the most basic outputs from an ellipsometric study of macromolecular adsorption, the surface mass density of the adsorbed layer, a few comparisons with radiolabeling techniques have been carried out.[10–12] Figure 12.1 shows one example from such a quantitative comparision using the protein human serum albumin (HSA). Even though a lot of different factors may explain the observed differences, it is clear that the methods agree very well.[11] In fact one cannot judge which one is the most accurate method of the two. The conclusion is that for the determination of surface mass density or mass per unit area of thin biolayers, ellipsometry provides data corresponding well to the state of the art in measurement technology.

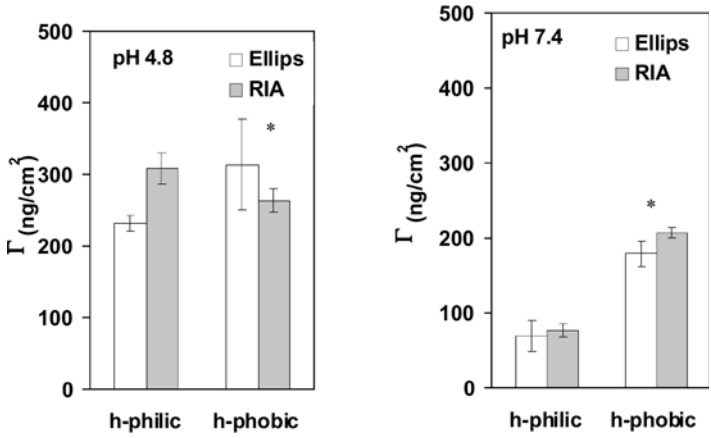


Figure 12.1 Adsorbed surface mass density Γ of human serum albumin on hydrophilic and hydrophobic oxidized silicon at pH 4.8 and 7.4 as determined by ellipsometry and radioimmunoassay techniques (RIA). Reprinted with permission from Ref. 11. Copyright (2000) Elsevier Science.

12.5 Methodology for Data Evaluation – from ψ and Δ to Biologically Related Parameters

The ellipsometric measurement provides a data set of (ψ, Δ) pairs. Among the most common experimental variables are wavelength, angle of incidence, time, and spatial position on the sample, which correspond to spectroscopic, variable angle, dynamic, and imaging ellipsometry, respectively. However, in simple routine measurements, often only one (ψ, Δ) pair per sample is determined. From ψ and Δ, the parameters of interest are obtained by evaluation following the normal procedures in ellipsometric analysis. In the context here typical optical parameters are the thickness d, the refractive index $N = n + ik$, and the fill factor f of a layer, often with thickness in the 1 to 30 nm range. A complementary parameter that is often used is the surface mass density Γ also called the surface excess. However, using Γ as the final result implies a reduction of the information as it combines thickness and refractive index into one parameter. Nevertheless, the surface mass density is conceptually very clear and is very useful in many cases. It is therefore well motivated to include Γ in the discussion.

We will describe, with increasing complexity, some different strategies to evaluate ellipsometric data. A limitation is made to layers of biomolecules on solid substrates and three cases are discussed: a thin layer on a flat ideal substrate, a thick layer on a flat ideal substrate, and adsorption into porous layers. Finally surface roughness, the use of dispersion models in the analysis and effects of anisotropy, will be briefly discussed. The assumption is in all cases that the substrate is well known or can be characterized by ellipsometry before the biolayer formation.

12.5.1 A Thin Biolayer on a Flat Ideal Substrate

Adsorption of biomolecules to a flat surface is one of the most common measurement situations. One representative example is adsorption of the protein ferritin to gold as shown in Fig. 12.2.[13] Two experiments using different protein concentrations in solution are presented as changes in Δ with time. The change in Δ from the baseline, obtained before the addition of protein, to the steady state level at the end of the experiments provides information about the adsorbed surface mass (or thickness of the layer). We see immediately that the adsorption kinetics is different for the two concentrations and that the total change in Δ is smaller for the lower concentration. In such dynamic studies the adsorption can thus be followed from an empty surface until steady state, normally a partially filled surface. When the surface is completely filled we talk about formation of a protein monolayer. The definition of a monolayer is not fully clear as the maximum number of

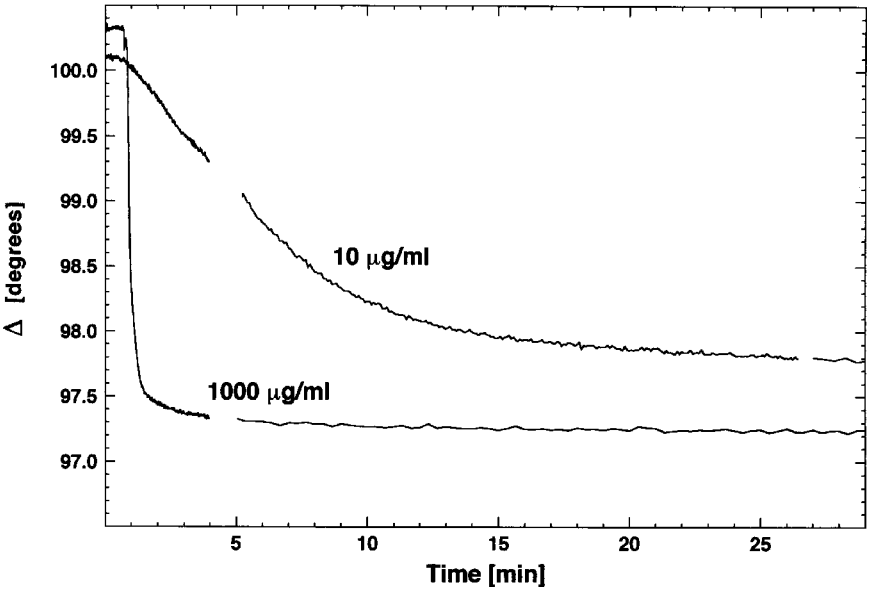


Figure 12.2 Change in Δ versus time during adsorption of ferritin on gold at λ = 620 nm at two different concentrations of ferritin in solution.[13]

molecules adsorbing per unit area depends on the surface chemistry, protein concentration, and even on the way the adsorption is carried out. The orientation of the molecules can also vary in different types of experiments. Even though there still exist "empty slots" on the surface, additional molecules may not fit into these due to no or limited lateral mobility of the already adsorbed molecules and electrostatic repulsion. Some proteins can also form bilayers or multilayers but these effects will not be discussed here.

In the region below monolayer surface coverage we can illustrate the adsorbed layer with a simple physical picture as in Fig. 12.3. This is an oversimplification but will serve here as our model for reality to be compared with our basic models for analysis also given in Fig. 12.3. The two models in Fig. 12.3 represent two extremes. In one of them (top right) the layer thickness corresponds to the physical extension of the biomolecules into the ambient medium with a refractive index corresponding to the equivalent index of a layer composed of biomolecules and the ambient. In the other extreme (bottom right) the thickness is an equivalent thickness of a smeared out (collapsed) layer and the refractive index is the intrinsic index of the biomolecules. We denote these parameters d and N in the first model because they are properties of the layer itself. In the second model we use d_{eq} and N_i. For an ideal compact layer $d = d_{eq}$ and $N = N_i$. Obviously

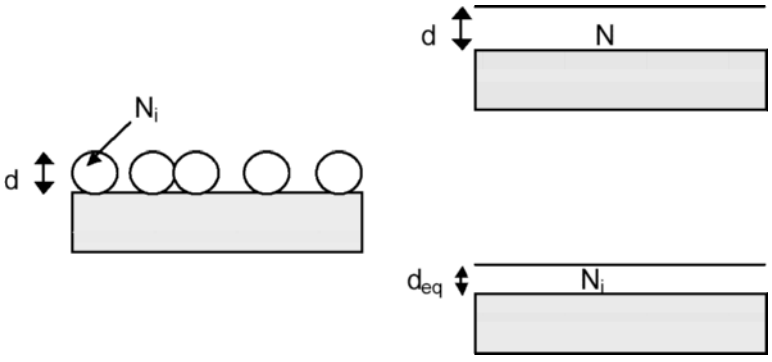


Figure 12.3 Simplified physical model of an adsorbed layer of spherical protein molecules with intrinsic refractive index N_i and diameter d on a flat ideal surface (left) and two ideal three-phase models (right).

we have a few conceptual problems and some immediate thoughts are: How do we explain a thickness value of, say, $d_{eq} = 4$ nm if the adsorbed macromolecules in the partially filled layer have a diameter of 9 nm? What about microstructure? Will the arrangement of the molecules affect the thickness value? What about form birefringence? Should the analysis be based on an anisotropic layer model? Should one use the monolayer thickness and vary the refractive index during the formation of the layer, use a constant index and vary the thickness, or even vary both index and thickness?

Many of these questions may be irrelevant, however, if we consider the case when the resolving power of the instrument in the experimental situation studied is insufficient to allow determination of both N and d. This is in fact very common in thin film analysis in general and can be understood by considering the phase factor β

$$\beta = \frac{2\pi d}{\lambda} N \cos\phi_1 \tag{12.1}$$

where λ is the light wavelength and ϕ_1 the angle of refraction in the layer. This is the only formula where d enters in the ellipsometer expressions in the ideal three-phase model. Furthermore it enters as a product with N which implies that only the product dN can be determined. N also enters into the Fresnel coefficients in the overall reflection coefficients but will only influence ψ and Δ to second order for small d. We conclude that for very thin layers, only the product dN can be determined. For a transparent layer this reduces to the product dn, which sometimes is referred to as the optical thickness. Let us now systematically describe some different strategies for evaluation of ellipsometric data on thin films bearing in

Table 12-2. Overview of methods to evaluate ellipsometric data on thin films of biomolecules.

Method	Description	Input	Independent parameter	Output	Comments
PRIM	Primary data used as result	Δ and ψ		d or Γ in a.u.	The thin film approximation for linearity is assumed. Provides only relative values.
THICK	Thickness only	Δ and ψ	N_i N	d_{eq} d	
GTI	General thickness and index	Multiple Δ and ψ	–	d and N	Optical model parameters obtained only.
FILL	EMA analysis	N	N_i	f	N from e.g. a GTI analysis.
CSM	Surface mass	d d and N	ρ_{1ay} ρ_i and N_i	Γ Γ	
FEIJ	de Feijters	d and n	dn/dc	Γ	dn/dc independent on c is assumed.
CUYP	Cuypers	d and n	A, M, V_{20}	Γ	Based on Lorentz-Lorenz.
COMP	Stenberg	Δ	A, M, ρ_i, d_{ref}, n_{ref}[a]	Γ	The thin film approximation for linearity is assumed.

[a] If silicon dioxide is used $d_{ref} = d_{ox}$ *and* $n_{ref} = n_{ox}$

mind the difficulty of decoupling d and N. We will find that the use of surface mass density Γ will save us in many cases. A three-phase model is assumed throughout our analysis but the same procedures are applicable in a general n-phase model. The different strategies are refered to by acronyms, PRIM, CSM, GTI, etc. and are summarized in Table 12-2.

12.5.1.1 Using Primary Data Only (PRIM)

In some applications it is sufficient to compare Δ-values (or ψ-values) between different experiments. The change in Δ from the baseline is then the measure of adsorption. The rate of adsorption or steady state information

can be evaluated without determination of absolute values of d or Γ. The underlying assumption is here that the change in Δ is linear with Γ. The data in Fig. 12.2 illustrate this and one can conclude that the kinetics of adsorption was faster for the higher protein concentration than for the lower and the relative difference in the steady state value of Γ can be quantified. This type of evaluation was extensively used before powerful evaluation programs became readily available.

12.5.1.2 Determination of Thickness Only (THICK)

The film thickness d can be obtained in the three-phase model by use of an independently determined or a literature value on N. This is a straightforward approach but in most cases the absolute accuracy is low and unknown because of the uncertainty in N. The interpretation of the meaning of d is also unclear. If the value of N corresponds to the intrinsic N_i of the macromolecules (assuming them to be large enough to possess their own dielectric identity), the value of d will correspond to the thickness d_{eq} of an equivalent homogeneous layer of the molecules smeared out on the sample surface - a collapsed layer. On the other hand, if N corresponds to the effective index of an equivalent layer including the macromolecules, the ambient medium, and eventually contacts adsorbed ions, then d represents the extension of this layer into the ambient. The two models are illustrated in Fig. 12.3.

12.5.1.3 Determination of Both Thickness and Refractive Index (GTI)

In very accurate experiments it is possible to obtain both N and d using the standard three-phase models or the n-phase model for more complicated samples. Often more advanced methodologies like spectroscopic ellipsometry, variable angle ellipsometry, variable angle spectroscopic ellipsometry, multiple samples, or multiple ambients are used. Also dispersion models can be included to reduce the number of unknown parameters. Figure 12.4 illustrates one type of analysis and shows the dielectric function ε and N versus photon energy of a layer of the protein bovine serum albumin (BSA) as determined by spectroscopic ellipsometry assuming a transparent layer ($k = 0$) for long wavelengths.[14] The thickness determined in this way is to be considered as a representation of the extension of the layer and the index is the equivalent layer index. At the high energy end (short wavelengths) the UV-absorption of the protein is resolved and seen as a finite value of the imaginary part ε_2 of the dielectric

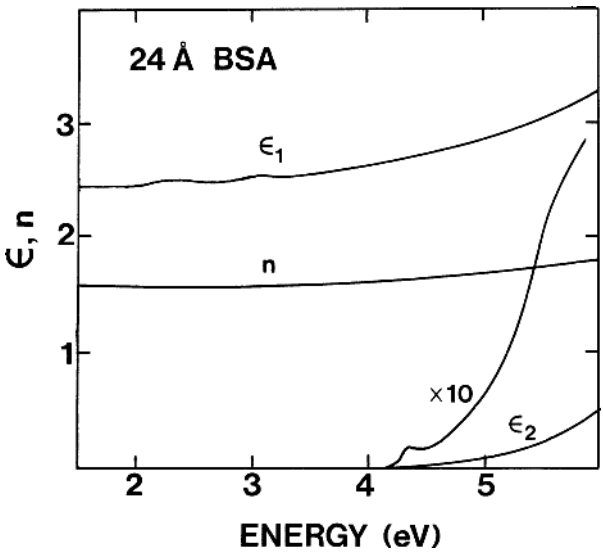


Figure 12.4 The complex dielectric function $\varepsilon = \varepsilon_1 + i\varepsilon_2$ and the real part n of the refractive index of a 24 Å layer of bovine serum albumin (BSA) adsorbed on a HgCdTe substrate. The shown ε_2 is also magnified by a factor ten.[14]

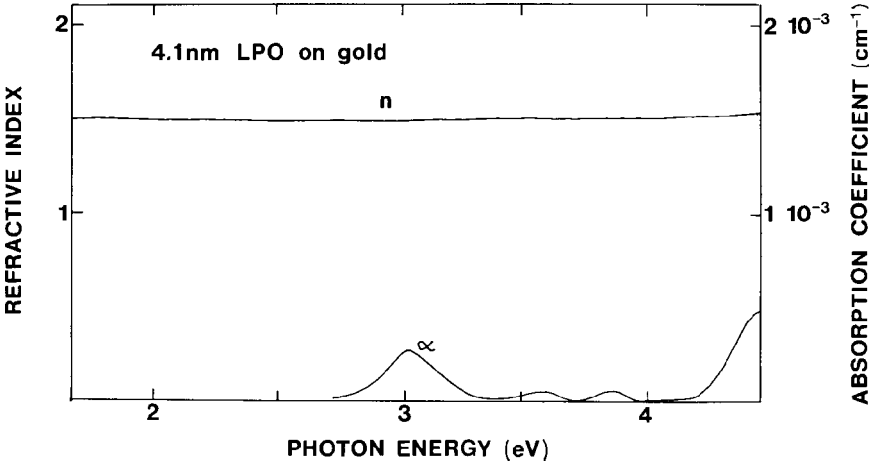


Figure 12.5 Ellipsometrically determined refractive index n and absorption coefficient α for a 4.1 nm layer of lactoperoxidase (LPO) measured *in situ* on gold.[15]

function. In Fig. 12.5 a similar example is shown for a layer of lactoperoxidase on gold determined *in situ* at a solid liquid interface.[15] Here the absorption coefficient $\alpha = 4\pi k/\lambda$ is also shown and the absorption of the heme group of the protein is seen around 3 eV.

It has also been shown that both n and d can be determined using analytical inversion resulting in a fifth degree polynomial equation. The derivation is, however, restricted to cases with real values on the film refractive index ($k = 0$). This methodology does not require numerical fitting procedures and was applied at the air/water interface to layers of arachidic acid and valine gramicidin A with thickness in the range of 2 to 3 nm.[16]

12.5.1.4 Determination of Volume Fill Factor (FILL)

If the equivalent index N of a layer can be measured and the intrinsic index N_i is known, it is possible to determine the fill factor f of the layer by using simple volume averaging or by the use of effective medium theories. In simple geometries we would have

$$\frac{1}{\varepsilon} = \frac{f}{\varepsilon_i} + \frac{1-f}{\varepsilon_{amb}} \tag{12.2a}$$

or

$$\varepsilon = f\varepsilon_i + (1-f)\varepsilon_{amb} \tag{12.2b}$$

where we have introduced the dielectric function $\varepsilon = N^2$ with ε, ε_i and $\varepsilon_{amb} = n_{amb}^2$ corresponding to dielectric functions of the equivalent layer, the layer material and the ambient, respectively. Equation (12.2b) here corresponds to simple volume averaging. Equations (12.2a) and (12.2b) are, however, strictly valid only in simple geometries where the electric field in the probe beam is perpendicular and parallel, respectively, to the interfaces between two media. This is normally not the case in ellipsometric measurements at oblique incidence and certainly not for layers containing molecules of irregular shape. As this is a case of form birefringence an anisotropic approach would be appropriate but for very thin layers the anisotropy is hard to resolve. The use of Eqs. 12.2 is sometimes seen but should be avoided. It is then better to employ an effective medium approximation (EMA), and the Bruggeman theory is most frequently used.[17] The fill factor is then obtained from

$$f\frac{\varepsilon_i - \varepsilon}{\varepsilon_i + 2\varepsilon} + (1-f)\frac{\varepsilon_{amb} - \varepsilon}{\varepsilon_{amb} + 2\varepsilon} = 0 \tag{12.3}$$

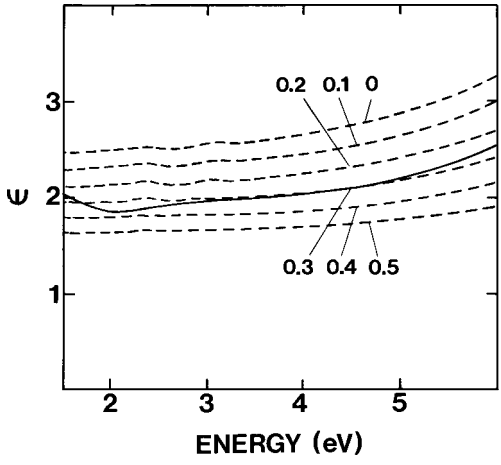


Figure 12.6 Real part of the dielectric function (solid curve) of a 24 Å thick bovine serum albumin layer on platinum determined by spectroscopic ellipsometry. The dashed curves show calculated ε_1 obtained with the Bruggeman EMA assuming void fractions as given by the numbers at each curve. The top curve (void fraction = 0) is identical with that in Fig. 12.4.[14]

A simple calculation shows that if f is solved for in Eqs. (12.2a), (12.2b), and (12.3) for, e.g., a value of $\varepsilon = 2$ ($n = 1.414$) using $\varepsilon_{amb} = 1$ and $\varepsilon_i = 2.25$, one finds f equal to 0.9, 0.8, and 0.89, respectively. We notice that f from the EMA approach is close to that from Eq. (12.2a), that is the case with interfaces perpendicular to electric field. In addition this is close to the real situation in ellipsometric measurements on thin partially filled layers (submonolayers) as the s-component and most of the p-component of the probe beam are parallel to the sample surface.

Figure 12.6 shows one example of how the density deficiency ($1 - f$) of a BSA layer on platinum can be determined by Eq. (12.3).[14] As reference data on BSA, that is ε_i, the spectrum in Fig. 12.4 was used and a density deficiency of around 30% was found.

12.5.1.5 Conversion to Surface Mass (CSM)

Some of the difficulties with the correlation between N and d can be avoided if surface mass per unit area Γ is used as the evaluation output. Γ is easy to understand and can be compared with results from other methods like radio immunoassays[11] and gravimetric methods. The drawback with Γ is that microstructural information is lost.

If the layer thickness can be determined as described in Sections 12.5.1.2 or 12.5.1.3 above, one can simple multiply with the layer density ρ_{lay} to obtain Γ,[†]

$$\Gamma = d\rho_{lay}. \tag{12.4}$$

However, the accuracy will never be better than that of d and an uncertainty due to ρ_{lay} is added. Another possibility, if both d and n (or N) are available, is to use EMA methods to determine the fill factor f. This requires that the intrinsic refractive index N_i of the layer material is known. If its density, that is, the intrinsic density ρ_i is also known, the surface mass is obtained from

$$\Gamma = d\rho_i f \tag{12.5}$$

12.5.1.6 Determination of Surface Mass According to De Feijter (FEIJ)

If both d and n can be determined[‡], a common approach to obtain Γ[§] is to use de Feijter's formula[18]

$$\Gamma = \frac{d(n - n_{amb})}{\mathrm{d}n/\mathrm{d}c} \tag{12.6}$$

where $\mathrm{d}n/\mathrm{d}c$ is the refractive index increment with protein concentration in the liquid for the layer material. The assumption is that n of the layer material is a linear function of its concentration. The refractive index increment $\mathrm{d}n/\mathrm{d}c$ can be determined independently by using an Abbe refractometer. In Eq. (12.6), the optical mass of the layer, the product dn, enters in the numerator (with an offset of dn_{amb}). This is very advantageous because often the variations observed in d and n are correlated with a lower d resulting in a higher n and vice versa. This implies that Γ looks less noisy than d or n. In many reports one can indeed observe that the noise level in Γ is substantially reduced compared to that in d or n, especially for very thin layers.

[†]Several units for Γ can be used. If d is given in nm and ρ in g/cm^3 a prefactor 100 is needed in Eq. (12.4) to give Γ in ng/cm^2. Other prefactors will apply for other units (µg/cm^2, mg/m^2, ng/mm^2 etc)

[‡]Only transparent layers with $N = n + i0$ are normally treated in de Feijter formalism.

[§]To obtain Γ in ng/cm^2, a prefactor 100 is needed in Eq. (12.6) if d is given in nm and $\mathrm{d}n/\mathrm{d}c$ in cm^3/g.

12.5.1.7 Determination of Surface Mass According to Cuypers (CUYP)

Cuypers, et al.[19] have developed a method based on d and n which requires independent knowledge about the molecular weight M, molar refractivity A, and partial specific volume V_{20} of the molecules in the layer. Their derivation is based on the Lorentz-Lorenz relation and for a mixed layer (ambient and biomolecules) yields Γ according to†

$$\Gamma = d\rho_{lay} = \frac{0.3}{\dfrac{A}{M} - V_{20}\dfrac{n_{amb}^2 - 1}{n_{amb}^2 + 2}} \frac{(n + n_{amb})}{(n^2 + 2)(n_{amb}^2 + 2)} d(n - n_{amb}). \qquad (12.7)$$

There exists also a one-component formula for a homogeneous layer.

12.5.1.8 Determination of Surface Mass According to Stenberg (COMP)

Stenberg, et al.[20] have developed a method to determine Γ of a thin organic layer on a silicon substrate based on comparison of the ellipsometric response with those of thin silicon dioxide layers on silicon. This can be done directly in a comparison ellipsometer[21] or indirectly in an ordinary ellipsometer by calibration with a set of oxidized samples. The procedure is based on the observation that mainly Δ changes upon adsorption of a thin layer on silicon. By equating Δ for measurements on oxides on silicon and on protein layers on silicon it can be shown[20] that the thin film approximation gives

$$d = \frac{n_{ox}^2 - 1}{n_{ox}^2} \frac{n^2}{n^2 - 1} d_{ox} \qquad (12.8)$$

where n_{ox} and d_{ox} are the refractive index and thickness, respectively, of the oxide. If n is known, d is obtained from Eq. (12.8) and by using the layer density, Γ follows from Eq. (12.4). In case n is unknown, it is suggested to be determined using the Lorentz-Lorenz approach similar to Cuypers, et al.[19] and the following expression for Γ results

$$\Gamma = \frac{1}{3}\left(2\rho_i - \frac{M}{A}\right)\frac{n_{ox}^2 - 1}{n_{ox}^2} d_{ox} \qquad (12.9)$$

† The prefactor 0.3 requires d in nm and V_{20} in cm^3/g.

For a layer of albumin Stenberg et al[20] calculated the prefactor in Eq. 12.9 and with d_{ox} in nm it follows $\Gamma = 0.12\ d_{ox}$ [μg/cm^2]. In the derivation of the Stenberg method it is assumed that the layer is at least a monolayer thick but they propose its use also for submonolayers.

12.5.1.9 Comparison of the Methods

Depending on amount and quality of the available data, different methods can be chosen for evaluation of ellipsometric data on thin organic films. In Table 12-2 a summary of the methods discussed above is given. Before selecting a method the objectives should be defined. The measurements provide the (ψ, Δ)pair(s) that are used to determine optical parameters using an optical model. These parameters are not always of primary interest and are further converted to other parameters.

The most simple method, PRIM, is sufficient in, e.g., sensor applications because an independent calibration must always be performed towards standards. The calibration will then provide a measure of Γ if required. However, often other measures like biological activity, concentration of a reagent in solution or binding constants may be of interest. If direct mass quantification is needed, the COMP method is very stable and easy to use. It can be regarded as a further development of the PRIM method and has been developed for studies of adsorption of biomolecules on oxidized silicon.[20] Applications on other substrates should be possible as well. If quantification is needed, the FEIJ and CUYP methods are usually employed and in most cases FEIJ is used, probably because the refractive index increment dn/dc is assumed to be more reliable than the molar refractivity A. The advantage with FEIJ, CUYP and also with the CSM method is that the noise level in Γ is considerably lower than in n or d due to the fact that the correlation between n and d to a large extent vanishes in Γ. This is especially obvious in *in situ* studies on kinetics of adsorption. The surface mass density Γ is also very easy to understand for a layman compared to n and d and these methods also work with single-wavelength data, which often are the only data available.

If the objective is to perform basic studies of layer optical properties and/or microstructure, spectroscopic data are in general required. Simple density analysis in well-defined systems can be done with the FILL method. The THICK method is seen very often, but if N_i is used, the obtained parameter d_{eq} should be used with care as it is not a physical parameter. The most general method of determining the adsorbed mass is the CSM method in the case where N_i is used. If instrumental resolution and experimental range in

terms of wavelength and angle of incidence are sufficient, the CSM method should be the starting point for more precise analysis.

From a fundamental point of view it is worth mentioning that the CUYP method is based on the Lorentz-Lorenz model, which assumes that the dipoles of the materials in the layers are in vacuum. This is a drawback since the Lorentz-Lorenz model is strictly correct only for low dipole densities. However, the CUYP method is used in spite of this for layers with ambient contents from zero to unity. In the first approximation, the three methods CUYP, FEIJ, and CSM all can be expressed in the form

$$\Gamma = F(n)d(n - n_{amb}) \tag{12.10}$$

where F(n) is a slow function of n obtained from Eqs. (12.5) and (12.7) for the CSM (with the Bruggeman EMA used) and CUYP methods, respectively, whereas F(n) is a constant in the FEIJ method as seen in Eq. (12.6).

12.5.2 A Thick Biolayer on a Flat Ideal Substrate

If the biolayer is thicker, that is, more than a few monolayers, the methodology for data evaluation is more or less the same as for ellipsometry in general and is described elsewhere in this handbook. Examples of such layers are stacked Langmuir-Blodgett films, self-assembled multilayer systems, protein multilayer systems and films deposited in fouling processes. These applications will be discussed later on. For many of the thick layers one must consider thickness and/or density inhomogeneities as well as anisotropic effects.

12.5.3 Adsorption of Biomolecules into Porous Structures

When biomolecules interact with a porous bulk or thin film material, the plane parallel boundary model breaks down. Microscopically, the interaction between the molecules and the surface should be considered as adsorption onto the inner walls of the pores but from a macroscopic (ellipsometric) point of view this is more similar to absorption into a porous matrix. The interpretation should then be based on volume (pore) filling by using effective medium theories. Before the biomolecules are introduced, the porous structure must be properly characterized which normally requires spectroscopic data. The porosity can in the general case be anywhere between 0

and 100% and the EMA of choice is then the Bruggeman theory.[17] One alternative would be to use the Maxwell-Garnett EMA but this works only for small or large porosity values. A minimum of three constituents must be used: the bulk material, the ambient medium and the biomolecules. The Bruggeman EMA can handle three constituents directly if a third order equation is implemented. Sometimes more constituents can be of interest to involve, e.g., if an oxide layer on the interior surfaces of the pores must be modeled. A higher order equation must then be solved which, however, is not straightforward. A simple approach is then to let one of the constituents in the EMA be an EMA itself.

Many porous layers also show gradients in porosity versus depth, which are necessary to model. This can be done in a continuum approach by defining a function describing the porosity profile similar to what is done for compositional profiles.[22, 23] The parameters of this function are then fitted in the modeling of the ellipsometric data. In practice the porous layer is then mathematically subdivided in sufficiently large number of slices and each slice is assigned optical properties using an EMA. Another approach is to model the layer with a discrete number of layers. Each layer is then described by a thickness and a porosity which are the parameters to be determined by fitting to ellipsometric data.[24] As usual in multiparameter fits one has to be careful not to involve more parameters than the amount and quality of the data allows and correlation effects and confidence intervals must be checked.

Another important aspect to consider is anisotropic effects due to form birefringence, which can be of importance when the porous structure lacks a three-dimensional symmetry. One example is porous silicon (PS) prepared by anodization. The microstructure of PS can be controlled by varying the fabrication parameters but most PS samples used to study biointeractions have a branched pore structure with a preferential orientation of the pores along the etching direction. An isotropic model is then sufficient.

When the dimensions of the remaining bulk material become too small one should also be aware of quantum confinement effects, and therefore the bulk optical properties of the material may not be appropriate to use. One example is porous silicon carbide (SiC) for which the optical properties of the solid phase in a porous structure have been found to be considerably different from those of bulk SiC.[25] Quantum confinement effects in porous materials with a preferential pore direction can also lead to intrinsic anisotropic optical properties.

Once a porous layer has been characterized in terms of its microstructure, one assumes this to be invariant during a subsequent exposure to biomolecules. Then only the differences due to adsorption and penetration of biomolecules are modeled. We will return to this in the applications section.

12.5.4 Surface Roughness

Some surfaces can be considered to be very flat, e.g. silicon wafers with a thin native or thermally grown oxide. Also thin metal layers on glass substrates prepared under well-controlled conditions have a very small roughness. However, for real biological surfaces like tissue or bone, the surfaces can be very rough and the interaction with light is viewed more as scattering rather than specular reflection. Most of the light will become depolarized and normal ellipsometry is not easily performed in a quantitative way. In between these extremes there is what we could call technical surfaces with a surface roughness of the order of or larger than the biomolecules. The traditional way to perform ellipsometric measurement is to make use of the polarized part of the specular beam only. The roughness itself is not often modeled but assumed to be taken care of by including its effects in the pseudo-optical properties measured on the surface before biomolecules are deposited.

If the lateral dimension of the roughness is much larger than the wavelength of the ellipsometric probe (a wavy surface), the main effect is that a substantial part of intensity is lost by scattering and only those parts on the surface which locally are parallel with the assumed surface plane will be measured. An angular spread of the beam, depending on the aperture size of the detector, should be included in the analysis.

If the surface has a roughness with dimensions smaller than the interacting molecules they cannot penetrate into the surface region and the surface is then more to be considered as nanoporous. An appropriate approach would then be to model the surface with an interfacial layer using EMA models. The biolayer is modeled as a layer on top of this interfacial layer. Neglecting the interfacial layer and assuming it to be included into the pseudo-optical properties is a first order approximation.

A third case appears when the roughness dimensions are considerably smaller than the wavelength but larger than the dimension of the biomolecules. The molecules then have access to a large surface area. The interface could then be modeled as a thin porous layer into which biomolecules penetrate.

Several investigators have addressed the ellipsometric surface roughness problem[26–29] but in most cases with the objective of clarifying its influence on determination of bulk properties or how to model the roughness with an equivalent layer. The correctness of the various approaches has not been sufficiently proven and much work remains, especially with regard to adsorption of biomolecules to rough surfaces. One clear example that more work is needed is the results by Benesch, et al. showing that the ellipsometrically determined surface mass of HSA on pyrolytic carbon with a root mean square roughness of 20 nm was less than 20% of that determined with RIA.[11]

12.5.5 Use of Dispersion Models

When spectroscopic data are available it is possible to model the optical properties of a layer of biomolecules using dispersion models. For transparent layers the Cauchy model is the basic model. This is an empiric model and is given by

$$n = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} \tag{12.11}$$

where A, B and C are constants fitted in the modeling. In the ultraviolet region one often sees a broad background adsorption in many organic materials. In proteins this is due to the peptide chain. In addition some materials possess functional groups like hemes, porphyrines etc which show resonances that can be modeled with simple Lorentzian lines. This also holds for the infrared region where all the absorption bands in ordinary infrared spectroscopy show up as resonances in the optical functions. The advantage of using dispersion models is that they are parameterizations of the optical functions of the materials and provide filtering. Preferably dispersion models should be applied on optical data after extraction from ellipsometric data, because if used in the ellipsometric evaluation directly, some optical features may be missed and model errors may be overlooked more easily.

12.5.6 Anisotropy

Effects due to anisotropy in thin films of biomolecules are often ignored. One motivation may be a lack of order in the film as the molecules are assumed to be randomly oriented. Another argument used is that even if there exists a certain order, the intrinsic optical properties of the biomolecules are isotropic. It is also stated that the resolving power of ellipsometry on very thin films is not sufficient for the analysis of anisotropy. The arguments above do certainly not apply to Langmuir-Blodgett films of fatty acids, which are both ordered and have intrinsic molecular anisotropy. Several ellipsometric investigations of such systems have been performed and demonstrate the importance of anisotropic modeling.[30, 31] On polymer films optical anisotropy has been resolved.[32, 33] Interestingly, it has been shown that biomolecules like BSA have an intrinsic structural uniaxial anisotropy with refractive indices of 1.744 and 1.563 along the major and minor axes, respectively.[34] This implies that an ordered layer of BSA should be anisotropic which, however, seldom is accounted for in ellipsometric analysis. In conclusion, anisotropic effects in thin biofilms should be considered

more seriously. The theoretical framework is clear and the methodology using generalized ellipsometry can be applied.[35]

12.6 Methodology – Experimental

In this section experimental methodology that is used in biorelated ellipsometric investigations is reviewed. Instrumentation is discussed elsewhere in this book and will not be described in much detail here. Earlier it was common to build ellipsometers in research laboratories and they could then be designed to include a cell for measurements in liquids. Today commercially available instruments are used in most cases and only the cell is home made and the sample holder is adapted for bioadsorption studies. So far there are very few cells available on the market, motivating a brief discussion about cell design. However, most of this section will be devoted to discussions about model surfaces, their treatment, and experimental design to improve data quality. During the years many research groups have worked with biomolecules and ellipsometry but only a few have survived in a longer perspective. This means that the collective experience of the scientific community is very scattered. The experience communicated here is a combination of the experience from the author's research environment and from the literature. It should be regarded as very incomplete and empirical, and a systematic approach to handling of surfaces and biomolecules with respect to ellipsometer studies is welcomed.

12.6.1 Instrumentation

Ellipsometers are often classified as null ellipsometers or photometric ellipsometers. The traditional configuration used in the present context is null ellipsometry in the PCSA (polarizer-compensator-sample-analyzer) setup. In routine measurements performed in air with an objective to quantify surface deposited mass, the PCSA technique is dominating. Some merits of PCSA instrumentation are stability, ease of operation, high resolution, and simple data collection. Several instruments are also offered on the market at affordable prices. Some drawbacks are limitations to a single wavelength and its low speed. When high speed is needed one can operate a PCSA instrument in off-null mode,[36] which, however, is a photometric mode and some of the advantages with the null principle are lost. In more advanced applications involving spectroscopy, photometric instruments like the rotating analyzer configuration are normally employed although null ellipsometry using multiple sources is also available. For imaging ellipsometry, off-null designs

have been described[37] and recently several instruments have become available on the market.

12.6.2 Cell Designs

The possibility of performing *in situ* measurement is one of the major advantages of ellipsometry in the life science area. An appropriate measurement cell is then needed. Some important aspects to consider during cell design are volume, stirring, way to exchange the liquid, angle of incidence, temperature control, window properties, space for electrochemical probes, and sample mounting to mention some of the more obvious. However, experiments can be of so many types that it is hard to design a universal cell.

Here we discuss cell design aspects by relating them to the multipurpose open cell shown in Fig. 12.7. The first and very important aspect is the *volume*. The cell in Fig. 12.7 can be used with volumes down to about 3 ml. This is a large volume if the cost of the molecules under study is an issue. If fast adsorption kinetics is of interest, a large volume significantly

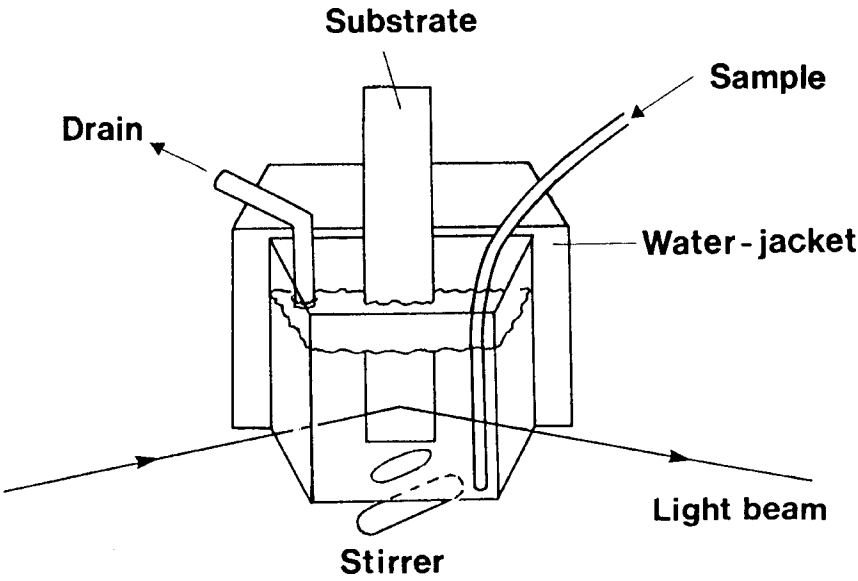


Figure 12.7 Ellipsometer measurement cell for *in situ* studies at the solid/liquid interface. The sample (substrate) is mounted on the sample holder in the ellipsometer and inserted into the liquid.

increases the time before molecules reach the sample surface and stirring efficiency and diffusion are likely to dominate the kinetics. This brings us into *stirring* which is a must. In a typical experiment baselines in ψ and Δ are first established in solution. Then a small volume of the solution containing the biomolecules under study is added into the cell. A homogeneous concentration in the cell is obtained by stirring, e.g., by a rotating Teflon embedded magnet driven by an external rotating magnet as illustrated in Fig. 12.7. The smaller the cell, the faster the mixing can be obtained although the stirring will be technically more complicated. In some applications stirring is solved by use of a flow cell. This will allow a rapid *exchange of the liquid* and with flow cells very small volumes can be used. However, with a flow cell one has to consider the effect of the flow on the interaction between the surface and the biomolecules. Fouling phenomena may occur and it is hard to determine whether the flow is laminar or turbulent. Some proteins are also sensitive to shearing forces and may denaturate at sharp edges in the flow system.

Further considerations of cell volume, stirring, and ways to exchange the liquid have to be considered when *electrochemical electrodes* are used, e.g., when ellipsometry is combined with an electrochemical analysis or when the adsorption should be controlled by an applied potential.[38] Another important aspect is *sample mounting*. The cell in Fig. 12.7 is open and the sample is immersed into the liquid from above. This allows the sample and the cell to be separately aligned which is preferable. The drawback is that very long (several cm) samples are needed or alternatively a smaller sample can be mounted on a holder that is immersed in the liquid. The latter implies, however, that additional parts have to be cleaned and several materials will come into contact with the solution. In closed cell designs the sample can be mounted directly on one of the walls inside the cell. Another possibility is to arrange a hole in one of the cell walls (backside) and press the sample onto the cell from the outside. This is very attractive to work with if leakage can be avoided. The drawback is that the cell and the sample cannot be aligned separately.

The *windows* in the cell are very important as they may introduce systematic errors in the ellipsometric data due to imperfections. The procedures for correcting such effects are in principle the same as for windows used in ellipsometer studies in vacuum chambers. However, in many cases small errors are accepted as they affect the results to second order, especially if a simple evaluation like the PRIM method is used. More important is that the window imperfections must be time invariant. The cell in Fig. 12.7 is made of quartz and is glued together to minimize stress-induced birefringence. In other designs small windows are mounted in the walls of a metal or Teflon cell. The placement of the windows also defines the *angle of incidence* as the beam should be perpendicular to the windows. In ellipsometry

analysis, a variation of the angle of incidence can be a powerful way to increase the database for the analysis. Such a variation is not trivial to perform in *in situ* studies. The use of a circular cell could in principle allow angle variations regarded that the beam convergence can be corrected for. In some studies a fixed number of angles are used in cells with multiple windows. Recently a setup using light guides to facilitate the passage of light through an air/solution interface was presented.[39] With this arrangement multiple angle of incidence ellipsometry can be performed at a solid/liquid interface. Finally it may be of interest to control the *temperature* of the solution. Different approaches are found including water heating (or cooling) or resistive heating.

If measurements are done in opaque liquids an internal reflection approach can be used. This is referred to as attenuated total reflection (ATR) ellipsometry[40] or total internal reflection ellipsometry (TIRE).[41, 42] Such designs require more complicated cells with special arrangements of the optical path to couple the light to the sample surface. The sample must be semitransparent to allow the probe beam to become reflected at its second interface to achieve sensitivity for layers on the surface. Internal reflection ellipsometry is from a configuration point of view similar to surface plasmon resonance setups and similar cell designs and flow systems can be employed.[43] Cells developed for controlled flow conditions are also described in the literature.[44–46]

12.6.3 *In situ* Considerations for Biological Interfaces

In many cases layers with thickness in the order of nm are studied and it is obvious that the cleaning process and the maintenance of a clean surface is extremely important. Cleaning protocols for the different types of surfaces used can be found in references given in the applications section. The last step in the cleaning procedure is normally that the sample is blown dry in air or nitrogen gas which then must be of high quality and oil free in order not to contaminate the sample surface. Once a sample has been cleaned it must be inserted into the cell without contamination of the surface. The procedure involves transfer of the sample through an air/liquid interface. At such an interface a contamination layer easily can accumulate which then can transfer to the sample by the Langmuir-Blodgett (LB) effect. It is therefore extremely important to use highly purified solutions. By special arrangements it is possible to keep the sample wet all the time to avoid the LB contamination. Once in the cell, one must be aware that the surface eventually adsorb ions from the liquid as discussed above. Furthermore, the liquid close to the surface is stagnant even during stirring

conditions. The thickness of the unstirred layer is of the order of micrometers and macromolecules diffuse across this layer and diffusion delays must therefore be taken into account in adsorption studies.

12.6.4 Some Model Surfaces

Ellipsometry has been applied to numerous materials in the life science area. Among them are titanium and its oxides, steel, copper, silver, chromium, and many more. Many of the materials have also been modified by chemical treatments or functionalizations. However, the majority of studies have been carried out on oxidized silicon or gold samples and here we limit the discussion to these two materials.

12.6.4.1 Oxidized Silicon

Numerous ellipsometric studies on protein adsorption onto silicon wafers can be found in the literature. Most are made at a single wavelength using null ellipsometry. Many of the studies are performed *in situ* at a solid/liquid interface with the objective to follow the kinetics of adsorption or biomolecular reactions. Silicon substrates are very attractive model surfaces for several reasons. They are cheap and easy to obtain with an almost atomically flat surface. The material quality is extremely high and its optical functions are well known. Silicon has a high refractive index and therefore provides a large optical contrast to organic layers formed on its surface. This implies high sensitivity and with rather simple instrumentation a thickness resolution well below one nm is obtainable. From a chemical point of view silicon is very stable which makes it suitable for studies in a variety of ambients. However, silicon oxidizes in air and water-based ambients so what biomolecules actually see is the oxidized silicon surface. This implies additional advantages since silicon dioxide is in principle glass, a very common material used, e.g., in clinical laboratories. Studies of biointeractions on oxidized silicon will therefore have immediate relevance for many surface phenomena encountered in biochemistry and medicine. Furthermore, the surface chemistry of silicon can readily be modified through silanization.[47, 48] A conversion from an extremely hydrophilic surface to an extremely hydrophobic surface can be done by covalent binding of dichlorodimethylsilane. Other types of functionalizations, like surfaces exposing groups of −COOH, −NH3, etc., can also be obtained through silanization. Also gradients in surface energy and functionality have been developed[49] and used in studies of biointeraction using lateral scanning ellipsometry.[50]

One drawback of silicon is that it is almost dielectric in the visible wavelength region which implies that Δ of a silicon surface with a thin native oxide will be close to 180 degrees. From an instrumental point of view this is not a problem in null ellipsometry or when a rotating compensator ellipsometer is used but for a rotating analyzer ellipsometer this results in very low accuracy due to the inherent low accuracy in the latter configuration. However, more fundamental is that upon adsorption of a monolayer of biomolecules, the dielectric nature of silicon leads to changes mainly in Δ and a very small change in ψ is obtained. As an example, the change in Δ upon deposition of a 3 nm layer of a protein ($n = 1.5$) in air at an angle of incidence of 70 degrees and wavelength of 633 nm is $\delta\Delta = -8.8$, whereas the change in ψ is only $\delta\psi = 0.19$ degrees. In principle it means that we essentially have only one parameter that carries information about the layer. This is not a problem if only surface mass is of interest but when more detailed layer information is needed it is a severe drawback. If spectroscopy is available one can move to shorter wavelengths but the most common strategy is to oxidize the silicon substrate prior to the measurements. A 30 nm oxide yields $\delta\Delta$ and $\delta\psi$ of -3.9 and 0.96 degrees, respectively, for a 3 nm protein layer. The conclusion is that additional information is obtained as we see large changes in both ψ and Δ and the possibility of resolving both d and n increases. However, the dependence of the various parameters in the optical model of the sample is generally complicated and the actual separation of parameters must be simulated from sample to sample. In Fig. 12.8 the change in ψ and Δ in the wavelength region 300 to 1000 nm is shown for the example above to further illustrate the effects of choice of wavelength and oxide thickness.

12.6.4.2 Gold

One advantage of gold as a model surface for ellipsometric studies of bioadsorption is the absence of an oxide. On the other hand, when measurements are carried out *in situ* in a liquid, one has to consider electrochemical phenomena which may influence the adsorption itself but also may change the optics of the near surface region of the gold substrate, most probably due to free carrier effects induced by the adsorbed layer.[51, 52] Gold is chemically very stable in most liquids, an advantage for *in situ* studies. Almost all ellipsometric studies on gold are done on thin films evaporated on glass or silicon substrates. These polycrystalline films will have a microstructure which depends on the deposition conditions and shows variations in grain size and surface morphology. These variations have to be taken into account and each sample should be carefully characterized ellipsometrically before use in bioadsorption experiments. A major advantage

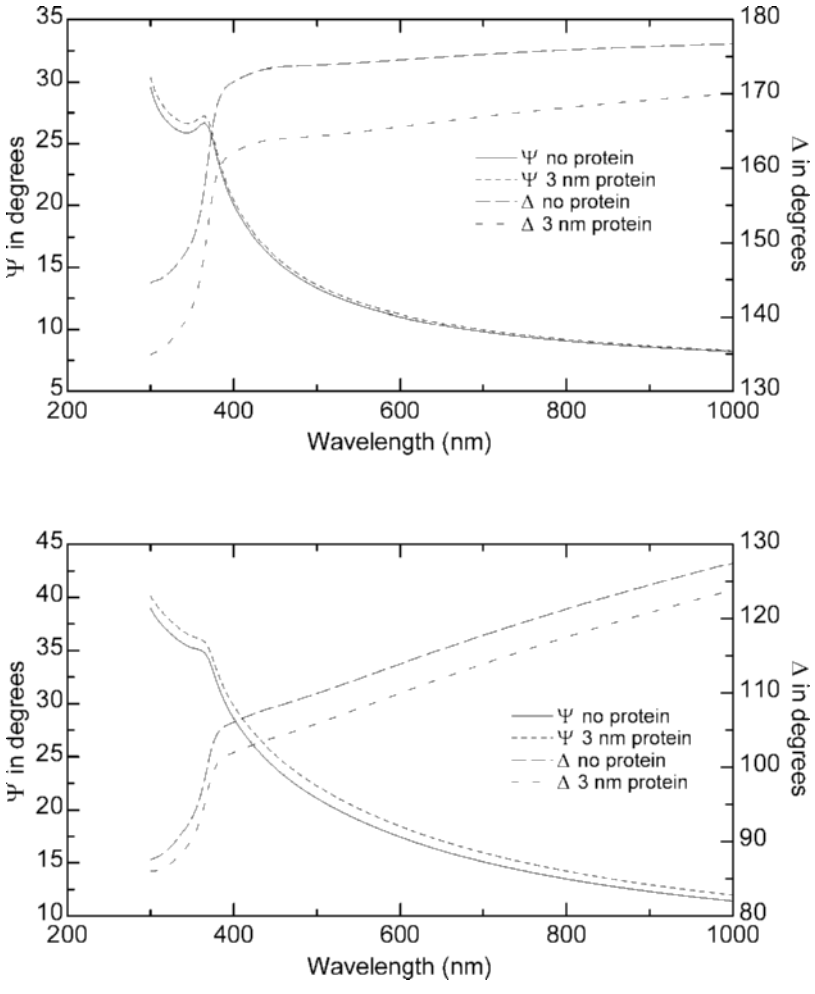


Figure 12.8 Simulation of changes in ψ and Δ due to adsorption of a 3 nm layer with $n = 1.5$ on oxidized silicon with a 1.5 nm thick native oxide (top) and with a 30 nm oxide layer (bottom).

with gold is that surface modifications can be done by use of thiol chemistry to change the surface energy and chemical functionality.[53, 54]

12.6.5 Studies on Real Biological Surfaces

There exist very few biological surfaces that have enough specular reflection to allow precise ellipsometric studies in the visible spectral region.

Biological surfaces are normally not flat over an area of a few mm^2 which is another requirement. Teeth and eventually bone could be used but to make ellipsometric measurements on these materials, they have to be polished. Furthermore, as these materials are living materials and once they are removed from a body and especially if they are polished, they will change and much of the relevance of such studies will be lost.

12.6.6 Complementary and Independent Information

It is very common to combine ellipsometry with other optical as well as non-optical techniques. One objective may be to get independent microstructural information about layer thickness and surface morphology, e.g., by using atomic force microscopy. Another objective may be to extract complementary information to sort out fine details in layer structure and dynamics. One illustrative example is a comparative study of protein adsorption on titanium oxide surfaces using ellipsometry, optical waveguide light mode spectroscopy, and quartz microbalance.[55] This study shows that ellipsometry gives the surface mass whereas the quartz microbalance provides complementary information about bound water and viscoelasticity of the protein layer. Ellipsometry can also be combined with capacitance measurements. By assuming that ellipsometry provides surface mass and capacitance measurements of the surface coverage, the surface area per molecule could be determined for small peptides.[56] These and similar approaches will not be reviewed here but it should be pointed out that it is often fruitful to combine results from different techniques.

12.6.7 Experimental Design

The data bank for analysis should be as large as possible and with experimental design one strives for optimizing the composition of the data bank and the quality of its data with respect to the objectives. Here we discuss a few useful strategies.

12.6.7.1 Variation of Ellipsometric Settings

When the sample is given one can only vary the instrumental settings, in principle the wavelength, angle of incidence, and the polarization of the incoming beam. Spectral data are in general an advantage since more information is available. However, for each wavelength the optical function of the material under study is different so spectral data are not redundant

data but will provide a more complete picture of the sample than single-wavelength data will do. During measurements in air, the angle of incidence can easily be varied whereas studies in liquids require special cell designs.[39] However, for very thin films, ellipsometry data at different angles may become linearly dependent which sometimes is used as an argument against variable angle ellipsometry. This effect is not universal and should be checked for in each case by simulations.

12.6.7.2 Variation of Sample or Media Parameters

If the film under study can be deposited with different thicknesses on different substrates or if the sample can be measured in different ambients one can use multiple sample analysis to extract optical and microstructural data with a larger confidence.[57] Tiberg, et al.[58] successfully determined the thickness and surface mass density of adsorbed surfactant films using multiple ambients. This requires that the layer material really is unaffected by such changes which probably is not the case very often in systems involving biomolecules. One important exception is when the oxide thickness of silicon is varied to tune the sensitivity of the ellipsometric measurements as described above. Once again the advice is to run simulations for the specific sample geometry and instrumental settings used.

12.7 Applications

12.7.1 Introduction

The objective of this section is to demonstrate how ellipsometry can be applied in different fields in the life science area. A majority of the reports describe its use for determination of surface mass density and/or thickness of an adsorbed layer of biomolecules on flat surfaces. Monitoring of the kinetics of adsorption *in situ* is also very common although the actual time dependence for mass deposition is made use of only in a few cases. This is most probably because the dynamics of macromolecular interactions with surfaces are complicated to describe as it involves mass transport, diffusion, sticking efficiency, and other phenomena.[59–61] Here we will discuss selected examples of adsorption to flat surfaces. Several reviews can be found that deal with these types of applications.[62–64] As a complement we will broaden the description by discussing a variety of applications including interactions with porous materials, spectroscopy, imaging, sensors, and engineering applications.

12.7.2 Adsorption of Biomolecules to Model Surfaces

One of the most common applications of ellipsometry in the life science area is to measure the deposition of mass due to macromolecular adsorption to surfaces. The measurements are often performed *in situ* to record the adsorption kinetics. Most studies involve proteins but other biomolecules such as triglycerides, DNA fragments, drugs, etc. have been studied. Single-wavelength null ellipsometry is used in most studies due to its simple operation and stability. For thin layers it is not always possible to resolve both n and d with reasonable accuracy and therefore the results often are converted to Γ using one of the methods described above. Here we present a few examples illustrating various types of experiments on different types of surfaces.

12.7.2.1 Adsorption of One Type of Protein to Simple Model Surfaces

Protein adsorption from single protein solutions to simple model surfaces has been performed on a larger number of different surfaces including metals and their oxides.[64] The adsorption of human serum albumin to silicon substrates in Fig. 12.1 may serve as an illustration of this type of application performed *ex situ*. Albumin has been widely used as a model protein and studies are found on a variety of materials including TiO_2,[65] plasma polymer films,[66] polysaccharide films,[67] HgCdTe,[14] tantalum oxide,[68] gold and stainless steel,[69] platinum and zirconium,[70] and many more. In Fig. 12.9, one example on an *in situ* study is shown.[71] In this particular experiment the adsorption of IgG was followed on hydrophilic and hydrophobic oxidized silicon surfaces in phosphate buffer at pH 7.4. The adsorbed amount was calculated from d and n using de Feijter's model.[18]

Figure 12.2 shows a second example of kinetic studies of protein adsorption on a model surface - in this case ferritin on gold.[13] Initial adsorption processes are other examples of important aspects of protein adsorption. The question is what happens during the first few seconds after a surface has come into contact with a protein solution. To avoid mass transport limitations a special flow cuvette with 50 µl volume was developed to enable a fast exchange of solution close to a substrate surface, and off-null ellipsometry was used to allow measurements with a time resolution of 0.1 s.[36] In Fig. 12.10, the initial adsorption of ferritin is shown where a delay in the initial kinetics is clearly seen. This delay

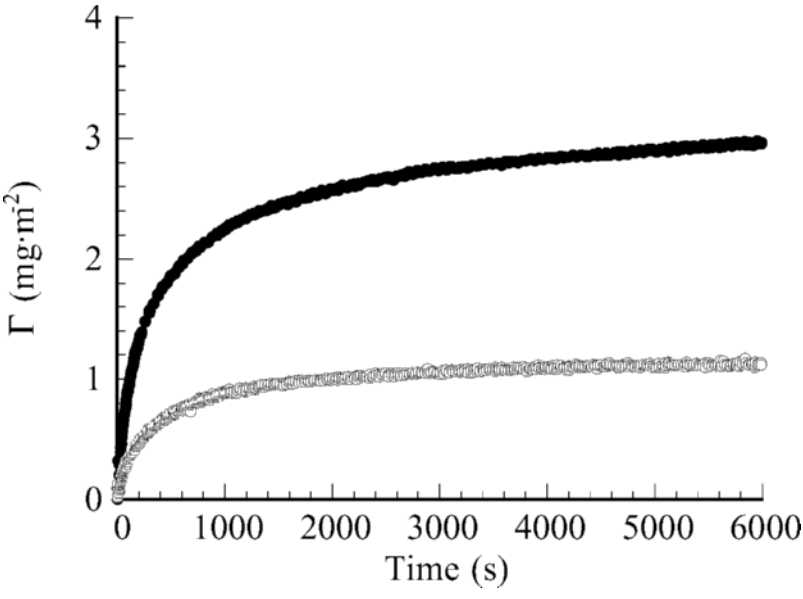


Figure 12.9 Adsorbed amount of IgG versus time on hydrophobic (top curve) and on hydrophilic (bottom curve) silicon dioxide. Reprinted with permission from Ref. 71. Copyright (1995) Elsevier Science.

cannot be explained by diffusion over an unstirred layer. Assuming that the initial lag-phase corresponds to a nucleation process, a kinetic model including cluster formation was developed.[72]

12.7.2.2 Adsorption of Proteins to Functionalized Surfaces

Through the attachment of chemical functional groups by silanization or by thiol chemistry, more detailed studies of the interaction between complex protein solutions and model surfaces can be performed. Figure 12.11 shows the ellipsometric response upon exposure of human plasma to a gold surface that is modified with a self-assembled monolayer of $HS\text{-}(CH_2)_{15}\text{-}CH_3$.[62] Upon subsequent exposure to various polyclonal antibodies, protein adsorption profiles can be mapped. Also the coagulation system has been studied and upon exposure of blood plasma to surfaces that contact activate this system, events during this process can be followed by ellipsometry.[73] Another example is ellipsometric studies of serum protein adsorption onto phospholipid surfaces, addressing issues related to intravenous drug delivery.[74] Further examples and references to

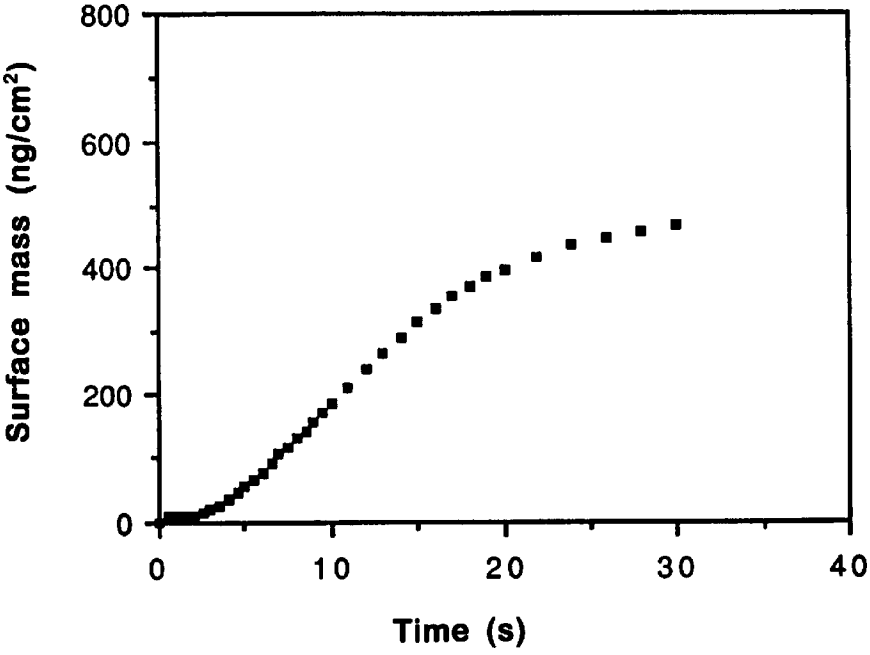


Figure 12.10 Surface mass Γ versus time for ferritin adsorbing on hydrophobic silicon dioxide measured with off-null ellipsometry.[72]

protein adsorption on model organic surfaces using ellipsometry is given by Tengvall, et al.[62]

Let us finally mention one example of a study of another complex biological system, the complement system. Figure 12.12 shows the deposition of complement proteins from human serum onto silicon surfaces "functionalized" with IgG. Depending on the additives to the serum, temporal deposition in different activation pathways as well as quenching of the complement system could be monitored. Thus, by using ellipsometry, not only can the activation itself be ascertained but also the activation pathway can be probed by time course experiments. In addition polyclonal antibodies can be used to identify surface active proteins, or convertases (active enzyme complexes).[75]

12.7.2.3 Adsorption of Proteins into Porous Surfaces

Besides the physical and chemical properties of an interface, its microstructure influences the interactions with biomolecules. Textured

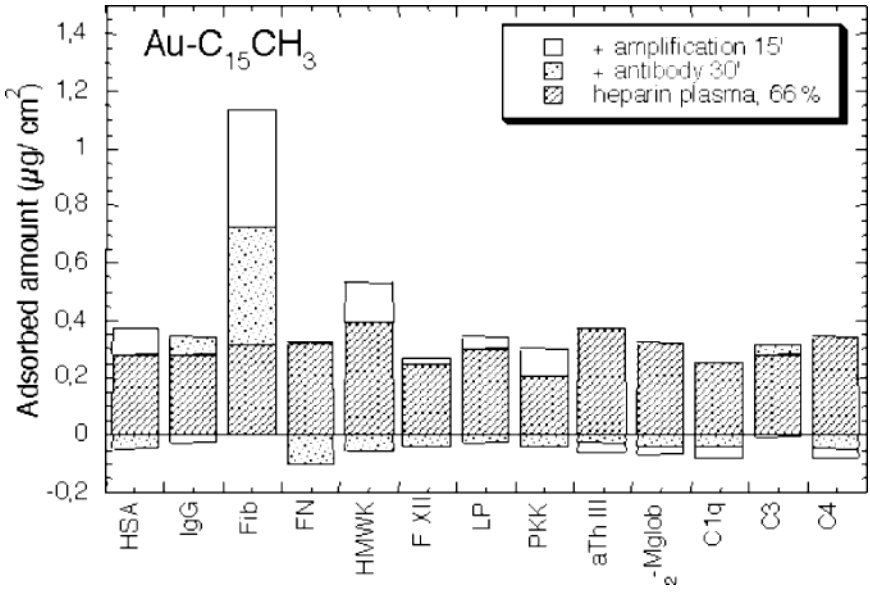


Figure 12.11 Adsorption of proteins from heparinized human plasma followed by antibody binding on hydrophobic HS–$(CH_2)_{15}$–CH_3 surfaces on gold. The protein layer bound mainly anti-fibrinogen. Reprinted with permission from Ref. 62. Copyright (1998) Elsevier Science.

surfaces are, for example, developed based on the hypothesis that surface features with dimensions of the same order as cells or macromolecules may improve their "biocompatibility".[76] Also porous surfaces are of interest in this context. A porous interface may provide a smooth transition between the organic and the inorganic phases. Possibilities for mechanical anchorage and compliance are other potential advantages in the case of bone implants, and controlled drug release from porous layers can also be envisaged.

We have studied protein adsorption into porous silicon dioxide layers[77] as well as into porous silicon layers.[78] Ellipsometry was then used to quantify the amount of adsorbed protein and also to follow the kinetics of adsorption from a liquid. Furthermore, depth profiling is possible by employing spectroscopic ellipsometry. Figure 12.13 shows the kinetics recorded during adsorption of the human serum albumin into a porous silicon layer. The rate of adsorption in such layers is generally lower than that on a flat surface (c f Fig. 12.2), which most probably is due to a slow diffusion of large molecules into the porous structure.

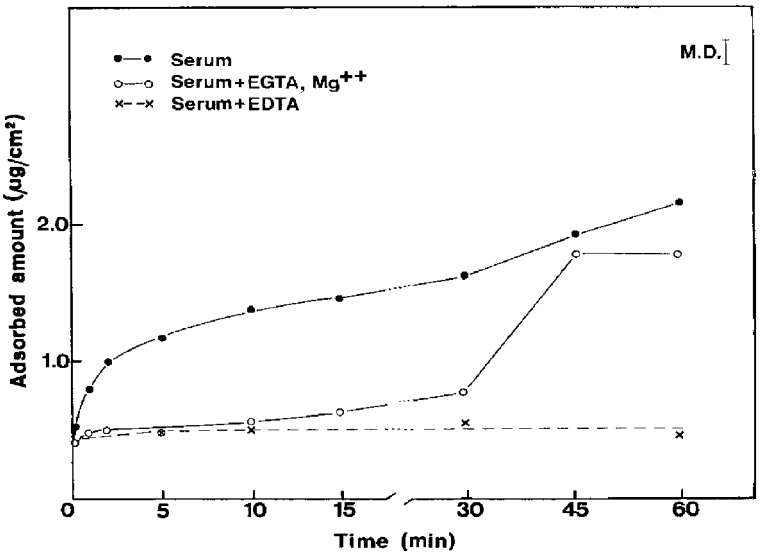


Figure 12.12 Deposition of human serum onto IgG coated hydrophobic silicon during up to 60 minutes incubations. The black dots and solid line show deposition from normal serum and indicates a (rapid) classical complement activation and increase in serum deposition. The solid line with open rings shows serum deposition from normal serum with added EGTA, and indicates the typical deposition profile during an alternative complement activation. Observe the 10 minutes delay in serum deposition. The dashed line with crosses show serum deposition with EDTA added to serum. The addition of EDTA quenches complement activation completely.[75]

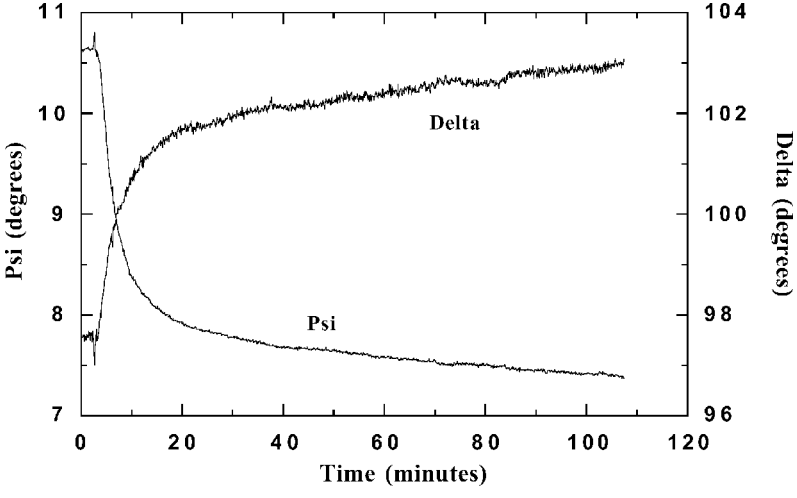


Figure 12.13 ψ and Δ recorded during adsorption of human serum albumin into a 288 nm thick porous silicon layer. Albumin was added to the ellipsometer cell after 3 minutes to a concentration of 1 mg/ml.[76]

In Fig. 12.14a ellipsometric spectra measured before and after protein adsorption into the porous silicon layer are shown. The difference between them gives information about the adsorbed amount and also about the adsorption depth. Quantification was achieved using multilayer models and EMA modeling as described earlier in this chapter and a microstructural parameterization as shown in Fig. 12.14b can be obtained. These results indicate that albumin penetrates into the porous matrix. Similar studies using fibrinogen showed adsorption only on top of the porous layers. Several factors influence this but of major importance is the larger size of fibrinogen molecules (MW = 340 kDa, typical dimension 46 × 7 nm) compared to albumin molecules (MW = 66 kDa, typical dimension 8 × 3 nm). Other factors are pore size, surface charge, and energy of the internal walls of the pores and the flexibility of the protein molecules.

12.7.2.4 Ellipsometric Studies Involving Bimolecular Interactions

In life sciences, bimolecular reactions are very common and may involve biological specificity and recognition. Two examples are antigen-antibody binding and enzymatic catalysis. If one of the two components is adsorbed or bound to a surface, the interaction with the counterpart can be studied with ellipsometry. Some examples of early work on immunological surface interactions were carried out by Rothen and Mathot,[2] Azzam et al.,[79] and Cuypers et al.[12] and during the 1980s and 1990s numerous reports can be found in this area. A commonly used technique to investigate which biomolecules are present on a surface is to expose a surface to a serum or a polyclonal antibody solution, preferably an IgG fraction. The antibodies will subsequently bind to the surface, given that the corresponding antigen is already adsorbed. This methodology was used by Lestelius, et al.[80] to map the adsorption of plasma proteins on gold that was functionalized with self-assembled alkanethiolate monolayers terminated by groups of $-CH_3$, $-OCOCF_3$, $-SO_3^-$, $-COO^-$ and $-OH$. A biological system involving complex protein surface dynamics is surface activated coagulation. Early work in this area was done by Vroman.[3] Tengvall, et al.[73] used scanning ellipsometry on surface energy gradients on silicon to investigate the surface dynamics of several proteins in the coagulation cascade.

12.7.2.5 Internal Reflection Ellipsometry in Life Science

Several biological fluids are opaque and will not allow ellipsometric studies at a solid/liquid interface in external reflection mode. However, if

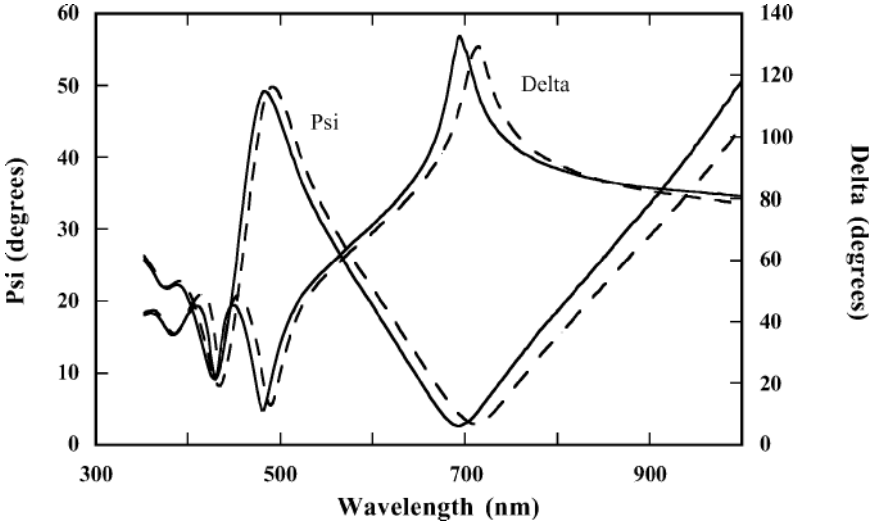


Figure 12.14a Ellipsometric spectra measured before (solid curve) and after (dashed curve) adsorption of HSA in a porous silicon layer.

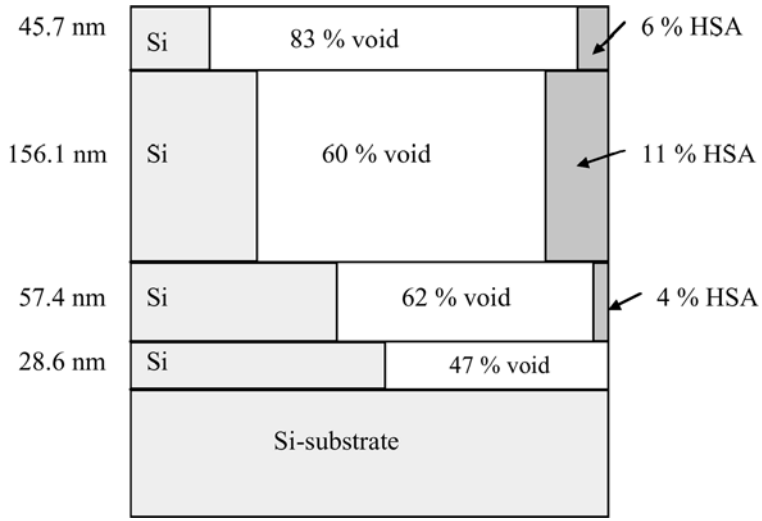


Figure 12.14b A microstructural model of a porous silicon layer that has been exposed to HSA showing the porosity gradient and the fill factors of HSA in the individual sublayers.

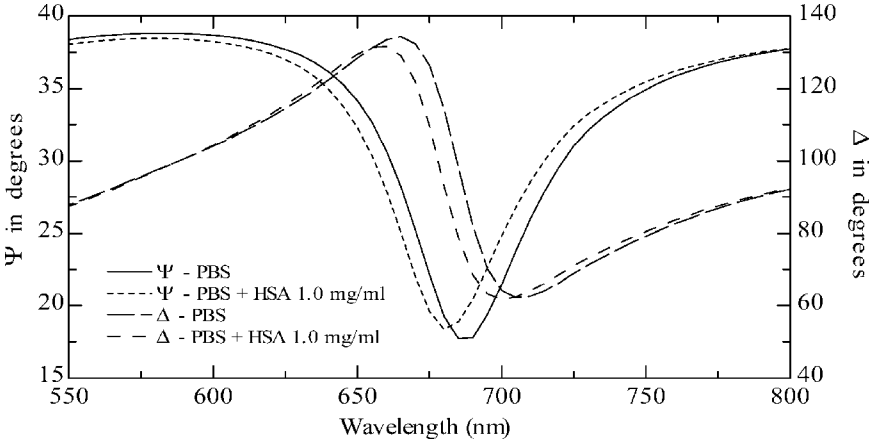


Figure 12.15 TIRE spectra measured before and after adsorption of HSA on a 50 nm gold layer.[41]

the substrate is transparent the interface is accessible by internal reflection. In the infrared spectral region, internal reflection ellipsometry in combination with an attenuated total reflection cell can be used to determine optical spectra of lubricants and water.[40] Rekveld developed this technique at single wavelength to study protein adsorption under a controlled flow.[42] Recently internal reflection setups in combination with spectroscopic ellipsometry have been used to study protein adsorption, which in total internal reflection ellipsometric mode was coined TIRE.[41] Figure 12.15 shows how ψ and Δ changes due to an adsorbed monolayer of HSA on a 50 nm gold film. Two important observations can be made. First we see that ψ exhibits a dip similar to that seen in surface plasmon resonance (SPR) experiments. The phenomenon is in fact very similar as ψ is the ratio of the p- and s-reflection coefficients. The second observation is that the change in Δ close to the dip in ψ is very large – about 20 degrees. This is a remarkable change when compared for example to the change in external reflection which at maximum is a few degrees. The explanation is that we obtain an enhancement due to the SPR effect. Similar results are found on thin copper films as well.[81] The SPR enhancement has also been observed by Westphal, et al.[82]

However, TIRE can also be used to monitor thin films not showing SPR effects, like chromium and iron, but the changes in ψ and Δ are then of the same order as in external reflection mode. A potential application of TIRE is to monitor deposition (biofouling) of biomolecules at the inside of a tube in a flowing system. The liquid can then be opaque and

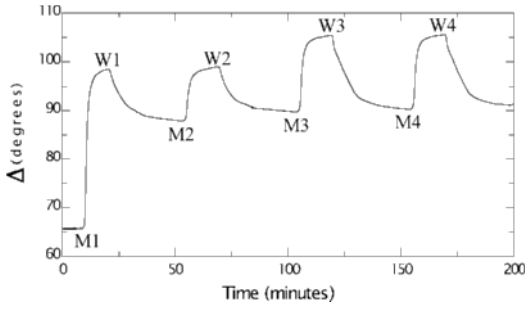


Figure 12.16 Δ versus time for a 50 nm gold film in a TIRE experiment under exposure to milk at times as indicated by M1, M2, etc. In between the milk exposures the surface was exposed to water starting at times W1, W2 etc.[41]

the probe beam monitors the deposition with monolayer sensitivity by using the internal reflection mode. A TIRE example of adsorption of biomolecules from milk onto a gold film and subsequent rinsing with water is shown in Fig. 12.16.[41] Again we observe a very large change in Δ. The adsorption was found to be irreversible when rinsing with water was done. By use of other solutions, such as sodium hydroxide, the surface can be cleaned and we can thus monitor both deposition and cleaning *in situ* in a flow system.[83]

12.7.2.6 Other Areas: Protein Electrochemistry, Enzymes, Surface Cleaning, and Biofouling

There are many other application areas of ellipsometry in the life sciences, and a few are mentioned here to illustrate the different possibilities. In electrochemical experiments the adsorption of biomolecules may lead to unexpected phenomena like changed interfacial capacitance or a general decrease in electrochemical activity. Ellipsometry can here serve as a diagnostic tool to detect and monitor the kinetics of growth of biofilms as well as of other layers.[84] However, a more direct use of ellipsometry is evident in redox-protein electrochemical studies, e.g., cytochrome c adsorption on gold.[84]

Enzymatic degradation of surface layers can also be monitored by ellipsometry. Berg, et al.[85] have studied the effects of the proteolytic enzymes krillase and trypsin on gelatin layers.

Ellipsometry has also been used to study effects of detergents on adsorbed layers. Engström, et al.[86] found that the removal of triglycerides

from polyvinylchloride and chromium surfaces could be monitored. More recently Lee, et al.[87] studied the interactions between lipase and the surfactants sodium dodecyl sulfate and tetradecyltrimethylammoniumbromide. In oral sciences, the accumulation and removal of dental plaque have been addressed. Proteolytic degradation of oral biofilms obtained by adsorption from saliva on oxidized silicon was investigated[88] as well as adsorption of salivary proteins onto hydroxyapatite.[89] Influence of saliva concentration on the rate of adsorption and total surface concentration on hydrophilic and hydrophobic silicon dioxide surfaces were studied by Lindh, et al.[90] who also performed similar measurements on selected proteins purified from human saliva.[91]

Biofouling is a large problem in process industry and ellipsometry can in this area be used to study cleanability of surfaces after adsorption of biomolecules. Studies on cleaning of stainless steel with adsorbed beta-lactoglobulin is one example.[92]

Protective coatings are of large interest in food and pharmaceutical processing as such coatings may exhibit antimicrobial activity. Lakamraju, et al.[93] have studied adsorption of the antibacterial peptide nisin and how well it can resist removal by selected milk proteins. Other work related to milk is the study of the structure of layers of β-casein.[94] Due to the surface affinity of casein it has been employed as an emulsifier and a foaming agent. The investigation of surface dynamics and structure of casein will therefore bear relevance for applications in food processing but also for performance of paints and glues and more areas.

One of the important factors for the bread making quality of wheat is considered to be the surface activity of proteins like gliadins and glutenins. Gliadin is known to be very surface active with possible consequences for the gas holding capacity of a dough. In view of this Örnebro et al.[95] have studied the behavior of gliadin at interfaces using ellipsometry.

12.7.3 Spectroscopy

More than 15 years ago we showed that it was possible to perform spectroscopy on monolayers of bovine serum albumin (BSA), γ-globulin, and hemoglobin on solid substrates.[14] These measurements were performed in air in the photon energy region 1.5 to 5 eV using substrates of evaporated films of gold and platinum as well as on semiconductor surfaces. Figure 12.17 shows the obtained dielectric function of a thin film of γ-globulin on gold (see also Fig. 12.4). The ellipsometric data allowed unambiguous determination of both n and d and even the weak polypeptide backbone UV absorption above 5 eV as well as the absorption from aromatic amino acid residues at 4.4 eV (280 nm) could be resolved. However,

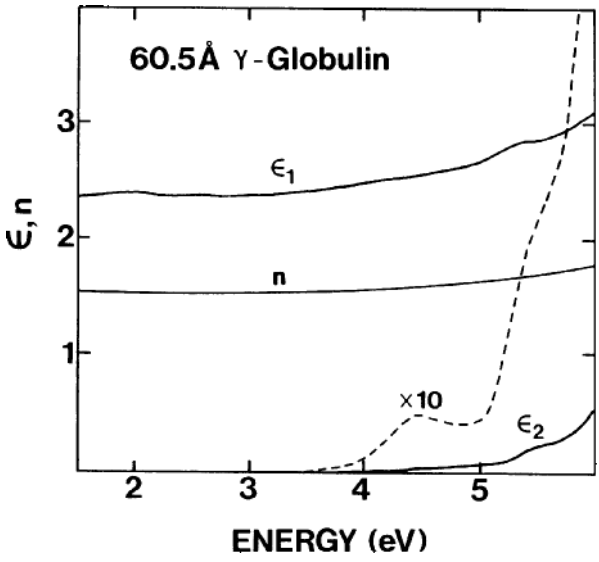


Figure 12.17 The complex dielectric function ($\varepsilon = \varepsilon_1 + i\varepsilon_2$) and real part of refractive index of a 60.5 Å thick layer of γ-globulin adsorbed on a gold substrate. ε_2 is also shown magnified by a factor of ten.[14]

it should be pointed out that these measurements were done in air and it is hard to justify that the obtained refractive indices represent the true intrinsic values of the protein molecules. Care was taken to prepare as dense layers as possible but the hydration state is unknown. Nevertheless, further analysis was done on BSA layers. By assuming that the best optical spectrum is that with the largest value of n (Fig. 12.4), it was possible to model the density of a less compact layer of BSA on a platinum substrate using the Bruggeman EMA as described earlier.

Later optical spectra for protein monolayers were determined at the solid/liquid interface.[15] In studies on gold it was found that a simple three-phase model could not be used to obtain satisfactory results for adsorption of ferritin.[13, 51] A four-phase model was developed to include protein-substrate interactions, which eliminated model artifacts in the evaluated optical spectra. In principle a very thin (less than a nm) depletion layer modeled with the Bruggeman EMA was included. The ellipsometric sensitivity to electrochemical changes in the near-surface region in gold was pointed out by Chao and Costa[96] and was proposed to be the physical background.[51] Although speculative, the four-phase approach gave consistent results and completely eliminated artifacts in the protein optical spectra. Figure 12.18 shows the complex-valued dielectric function of a monolayer of ferritin on gold. It was also possible to follow

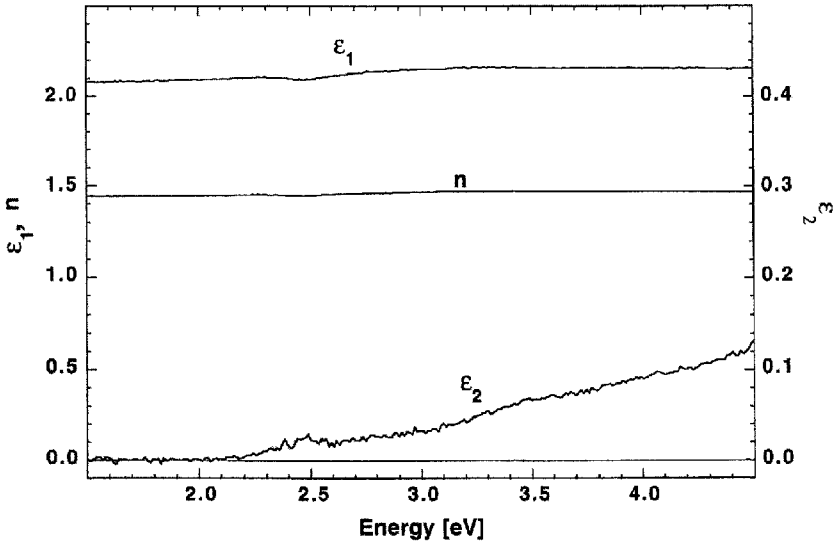


Figure 12.18 The dielectric function ε and refractive index n measured *in situ* on a 9.5 nm layer of ferritin on gold.[51]

the development of the protein-surface interaction and a two-state protein adsorption model was proposed.[51] The methodology has later been refined by Reipa, et al.[52] who determined the complex-valued refractive index of 2.9 nm layers of putidaredoxin as shown in Fig. 12.19.

Spectroscopic ellipsometry has also been employed to study adsorption of HSA and fibrinogen on fluorohydrocarbon polymers,[97] but in these studies the wavelength dispersion was mainly used to increase the database for evaluation of Γ. The protein layers were modeled with the Cauchy formula for the evaluation of n and d. The Maxwell-Garnett EMA was suggested to model the microstructure (density effects) of the layer to obtain more realistic results. In another investigation the globular protein β-lactoglobulin was studied on hydrophilic and hydrophobic silicon with the objective of studying displacement upon rinsing with buffer.[98] Also in this case the spectral information was used to increase the database for evaluation. Similar types of measurements are also found on stainless steel incubated in growing cell cultures.[99]

12.7.4 Imaging

In most ellipsometric studies, mm-sized light beams are used and beam-profile weighted average properties over the area projected on a surface are thus obtained. Microellipsometry is possible down to a lateral

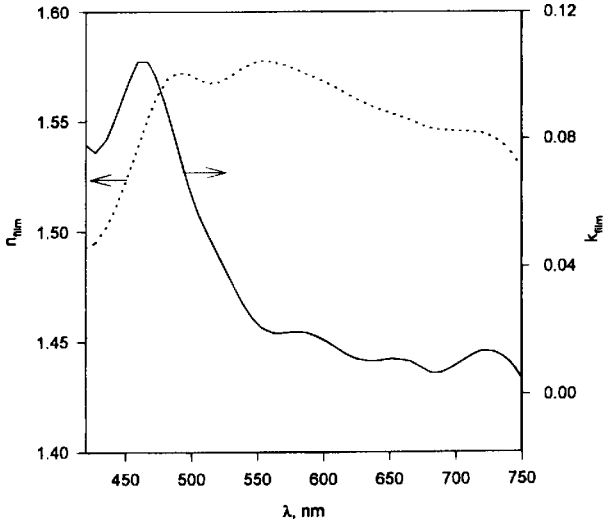


Figure 12.19 Complex-valued refractive index $N = n + ik$ for a 2.9 nm layer of putidaredoxin on gold obtained by using a fit to a model including a transition layer at the gold surface. Reprinted with permission from Ref. 52. Copyright (1997) American Chemical Society.

resolution of 50 μm but requires scanning methods for mapping of a larger area. An alternative is ellipsometric imaging whereby a large diameter beam is employed. Using a CCD-camera and a beam diameter of the order of 20 mm, it is possible to image areas up to 15 × 25 mm.[37] The lateral resolution depends on the number of pixels and the optics used but lateral resolutions down to 5 μm with nm thickness resolution was achieved. Figure 12.20 shows an image of a silicon surface on which three different types of proteins have been adsorbed in selected areas. This image is recorded with null ellipsometry operated in the so-called off-null mode.[36] The intensity from each pixel is then a measure on the layer thickness in the corresponding area on the surface. Observe that the image shows monolayers of the three proteins. The advantage with operation in off-null mode is speed but some precision is lost.

In commercially available instruments a spatial resolution of 1 μm is stated and such instruments work in the ordinary ellipsometric mode. Recently imaging ellipsometry measurements on affinity biochips with 900 targets were also demonstrated[100] and the use of imaging ellipsometry for quality control of biochips with protein and DNA spots has been suggested.[101]

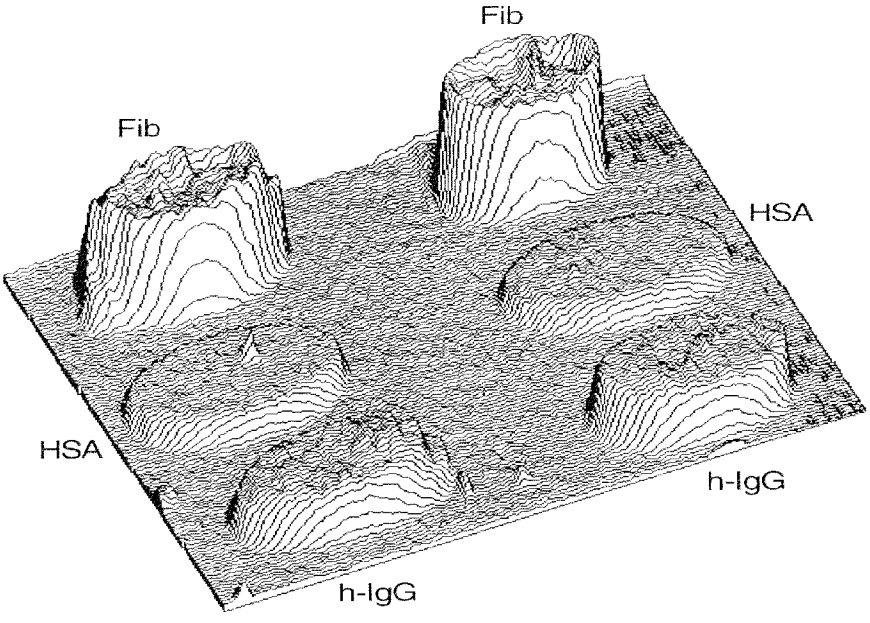


Figure 12.20 Ellipsometric image of a 15 × 25 mm hydrophilic silicon sample with monolayers of fibrinogen (Fib), human serum albumin (HSA) and human immunoglobulin G (h-IgG) adsorbed in $\phi = 4$ mm spots in duplicate.[37]

12.7.5 Biological Surfaces

Very few studies involving real biological surfaces or tissue are found in the literature. One reason may be that real biological surfaces are very inhomogeneous and will scatter and depolarize light, which makes ellipsometry studies hard to evaluate. A review of ellipsometric studies on cell coat materials has been given by Poste and Moss.[4] Some studies of uptake of salivary substances onto tooth enamel have also been carried out.[102] Recently Malmsten, et al.[103, 104] showed that adsorption of lipoproteins onto proteoheparan sulfate surfaces as well as onto endothelial cell layers grown on silicon can be followed by ellipsometry. They found a Ca^{2+} dependence on the lipoprotein binding.

Ellipsometry has also been applied to biological reflectors like the colored wing scales of Chrysiridia croesus and other moths. Using variable angle null ellipsometry at three wavelengths, the refractive index and thickness of different colored scales could be evaluated in a thin-film stack model.[105, 106]

In several disciplines one uses methods that are very similar or identical to ellipsometry but named differently. One such area in which real biological surfaces are studied is ophthalmology. Mueller-matrix imaging polarimeters have for example been used to study corneas[107] and laser polarimetry to measure the retinal nerve fiber thickness for diagnostics of glaucoma.[108]

12.7.6 Biosensors Based on Ellipsometric Readout

Ellipsometry as a sensor principle has recently been reviewed.[109] However, the idea of using ellipsometry in sensor applications is rather old, but ellipsometers are normally technically complicated and data evaluation is not trivial for non-experts. In most cases the development is limited to the sensing layers, only mentioning their potential use in clinical or industrial environments. Dedicated, robust, user-friendly and cheap instruments have not been commercially available even though several designs have been proposed. An early development was the Isoscope or comparison ellipsometer[21] designed for biosensor applications. Biosensor applications of an ellipsometric imaging system were demonstrated a few years ago.[110] Imaging systems have the advantage of allowing multisensing and high throughput systems can be developed. However, no complete ellipsometric biosensor system has, to the author's knowledge, so far been used widely on a routine basis although several imaging ellipsometer systems are available.

Another example of the use of ellipsometry in sensor systems in medical applications are monitoring of anesthetic gases by means of measuring gas induced optical changes in thin layers. Figure 12.21 illustrates this by showing ellipsometric responses from a thin hexamethyldisiloxane (HMDSO) layer on silicon for different concentrations of the anesthetic gas halothane.[111] However, gas induced changes are in most cases due to general physical effects like adsorption, layer swelling, pore filling, etc., and the selectivity is therefore in general relatively poor. In a development of a measurement system for clinical applications it is necessary to use a multisensor approach and a considerable improvement in selectivity can be obtained by combining the ellipsometric responses from several different layers and by employing pattern recognition algorithms like principal component analysis. Similar ellipsometric systems, so called optical noses, have been proposed for detection and discrimination of solvents.[112]

Current developments of biosensor systems for the physician's office include instrument designs with a fixed polarizer which is anticipated to soon enter clinical trials.[113] Among the examples recently found in the

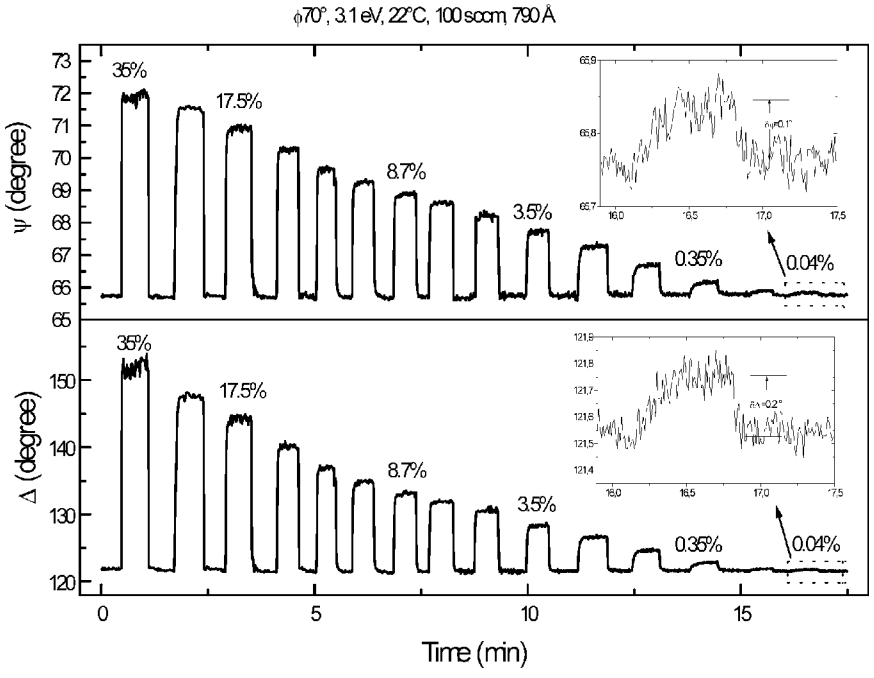


Figure 12.21 Ellipsometric response of a 79 nm thick HMDSO film for stepwise exposure to different concentrations of halothane measured at a photon energy of 3.1 eV.

literature is also an affinity biosensor system suggested for measurement of DNA hybridization and for immunoassays.[114] In immunoassays, sensitivities of 0.1 ng/ml of hepatitis B surface antigen and 0.01 ng/ml α-fetoprotein have been reported.[115] For detection of nucleic acid targets, detection limits of 5 pmol/l corresponding to 150 amol/sample are stated.[116]

12.7.7 Engineering Applications

Several authors have investigated the performance of ellipsometry for protein films of increased thickness that exceed monolayer coverage. The objectives are in general rather fundamental but it is easily understood that such multilayer systems may be important building blocks in the emerging bionanotechnology area. Ellipsometry is in this context an important tool as it can provide *in situ* feedback of layer growth.

Spaeth, et al.[117] have used spectroscopic ellipsometry to study protein multilayer systems on silicon substrates with up to 20 alternating layers of

biotinylated bovine serum albumin and streptavidin. An almost linear increase with number of incubation steps was observed. The refractive index of the multilayer was modeled with the Cauchy dispersion formula and found to be around 1.385 (at 590 nm) which is much lower than expected. For dry and compact protein films values above 1.5 have been reported[14] as a comparison.

Benesch, et al.[118] have used antigen-antibody recognition to prepare multilayer systems of 22 double layers of human serum albumin (HSA) and polyclonal antiHSA on silicon substrates. A total thickness of 103 nm was found with $N = 1.54 + i0.0009$ (at 633 nm) using a Cauchy dispersion for the protein film. AFM measurements gave a total thickness of 107 nm. They also found that the Cuypers one-component formula[19] failed to determine the surface mass and showed a difference of almost a factor of two compared to RIA measurements.

12.8 Outlook

If we go beyond simple determination of surface mass density we may appreciate that the potential of ellipsometry in life sciences is very large. Among the fascinating and advanced applications are surface dynamics in terms of the evolution of the microstructure of thin biological films, biosensor systems utilizing sensor arrays and imaging ellipsometry, imaging of molecular superstructures with predetermined biological functions, and questions related to interfacing biology with electronics. Let us make an outlook by first discussing which emerging ellipsometric instrumentation and methodology can be applied in life sciences. Finally we discuss technology transfer from the scientific community to colleagues at industrial and clinical laboratories.

Among the new possibilities are spatial imaging techniques with potential applications in high-throughput screening applications but also in surface mapping in general. A representative example is the investigation of light-activated affinity micropatterning of proteins using imaging ellipsometry.[119] Such spatially resolved immobilization of proteins may find applications in surface control of biomaterials and tissue engineering, multi-analyte biosensors, clinical assays, and genomic arrays. The use of microspot spectroscopic ellipsometry on real biological surfaces also remains to be explored. If confocal ellipsometry is developed it could open a new field for nondestructive characterization of the complex samples often encountered in life sciences. Time-resolved spectroscopic ellipsometry and time-resolved imaging ellipsometry have so far only a few applications. Internal reflection applications hold great promise due to the recently shown extreme sensitivity if the SPR phenomenon is utilized.

Infrared spectroscopic ellipsometry (IRSE) is a technique[120] with much promise in the life science area. IRSE will allow describing composition, structure, thickness and morphology in the same measurements. It has been shown that monolayers of self-assembled hexadecanethiol on gold can be resolved and analyzed by IRSE[121] as well as Langmuir-Blodgett films of arachidic acid.[122] It was also recently demonstrated that optical signatures of proteins adsorbed in porous silicon could be detected.[123]

The transfer of ellipsometric methodology to other environments proceeds slowly and so far ellipsometry is mainly used in research laboratories, especially in applications in the life science area. A few sensor-oriented systems are commercially available but presumably not widely used, and it remains to be seen if there is a market for ellipsometry in sensor applications. The competition is tough and many techniques are available including optical systems based on SPR and fluorescence. However, bionanotechnology is an emerging area where it is more likely that ellipsometry may play an important role. A nm thickness resolution for measurements at solid/liquid interfaces is relatively unique and imaging ellipsometry can provide spatial resolution. Infrared ellipsometry can add the chemical identification dimension. In applications described by words such as biomimetics, precision immobilization, self-assembly, nano fabrication, smart surfaces, and more, it appears that ellipsometry can contribute with methodology for *in situ* surface and thin film analysis and may even be central for such developments.

Acknowledgements

The author acknowledges the Knut and Alice Wallenbergs foundation for generous support to state of the art equipment over the years, and the Swedish Research Council for Engineering Sciences, the Swedish Research Council, and the Foundation for Strategic Research for project support. I also would like to thank my colleagues and students for fruitful collaboration and for many stimulating discussions regarding the use and development of ellipsometry in life science applications. A special thank to Pentti Tengvall for providing material and for valuable comments on the manuscript.

12.9 References

1. L. Tronstad, "The validity of Drude's optical method of investigating transparent films on metals," *Trans. Far. Soc.*, **31**, 1151–1156 (1935).
2. A. Rothen and C. Mathot, "Immunological reactions carried out at a liquid-solid interface," *Helvetica Chimica Acta*, **54**, 1208–1217 (1971).

3. L. Vroman, "Surface activity in blood clotting," in *Blood clotting enzymology*, (W.H. Seegers, Ed.) Academic Press, New York, 1967, pp. 279–323.
4. G. Poste and C. Moss, "The study of surface reactions in biological systems by ellipsometry," *Prog. Surf. Sci.*, **2**, 139–232 (1972).
5. M.K. Debe, "Optical probes of organic thin films: Photons-in, photons-out," *Prog. Surf. Sci.*, **24**, 1–282 (1987).
6. H. Arwin, "Spectroscopic ellipsometry and biology: recent developments and challenges," *Thin Solid Films*, **313–314**, 764–774 (1998).
7. H. Arwin and D.E. Aspnes, "Determination of optical properties of thin organic films by spectroellipsometry," *Thin Solid Films*, **138**, 195–207 (1986).
8. H. Arwin, "Ellipsometry on thin organic layers of biological interest: characterization and applications," *Thin Solid Films*, **377–378**, 48–56 (2000).
9. H. Arwin, "Ellipsometry," in *Physical Chemistry of Biological Interfaces* (A Baszkin and W Norde, Eds.) Marcel Dekker (2000) pp. 577–607.
10. U. Jönsson, M. Malmqvist, and I. Rönnberg, "Adsorption of Immunoglobulin G, protein A, and fibronectin in the submonolayer region evaluated by a combined study of ellipsometry and radiotracer techniques," *J. Coll. Int. Sci.*, **103**, 360–372 (1985).
11. J. Benesch, A. Askendahl, and P. Tengvall, "Quantification of adsorbed human serum albumin at solid interfaces: a comparision between radioimmunoassay (RIA) and simple null ellipsometry," *Colloids and Surfaces B*, **18**, 71–81 (2000).
12. P.A. Cuypers, W.Th. Hermens, and H.C. Hemker, "Ellipsometry as a tool to study protein films at liquid-solid interfaces," *Anal. Biochem.*, **84**, 56–57 (1978).
13. J. Mårtensson, H. Arwin, H. Nygren, and I. Lundström, "Adsorption and optical properties of ferritin layers on gold studied with spectroscopic ellipsometry," *J. Coll. Int. Sci.*, **174**, 79–85 (1995).
14. H. Arwin, "Optical properties of thin layers of bovine serum albumin, γ-globulin, and Hemoglobin," *Appl. Spectr.*, **40**, 313–318 (1986).
15. J. Mårtensson, H. Arwin, I. Lundström, and Th. Ericson, "Adsorption of lactoperoxidase on hydrophilic and hydrophobic silicon dioxide surfaces: an ellipsometric study," *J. Coll. Int. Sci.*, **155**, 30–36 (1993).
16. D. Ducharme, A. Tessier, and S.C. Russev, "Simultaneous thickness and refractive index determination of monolayers deposited on an aqueous subphase by null ellipsometry," *Langmuir*, **17**, 7529–7534 (2001).
17. D.A.G. von Bruggeman, "Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen," *Annal. der Physik*, **5**, 636–679 (1935).
18. J.A. de Feijter, J. Benjamins, and F.A. Veer, "Ellipsometry as a tool to study the adsorption behavior of synthetic and biopolymers at the air-water interface," *Biopolymers*, **17**, 1759–1772 (1978).
19. P.A. Cuypers, J.W. Corsel, M.P. Janssen, J.M.M. Kop, W.Th. Hermens, and H.C. Hemker, "The adsorption of prothrombin to phosphatidylserine multilayers quantitated by ellipsometry," *J. Biol. Chem.*, **258**, 2426–2431 (1983).
20. M. Stenberg and H. Nygren, "The use of the isoscope ellipsometer in the study of adsorbed proteins and biospecific binding reactions," *J. de Phys, Colloque*, **C10**, 83–86 (1983).

21. M. Stenberg, T. Sandström, and L. Stilbert, "A new ellipsometric method for measurements on surfaces and surface layers," *Mat. Sci. and Eng.*, **42**, 65–69 (1980).
22. P.G. Snyder, Y-M. Xiong, J.A. Woollam, G.A. Al-Jumaily, and F.J. Gagliardi, "Graded refractive index silicon oxynitride thin film characterized by spectroscopic ellipsometry," *J. Vac. Sci. Technol. A*, **10**, 1462–1466 (1992).
23. S. Zangooie, R. Jansson, and H. Arwin, "Electrochemical tailoring and optical investigation of advanced refractive index profiles in porous silicon layers," *Mat. Res. Soc. Symp. Proc.*, **557**, 195–200 (1999).
24. S. Zangooie, R. Jansson, and H. Arwin, "Reversible and irreversible control of optical properties of porous silicon superlattices by thermal oxidation, vapor adsorption and liquid penetration," *J. Vac. Sci. Technol. A*, **16**, 2901–2912 (1998).
25. S. Zangooie, P.O.Å. Persson, J.N. Hilfiker, L. Hultman, and H. Arwin, "Microstructural and infrared optical properties of electrochemically etched highly doped 4H-SiC," *J. Appl. Phys.*, **87**, 8497–8503 (2000).
26. T. Yamaguchi, J. Lafait, A. Bichri, and K. Driss-Kodja, "Effective media equivalent to an asymmetric multilayer and to a rough interface," *Appl. Opt.*, **30**, 489–494 (1991).
27. K. Brudzewski, "Ellipsometric investigations of a substate-thin film system with rough boundaries using the equivalent film theory," *Thin Solid Films*, **61**, 183–191 (1979).
28. D.E. Aspnes and J.B. Theeten, "Investigation of effective-medium models of microscopic surface roughness by spectroscopic ellipsometry," *Phys. Rev. B*, **20**, 3292–3302 (1979).
29. D. Rönnow, S.K. Anderson, and G.A. Niklasson, "Surface roughness effects in ellipsometry: comparison of truncated sphere and effective medium models," *Opt. Mat.*, **4**, 815–821 (1995).
30. D. den Engelsen, "Optical anisotropy in ordered systems of lipids," *Surf. Sci.*, **56**, 272–280 (1976).
31. A.Y. Tronin and A.F. Konstantinova, "Ellipsometric study of the optical anisotropy of lead arachidate Langmuir films," *Thin Solid Films*, **177**, 305–314(1989).
32. L.A.A. Pettersson, F. Carlsson, O. Inganäs, and H. Arwin, "Spectroscopic ellipsometry studies of optical properties of doped poly(3, 4-ethylenedioxythiophene): An anisotropic metal," *Thin Solid Films*, **313–314**, 356–361 (1998).
33. C.M. Ramsdale and N.C. Greenham,"Ellipsometric determination of anisotropic optical constants in electroluminescent conjugated polymers" *Adv. Mater.*, **14**, 212–215 (2002).
34. Y. Sano, "Optical anisotropy of bovine serum albumin," *J. Coll. Int. Sci.*, **124**, 403–406 (1988).
35. M. Schubert, "Polarization-dependent optical parameters of arbitrarily anisotropic homogeneous layered systems," *Phys. Rev. B*, **53**. 4265–4274 (1996).
36. H. Arwin, S. Welin-Klintström, and R. Jansson, "Off-null ellipsometry revisited: basic considerations for measuring surface concentrations at solid/liquid interfaces," *J. Coll. Int. Sci.*, **156**, 377–382 (1993).

37. G. Jin, R. Jansson, and H. Arwin, "Imaging ellipsometry revisited: developments for visualization of thin transparent layers on silicon substrates," *Rev. Sci. Instr.*, **67**, 2930–2936 (1996).
38. N. Holmström, A. Askendal, and P. Tengvall, "In vitro studies on blood protein adsorption to gold and pyrolytic carbon at pre-set electrical potentials," *Coll. Surf. B*, **11**, 265–271 (1998).
39. J-W. Benjamins, B. Jönsson, K.Thuresson, and T. Nylander, "New experimental setup to use ellipsometry to study liquid-liquid and liquid-solid interfaces," *Langmuir*, **18**, 6437–6444 (2002).
40. Th. E. Tiwald, D. W. Thompson, J. A. Woollam, and S. V. Pepper, "Determination of the mid-IR optical constants of water and lubricants using IR ellipsometry combined with an ATR cell," *Thin Solid Films*, **313–314**, 718–721 (1998).
41. M. Poksinski, H. Dzuo, J-O. Järrhed, and H. Arwin, "Total internal reflection ellipsometry," *Proceedings of Eurosensors XIV*, Copenhagen, August 27–30, 2000.
42. S. Rekveld, "Ellipsometric studies of protein adsorption onto hard surface in a flow cell," 1997, *Fedobruk*, Enschede.
43. I. Lundström, "Real-time biospecific interaction analysis," *Biosensors & Bioelectronics*, **9**, 725–736 (1994).
44. U. Jönsson, I. Rönnberg, and M. Malmqvist, "Flow-injection ellipsometry – an in situ method for the study of biomolecular adsorption and interaction at solid surfaces," *Coll. Surf.*, **13**, 333–339 (1985).
45. J.L. Ortega-Vinuesa, P. Tengvall, B. Wälivaara, and I. Lundström, "Stagnant versus dynamic conditions: a comparative adsorption study of blood proteins," *Biomat.*, **19**, 251–262 (1998).
46. C.A-C. Karlsson, M.C. Wahlgren, and A.C. Trägårdh, "Non-invasive monitoring of protein adsorption and removal in a turbulent flow cell," *Coll. Surf. B*, **20**, 9–25 (2001).
47. B. Arkles, "Tailoring surfaces with silanes," *Chemtech.*, **7**, 766–778 (1977).
48. E.P. Pluddeman, "Chemistry of silane coupling agents," in *Silylated Surfaces*, (D. E. Leyden, Ed.). Gordon & Breach, New York (1980) pp. 31–53.
49. S. Welin-Klintström, A. Askendal, and H. Elwing, "Surfactant and protein interactions on wettability gradient surfaces," *J. Coll. Int. Sci.*, **158**, 188–194 (1993).
50. S. Welin-Klintström, R. Jansson, and H. Elwing, "An off-null ellipsometer with lateral scanning capability for kinetic studies at liquid-solid interfaces," *J. Coll. Int. Sci.*, **157**, 498–503 (1993).
51. J. Mårtensson and H. Arwin, "Interpretation of spectroscopic ellipsometric data on protein layers on gold including substrate-layer interactions," *Langmuir*, **11**, 963–968 (1995).
52. V. Reipa, A.K. Gaigalas, and V.L. Vilker, "Spectroscopic real-time ellipsometry of putidaredoxin adsorption on gold electrodes," *Langmuir*, **13**, 3508–3514 (1997).
53. A. Ulman, "Formation and structure of self-assembled monolayers," *Chem. Rev.*, **96**, 1533–1554 (1996).

54. F. Schreiber, "Structure and growth of self-assembling monolayers," *Progr. Surf. Sci.*, **65**, 151–256 (2000).
55. F. Höök, J.Vörös, M. Rodahl, R. Kurrat, P. Böni, J.J. Ramsden, M. Textor, N.D. Spencer, P. Tengvall, J. Gold, and B. Kasemo, "A comparative study of protein adsorption on titanium oxide surfaces using in situ ellipsometry, optical waveguide lightmode spectroscopy, and quartz crystal microbalance/dissipation," *Coll. Surf. B*, **24**, 155–170 (2002).
56. H. Arwin, I. Lundström, S. Arielly, and G. Claeson, "Orientation of a tripeptide on platinum," *Langmuir*, **6**, 1551–1557 (1990).
57. K. Järrendahl and H. Arwin, "Multiple sample analysis of spectroscopic ellipsometry data of semi-transparent films," *Thin Solid Films*, **313–314**, 114–118 (1998).
58. F. Tiberg and M. Landgren, "Characterization of thin nonionic surfactant films at the silica/water interface by means of ellipsometry," *Langmuir*, **9**, 927–932 (1993).
59. I. Lundström and H. Elwing, "Simple kinetic models for protein exchange reactions on solid surfaces," *J. Coll. Int. Sci.*, **136**, 68–84 (1990).
60. I. Lundström, B. Ivarsson, U. Jönsson, H. Elwing, "Protein adsorption and interaction at solid surfaces," in *Polymer Surfaces and Interfaces*, John Wiley & Sons Ltd, (W.J Feast, H.S Munro, Eds), 1987, pp. 201–230.
61. M. Wahlgren, T. Arnebrant, and I. Lundström, "The adsorption of lysozyme to hydrophilic silicon oxide surfaces: comparison between experimental data and models for adsorption kinetics," *J. Coll. Int. Sci.*, **175**, 506–514 (1995).
62. P. Tengvall, I. Lundström, and B. Liedberg, "Protein adsorption studies on model organic surfaces: an ellipsometric and infrared spectroscopic approach," *Biomaterials*, **19**, 407–422 (1998).
63. H. Elwing, "Protein absorption and ellipsometry in biomaterial research," *Biomaterials*, **19**, 397–406 (1998).
64. B. Ivarsson and I. Lundström, "Physical characterization of protein adsorption on metal and metaloxide surfaces," *CRC Crit. Rev. Biocompat*, **2**, 1–96 (1986).
65. C.E. Giacomelli, M.J. Esplandiú, P.I. Ortiz, M.J. Avena, and C.P. De Pauli, "Ellipsometric study of bovine serum albumin adsorbed onto Ti/TiO_2 electrodes," *J. Coll. Int. Sci.*, **218**, 404–411 (1999).
66. B. Lassen and M. Malmsten, "Competitive protein adsorption at plasma polymer surfaces," *J. Coll. Int. Sci.*, **186**, 9–16 (1997).
67. M-R. Sierakowski, R.A. Freitas, J. Fujimoto, and D.F.S. Petri, "Adsorption behavior of oxidized galactomannans onto amino-terminated surfaces and their interaction with bovine serum albumin," *Carbohydrate Pol.*, **49**, 167–175 (2002).
68. S.M. Ma, D.L. Coleman, and J.D. Andrade, "Ellipsometry studies of albumin films on tantalum oxide and SiO_2," *Surf. Sci.*, **56**, 117–125 (1976).
69. D. Beaglehole, B. Webster, and S. Werner, "Ellipsometry study of the adsorption of molecules at electrolyte interfaces with gold and stainless steel," *J. Coll. Int. Sci.*, **202**, 541–550 (1998).

70. B.A. Ivarsson, P-O. Hegg, K.I. Lundström, and U. Jönsson, "Adsorption of proteins on metal surfaces studied by ellipsometric and capacitance measurements," *Coll. Surf.*, **13**, 169–192 (1985).
71. M. Malmsten, "Ellipsometry studies of the effects of surface hydrophobicity on protein adsorption," *Coll. Surf. B: Interfaces*, **3**, 297–308 (1995).
72. H. Nygren, H. Arwin, and S. Welin-Klintström, "Nucleation as the rate-limiting step in the initial adsorption of ferritin at a hydrophobic surface," *Colloids and Surfaces A*, **76**, 87–93 (1993).
73. P. Tengvall, A. Askendal, I. Lundström, and H. Elwing, "Studies of surface activated coagulation: antisera binding onto methyl gradients on silicon incubated in human plasma *in vitro*," *Biomaterials*, **13**, 367–374 (1992).
74. M. Malmsten, "Studies of serum protein adsorption at phospholipid surfaces in relation to intravenous drug delivery," *Coll. Surf.*, **159**, 77–87 (1999).
75. P. Tengvall, A. Askendal, and I. Lundström, "Temporal studies on the deposition of complement on human colostrum IgA and serum IgG immobilized on methylated silicon," *J. Biomed. Mat. Res.*, **35**, 81–92 (1997).
76. B. Kasemo and J. Lausmaa, "Material-tissue interfaces: the role of surface properties and processes," *Environ. Health Perspect.*, **102**, 41–45 (1994).
77. S. Zangooie, R. Bjorklund, and H. Arwin, "Protein adsorption in thermally oxidized porous silicon layers," *Thin Solid Films*, **313–314**, 825–830 (1998).
78. H. Arwin, M. Gavutis, J. Gustafsson, M. Schultzberg, S. Zangooie, and P. Tengvall, "Protein adsorption in thin porous silicon layers," *Phys. Stat. Sol.* (*a*), **182**, 515–520 (2000).
79. R.M.A. Azzam, P.G. Rigby, and J.A. Krueger, "Kinetics of protein adsorption and immunological reactions at a liquid/solid interface by ellipsometry," *Phys. Med. Biol.*, **22**, 422–430 (1977).
80. M. Lestelius, B. Liedberg, and P. Tengvall, "*In vitro* plasma protein adsorption on ω-functionalized alkanethiolate self-assembled monolayers," *Langmuir*, **13**, 5900–5908 (1997).
81. M. Poksinski, H. Dzuho, and H. Arwin, "Copper corrosion monitoring with total internal reflection ellipsometry," *J. Electrochem. Soc.* **150**, B536–539 (2003).
82. P. Westphal and A. Bornmann, "Biomolecular detection by surface plasmon enhanced ellipsometry," *Sens. Act. B*, **84**, 278–282 (2002).
83. M. Poksinski, J.-O. Järrhed, and H. Arwin, "*In situ* monitoring of adsorption from milk on metal surfaces using total internal reflection ellipsometry," *Sensors and Actuators*, **B 94**, 247–252 (2003).
84. A. Hamnett, "Ellipsometric techniques for the characterization of electrode surfaces," *J. Chem. Soc. Far. Trans.*, **89**, 1593–1607 (1993).
85. I.C. Hahn Berg, D. Muller, T. Arnebrant, and M. Malmsten, "Ellipsometry and TIRF studies of enzymatic degradation of interfacial proteinaceous layers," *Langmuir*, **17**, 1641–1652 (2001).
86. S. Engström and K. Bäckström, "Ellipsometry as a tool to study detergency at hard surfaces," *Langmuir*, **3**, 568–574 (1987).
87. L-T. Lee, B.K. Jha, M. Malmsten, and K. Holmberg, "Lipase-surfactant interactions studied by neutron reflectivity and ellipsometry," *J. Phys. Chem. **B***, **103**, 7489–7494 (1999).

88. I.C. Hahn Berg, S. Kalfas, M. Malmsten, and T. Arnebrant, "Proteolytic degradation of oral biofilms *in vitro* and *in vivo*: potential of proteases orginating from *Euphausia superba* for plaque control," *Eur. J. Oral Sci.*, **109**, 316–324 (2001).
89. I.C. Hahn Berg, U.M. Elofsson, A. Joiner, M. Malmsten, and T. Arnebrant, "Salivary protein adsorption onto hydroxyapatite and SDS-mediated elution studied by *in situ* ellipsometry," *Biofouling*, **17**, 173–187 (1998).
90. L. Lindh, P-O. Glantz, P-E. Isberg, and T. Arnebrant, "An *in vitro* study of initial adsorption from human parotid and submandibular/sublingual resting saliva at solid/liquid interfaces," *Biofouling*, **17**, 227–239 (2001).
91. L. Lindh, P-O. Glantz, N. Strömberg, and T. Arnebrant, "On the adsorption of human acidic proline-rich proteins (PRP-1 and PRP-3) and statherin at solid/liquid interfaces," *Biofouling*, **18**, 87–94 (2002).
92. C. Karlsson, A-C. Wahlgren, M.C. Trägårdh, and A. Christian, "Some surface-related aspects of the cleaning of new and reused stainless-steel surfaces fouled by protein," *Int. Dairy J.*, **8**, 925–933 (1998).
93. M. Lakamraju, J. McGuire, and M. Daeschel, "Nisin adsorption and exchange with selected milk proteins at silanized silica surfaces," *J. Coll. Int. Sci.*, **178**, 495–504 (1996).
94. T. Nylander, F. Tiberg, and N.M. Wahlgren, "Evaluation of the structure of adsorbed layers of β-casein from ellipsometry and surface force measurements," *Int. Dairy J.*, **9**, 313–317 (1999).
95. J. Örnebro, M. Wahlgren, A-C. Eliasson, R.J. Fido, and A.S. Tatham, "Adsorption of α-, β-, γ- and ω-gliadins onto hydrophobic surfaces," *J. Cereal Sci.*, **30**, 105–114 (1999).
96. F. Chao and M. Costa, "The electrochemical interface, a stratified medium studied by ellipsometry," *Surf. Sci.*, **135**, 497–520 (1983).
97. C. Werner, K-J. Eichhorn, K. Grundke, F. Simon, W. Grählert, and H-J. Jacobasch, "Insight on structural variations of protein adsorption layers on hydrophobic fluorohydrocarbon polymers gained by spectroscopic ellipsometry (part I)," *Coll. Surf. A: Physicochemical and Engineering Aspects*, **156**, 3–17 (1999).
98. R.J. Marsh, R.A.L. Jones, and M. Sferrazza, "Adsorption and displacement of a globular protein on hydrophilic and hydrophobic surfaces," *Coll. Surf. **B***, **23**, 31–42 (2002).
99. M. Vinnichenko, Th. Chevolleau, M.T. Pham, L. Poperenko, and M.F. Maitz, "Spectroellipsometric AFM and XPS probing of stainless steel surfaces subjected to biological influences," *Appl. Sufr. Sci.*, **8168**, 1–10 (2002).
100. D. van Noort, J. Rumberg, E.W.H. Jager, and C-F. Mandenius, "Silicon based affinity biochips viewed with imaging ellipsometry," *Meas. Sci. Tech.*, **11**, 801–808 (2000).
101. A. Eing and M. Vaupel, "Imaging ellipsometry in biotechnology," *Nanofilm surface analysis*, July 2002.
102. Th. Ericson, K.M. Pruitt, H. Arwin, and I. Lundström, "Ellipsometric studies of film formation on tooth enamel and hydrophilic silicon surfaces," *Acta. Odont. Scand.*, **40**, 197–201 (1979).

103. M. Malmsten, G. Siegel, and A. Becker, "A model substrate for ellipsometry studies of lipoprotein deposition at the endothelium," *J. Coll. Int. Sci.*, **240**, 372–374 (2001).
104. M. Malmsten, G. Siegel, and W.G. Wood, "Ellipsometry studies of lipoprotein adsorption," *J. Coll. Int. Sci.*, **224**, 338–346 (2000).
105. D.J. Brink, and M.E. Lee, "Thin-film biological reflectors: optical characterization of the *Chrysiridia croesus* moth," *Appl. Opt.*, **37**, 4213–4217 (1998).
106. D.J. Brink, and M.E. Lee, "Ellipsometry of diffractive insect reflectors," *Appl. Opt.*, **35**, 1950–1955 (1996).
107. J.M. Bueno, and J. Jaronski, "Spatially resolved polarization properties for *in vitro* corneas," *Ophthal. Physiol. Opt.*, **21**, 384–392 (2001).
108. J.A. Paczka, D.S. Friedman, H.A. Quigley, Y. Barron, and S. Vitale, "Diagnostic capabilities of frequency-doubling technology, scanning laser polarimetry, and nerve fiber layer photographs to distinguish glaucomatous damage," *Amer. J. Ophthalmology*, **131**, 188–197 (2001).
109. H. Arwin, "Is ellipsometry suitable for sensor applications?," *Sensors & Actuators A*, **92**, 43–51 (2001).
110. G. Jin, P. Tengvall, I. Lundström, and H. Arwin, "A biosensor concept based on imaging ellipsometry for visualization of biomolecular interactions," *Anal. Biochem.*, **232**, 69–72 (1995).
111. S. Guo, R. Rochotzki, I. Lundström, and H. Arwin, "Ellipsometric sensitivity to halothane vapors of hexamethyldisiloxane plasma polymer films," *Sensors & Actuators **B** Chemical*, **44**, 243–247 (1997).
112. G. Wang, H. Arwin, and R. Jansson, "An optical gas sensor based on ellipsometric readout," *IEEE Sensor Journal*, **3**, 739–743 (2003).
113. B. Trotter, G. Moddel, R. Ostroff, and G.R. Bogart, "Fixed-polarizer ellipsometry: a simple technique to measure the thickness of very thin films," *Opt. Eng.*, **38**, 902–907 (1999).
114. *Biophotonics International*, July 36–37 (1999).
115. R.M. Ostroff, D. Maul, G.R. Bogart, S. Yang, J. Christian, D. Hopkins, D. Clark, B. Trotter, and G. Moddel, "Fixed polarizer ellipsometry for simple and sensitive detection of thin films generated by specific molecular interactions: applications in immunoassays and DNA sequence detection," *Clin. Chem.*, **44**, 2031–2035 (1998).
116. R.M. Ostroff, D. Hopkins, A.B. Haeberli, W. Baouchi, and B. Polisky, "Thin film biosensor for rapid visual detection of nucleic acid targets," *Clin. Chem.*, **45**, 1659–1664 (1999).
117. K. Spaeth, A. Brecht, and G. Gauglitz, "Studies on the biotin-avidin multilayer adsorption by spectroscopic ellipsometry," *J. Coll. Int. Sci.*, **196**, 128–135 (1997).
118. J. Benesch, A. Askendal, and P. Tengvall, "The determination of thickness and surface mass density of mesothick immunoprecipitate layers by null ellipsometry and protein^{123} Iodine labeling," *J. Coll. Int. Sci.*, **249**, 84–90 (2002).
119. Z. Yang, W. Frey, T. Oliver, and A. Chilkoti, "Light-activated affinity micropatterning of proteins on self-assembled monolayers on gold," *Langmuir*, **16**, 1751–1758 (2000).

120. A. Röseler, and E-H. Korte, "Infrared Spectroscopic Ellipsometry," in *Handbook of vibrational spectroscopy*, John Wiley & Sons Ltd, Chichester (2002) pp. 1–26.
121. C.W. Meuse, "Infrared spectroscopic ellipsometry of self-assembled monolayers," *Langmuir*, **16**, 9483–9487 (2000).
122. A. Röseler, R. Dietel, and E-H. Korte, *Mikrochim. Acta (Suppl)*, **14**, 657 (1997).
123. L.M. Karlsson, H. Arwin, and J.A. Woollam, "Identification and quantification of human carbonic anhydrase II in porous silicon by infrared ellipsometry," *7th Int. Conf. on Nanometer-scale Science and Technology and 21st European Conference on Surface Science (NANO-7 ECOSS-21)*, 2002, abstracts p. 56.

Index

Index terms	Links
A	
a-Si	101 146 159 164 168 179 185 186 187 188 189 191 192 193
Abelés matrices	247 248 250 251
absorbing solids	110 119 123 229
absorption	76 78 88 94 99 107 120 123 128 136 138 143 149 155 159 162 165 169 172 174 179 180 186 194 196 201 205 227 234 256 260 494 659 683 696 698 723 725 727 733 739 741 745 748 749 757
absorption coefficient	99 101 107 143 156 159 162 172 174 284 589 776 811
absorption ellipsoids	120 123
absorption tensor	120
accuracy	3 13 16 44 251 254 255 259 435 466 467 476 478 482 494 496 497 685 689 697 729
achromatic compensator	485 496
acquisition time	486 489 490 494 495 509
adsorption kinetics	800 806 822 830 851
Airy formula	246
Al_2O_3	689 690 694 714 716 745
albumin	805 810 813 816 830 833 839 843 846 848 851
aluminum	98 100 159 164 173 177 209 220 221 224 716
ambient light	342 363 498

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
amorphous	29 101 143 159 164 170 178 184 189 190 191 193 228 233 253 254 256 760
amorphous materials	29 253 255 274 295
amorphous semiconductor	144 160 180 189 278
amorphous silicon	101 146 159 164 253 271 413
amorphous silicon-germanium	101 159
analyzer offset	522 523 534 544 557 558
anisotropic	3 10 28 102 105 110 114 119 121 123 128 137 157 167 227 238 241 244 245 248 249 250 252 451 454 461 462 637 646 647 650 651 658 670 671 676 680 684 685 688 691 692 695 696 701 705 711 713 740
anisotropic layered structures	637 652
anisotropic layered systems	637
anisotropic materials	3 28 102 238 251 285 562 650 652 685 710 715
anisotropic sample	245 426 454 462 637 640
anisotropic solids	105 110 114 119 123 128 137
anisotropic systems	248 637
anomalous dispersion	160
anti-reflective	741 743
arachidic acid	812 847
arbitrarily oriented anisotropy	638
Archer	480
array detector	342 357 481 496 499 502 505 524 541 547 562
artifact minimization	96
ATR	136
attenuated total reflection	136

This page has been reformatted by Knovel to provide easier navigation.

Index terms **Links**

azimuth modulation currents 485
azimuthal alignment errors 266

B

b-casein 839 853
b-lactoglobulin 841
Babinet-Soleil 117
Babinet-Soleil compensator 311 349 426
backside effects 676 680 682 685
band gap 128 145 179 180 189 196 203 211 229 253 254 259 689 696 697 698
band structure 94 97 99 150 153 176 201 202 207 209 210 211 213 217 220 227 229 234 258 695 733 749
band structure critical points 153
bandgap 6 97 145 161 179 194 212 502 514 727 749 753
barrier energies 608
barrier heights 608 610
basis function 457 463 469 472 473 476
basis function coefficients 466 469 472
basis set 463 465
Berek compensator 314
Berreman effect 787 791 793
Berreman equation 249
Berreman matrix 249
BESSY 721 726 739 740 749
beta-lactoglobulin 839

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
biaxial	29 113 117 120 123 638 650 651 657 661 663 669 672 683 696 697 702 713 714 716
bioadsorption	801 821 826
biocompatibility	833
bioellipsometry	802
biofouling	837 853
biolayer	803 806 817 819
biological	800 804 816 819 824 827 832 835 843 846 848 853
biomolecule	800
biosensor	844 854
birefringence	28 110 116 120 128 439 479 480 638 650 651 653 671 672 678 679 684 695 713 727 728 740 746
birefringent	110 480 652 653 655 684 691 702 710 726 746 747
birefringent crystal	309 316
Bloch wavefunctions	138 144 146
bond polarizability	573
Booker quartic	652
bound charges	20
bound electrons	141 176 177
bovine serum albumin	810 813 839 846 848 851
Brewster	75 78 87 677 678 691 725
Brewster angle	75 78 80 87 282 391 676 690 768 774 780 793 796
Brewster condition	725
Brillouin zone	148 152 155 194 221 696
Brillouin zone	148 152 155 695
broadening energy	146 160 179 260
broadening parameter	166 168 195 197 200 202 205 216 224 259

This page has been reformatted by Knovel to provide easier navigation.

Index terms **Links**

Bruggeman 191 192 261
Bruggeman effective medium approximation 273 287 289 616
Bruggeman EMA 261 813 817 840

C

c-Si 284 412 573 591
calcite crystal 301
calibrating 13
calibration 435 445 461 469 474 476 477 480 483 497 502 505 508 509 511 513 517 519 520 646 701 702 713
calibration sample 519 545
Canada balsam 305
Cauchy 25 255 441 445 470 471 683 694 706 708 739
Cauchy expansion 255 445 470
causality 20 22 24 100 124
charge density 20 103 131 175 655
chi-by-eye 263
chiral biaxial films 663
chiral liquid crystal 646 670 684 702 711
cholesteric liquid crystal 664 714
chromium 743
circular birefringence 110 121 124 552 561
circular polarizations 8 36 39 42 46 57 124
classical Drude models 172
Clausius-Massotti formula 599
coalesced film regime 225
collimated beam 18

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
compensator	10 16 433 469 483 485 497 502 513 647 677 679 682 691 701 713
complex conductivity	21 24 31
complex dielectric function	31 94 108 119 125 138 159 172 209 227 252 254 697 732
complex optical conductivity	99 108 130
complex permittivity	4 21 29
complex reflection coefficients	238 239 241 246 251
complex refractive index	5 24 26 29 67 70 246 247 252 734 735
compound quarter-wave plate	326
conduction band	101 144 148 155 165 169 175 176 179 696 698 700 732 733 749
conduction band electrons	609
conduction band states	101 144 146 148 157 165 695 732
conduction electron current	109
conduction electrons	101 104 108 126 163 166 227
conductivity	20 24 53 99 105 130 651 689
confined electron systems	159 164
consistency checks	529 551
constructive interference	319
convergence routines	268
correlated errors	270 273 275
correlated parameters	271 275
correlation coefficients	99 228 475
critical point	98 151 169 179 182 194 195 197 200 205 213 214 216 219 220 228 253 258 737
critical point features	253

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
cross correlation coefficients	270
cross-polarization	241 244 245 248 474 479 644
cross-polarization ratios	244
cross-polarization reflection coefficients	241 248
crystal-quartz	438
crystalline protrusions	619
crystalline quartz	316
crystalline silicon	190 197 202 509
CuPt-ordered films	672 697
current densities	19 103

D

D_2 lamp	495 727
damping parameter	166
dangling bonds	603
dark current	490 493 498
data analysis	93 102 237 238 239 255 435 466 475 488 499 637 646 676 678 680 682 690 729 751
data collection	483 502 504 513 637 727
data reduction	458 461 482 483 497 502 505 519 520
degree of polarization	16 58 62 65 520 723
dental plaque	839
depolarization	16 18 65 78 87 132 240 241 245 450 453 478 479 639 676 678 729 745
depolarization effects	16 18 266 290 295 453 678 745
depolarization elements	450
depolarize	66 239 240 241 243
depolarizing	458 642
depolarizing elements	59

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links									
detection systems	489	497	498	501						
detector dark current	339	493	498							
detector elements	481									
detector noise	386	388	390							
detector polarization	341	344	488	497	524					
Deuterium	495	727								
dextro-rotatory crystal	122									
diamond	498	749	758							
diamond-like	749									
dichroism	100	110	120	123	128	479	480	751		
dielectric function	21	24	28	93	96	100	102	105	107	110
	127	131	137	141	146	150	154	160	163	166
	175	188	192	197	200	209	213	214	216	225
	227	251	255	257	509	650	659	665	670	672
	674	676	678	686	689	697	703	710	713	
dielectric tensor	27	100	102	110	112	114	119	121	123	138
	157	651	659							
diffraction grating	318									
diffraction grating	768									
diffraction orders	435									
diffusivity	582	594	612							
digital waveform analysis	466									
digitization speeds	449									
direct band edges	260									
direct band-gap	260									
direct interband transitions	101	109	125	144	146	148	175	176	225	
dispersion	5	21	91	100	124	129	131	138	155	160
	227	232	234	477	501	699	700	735	737	741
dispersion equation	5									
dispersion relation	135	739	741	795						
dispersion relationships	124	132	134							

This page has been reformatted by Knovel to provide easier navigation.

Index terms **Links**

DLC 749
DNA 830 845 854
double refraction 301 327
doubly refracting crystal 302
driving element 436 440
Drude 100 125 129 138 159 166 172 176 221 222 228 232 233 255
Drude expression 255
Drude formulas 174
Drude free electron 129 138 166
Drude Free Electron Model 129
Drude model 125 129 172
Drude theory 130
DTGS 775
dual-beam photoconductive spectroscopy 186
dual-rotating compensator 550 552
dynode bias voltage 447

E

effective mass theorem 158
effective medium 166 191 192 225 226 261 685
effective medium approximation 260 262 290
effective medium approximation 812
electric charges 4 99
electric field intensity 6 19 22 27
electric induction 23
electric intensity vector 33 73
electric susceptibility 23 108 124

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
electrical isolation	569
electro-optic modulators	435
electrochemical interface	803 853
electromagnetic theory	4 99
electron cyclotron resonance	595 625
electron relaxation times	97
electron tunneling	621
electron-impurity scattering	147
electron-phonon coupling	154
electron-phonon scattering	147
electronic states	94 97 100 138 144 148 150 153 169 570 574 602 610 629
electronic transitions	97 101 142 146 178 194 195 197 198 200 207 211 212 228 721 732 740 749 754 755
ellipsometric period	576 588
elliptic polarizer	50 58 60
ellipticity angle	34 42 44 57 62 520
EMA	261 685 689
endpoint detection	482
energy flux	31 40
error analysis techniques	237
Euler angles	658 659 683 686 689 703 710
EUV	721 724 725 726 757
exact numerical inversion	95
extraordinary ray	302 304 310 316

F

Fabry-Perot resonators	683
Faraday cell	111 483 485
Faraday cell modulation	483

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
Faraday modulators	434 435
fast and slow axes	11 52 66
fatty acid	802
feedback loop	482
ferritin	806 830 832 840 848 852
ferrocholesteric	650
ferrosmectic	650
fiber optics	450 491 727
fibrinogen	833 835 841 843
figure of merit	238 239
figure of merit function	238 264 268 270
fill factor	806 812 814
film formation kinetics	588
film stress	597 599
fine-grain polycrystalline	286
finite band widths	237
fitted data representation	237
fitting algorithms	238 239 475
fluorocarbons	743
FNCO	623 630
focused beam	564
Forouhi and Bloomer amorphous model	272
Forouhi and Bloomer approximation	256 463 480 488 492 505 713
Forouhi and Bloomer model	273 280
Fourier analysis	336 347 355 463 480 488 492 505 712
Fourier coefficients	15 42 246 468 469 488 492 504 505 511 512 514 515 520 522 682 683 691
Fourier transform	15 461 462 466 493

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
fractal analysis	625
free charge carriers	20
free current density	104
free electron contribution	175 176
free electron densities	97
free electron oscillations	132 177
free electron plasma energy	167
free electron systems	172 175
free electrons	104 129 131 136 163 177 223 227
free-electron metals	125
frequency-dependent dielectric	
function	127
Fresnel equations	95 509
Fresnel reflection coefficients	246 247 734
Fresnel rhomb	313 326 770
functionalization	800 803
fundamental bandgap	97
fundamental dielectric response	143
fundamental gap	732
fused quartz	438
fused silica	136 253 441

G

g-globulin	839 848
GaAs	101 159 168 178 194 207 209 216 652 695 698 699 701 717 732 733 740
gallium arsenide	100 101 159 168 209
gallium-aluminum arsenide	98

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links									
gallium-arsenide	100									
gate dielectric	721	726								
Gaussian broadening	147	152	197	211	216	220				
Gaussian lineshape function	195									
Gaussian oscillators	739	740								
GE	146	150	189	637	644	656	657	671	672	675
	677	682	684	688	689	692	693	695	697	698
	701	710	711	717						
Generalized Critical Point Model	197	206	220							
Generalized Ellipsometry	277	295	426	480	637	641	643	645	647	649
	651	653	655	657	659	661	663	665	667	669
	671	673	675	677	679	681	683	687	689	691
	693	695	697	699	701	703	705	707	709	711
	717	747	751	754	802	821				
Generalized Oscillator Model	195	201	202	204	205					
germanium	100	146	159							
Glan-Foucault	119									
Glan-Taylor	306	347								
Glan-Thompson	119									
Glan-Thompson prism	306									
Glan-type prism	307									
gliadin	839									
gliadins	839	853								
glowbar	763									
glutenins	839									
gold	100	412	708	710	785	791	800	806	811	825
	830	833	835	837	847	850	854			
goniometer	324	352								
goodness of fit	238									

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
graduating the circle	321
graphite-like	749
grating coupler	134
grating spectrograph	491 492 501
gratings	435 711 712
growth dynamics	576
gyration vector	121
H	
H terminated	601
handedness	35 37 42 53 648 667 683 702 706 729 120
hardmasks	741
hemoglobin	839 848
highly anisotropic systems	637
human saliva	839
human serum albumin	805 830 833 843 846 848
Huygens	301
hydrogen terminated	285 601
hydrogenated amorphous silicon	146 164
hydrophilic surface	825
hydrophobic surface	825 852
hydroxyapatite	839 853
I	
Iceland Spar	303
ideal polarizer	9 16 50 55 62 64 365
Image persistence	499 502
imaging ellipsometry	333 428 806 821 842 846 850 853

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
imperfectly collimated beam	18
in situ ellipsometry	449 481 607 612 851 853
in-situ SWE	593
incident beam polarization ellipse	93
incoherent	9 18 55 58 479 639 676 678 680 682 685 690 703
index ellipsoid	113 119 123
index tensor	100 115 120
Indirect Interband Transitions	144 153
induced current	20 22 31
induced current density	20 22 31
induced dipole moment	23 126
infrared ellipsometry	102 390 599 651 661 715 763 769 771 773 775 777 779 781 783 785 787 789 791 793 795 797 820 847 855
infrared spectroscopy	599 763 820
in GaAs detectors	727
inhomogeneous layers	251
initial oxidation regime	583 601
insulators	20 25 100 125 128 159 164 252 282
intensity waveform	436 447 452 455 458 461 463 465 469 472
interband transitions	101 169 172 177 222 223 256 672 753
interface	67 70 75 78 88 101 116 119 136 163 191 192 237 246 508 509 655 669 703 705 715 726 734 735 743
interface charge	29
interface layers	237 262 597
interface roughness	619

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
interference fringes	96
interference region	289
intraband contribution	101 125 138 142 147 157 222 223 224
intraband transitions	176
intrinsic stress	597 606
inverse effective mass tensor	158
inverse Hessian method	268
inversion layer	570
ionic crystals	129 170
ionic transport mechanism	613
iron	837
irradiance waveform	488 491 499 502 504 514
irradiance-function	515 521
isotropic reflecting surface	53 64

J

joint densities of electronic states	97
joint density of states	94 97 138 150 153 156 212 214 218 228 229
Jones algorithm	638
Jones calculus	637 641
Jones matrices	45 48 52 59 240 243 638 639 640 652
Jones matrix	50 639
Jones vector	7 12 45 638 641 644

K

KBr	771 790
Kerr angle	709
Kerr effect	756

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
KK transformation	734
Kramers-Kronig	96 125 137 152 178 181 182 185 186 196 220 229 256 257 258 734 737
Kramers-Kronig consistency	737
Kramers-Kronig integration	256 277
krillase	838
KRS5	439

L

L-P oxidation model	584
Langmuir-Blodgett	671 684 713
large-area sheet polarizers	301
large-grain polycrystalline	286
leakage current	570
left-handed polarization	33
Levenberg-Marquardt	239
Levenberg-Marquardt algorithm	239 268 281
life science	800 804 822 825 829 835 847
linear and circular polarization	35
linear growth	581
linear parabolic model	581
linear polarizer	8 10 17 41 49 52 56 60 63 436 443 643 647
linear retarder	43 52 63 66
linear-parabolic kinetics	583
lipase	839 852
lipoprotein	843 854

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
lithography	721 726 737 739 740 742 743 745 757 759 761
lock-in detection	492
long-range chemical ordering	672 695
longitudinal mode	438 439
Lorentz model equation	254
Lorentz oscillator	101 125 128 132 137 141 159 165 173 179 181 182 193 210 228 229 254 255 257 258 689 701 738 739
Lorentz Oscillator Model	125 129 138 159 165 173 254 272 739
Lorentz-Lorenz	599 815 817
loss function	24
low index planes	605
LRO	695

M

macromolecular adsorption	804 830
magnesium fluoride	727
magneto-optic biaxial materials	661
magneto-optic phenomena	638
magneto-optical effects	651 717
magneto-optical films	661 683
magneto-plasmons	661
Malus	4
manual ellipsometers	311
material relations	19
Maxwell equations	3 19 23 26 30 653 657
Maxwell-Garnett	166 225 226 261
Maxwell-Garnett EMA	818 841

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
McCrackin	431 564 579 634
MCT	766 775
mean squared error	264
measurement errors	14 475
mechanical stress	111 433 434
mercury cadmium telluride	101 159
metal grating	768
metal grid	768
metallic substrates	497
MgF_2	116 439 495 723 724 727 745
milk proteins	839 853
misalignment	475 476 701 702
mixed polarization state	60
model-dependent	276
modulator amplitude	466 471
modulator control voltage	445 471
molar absorption coefficient	776
monochromator	433 446 450 477 481 486 488 489 492 724 727 729 741
monolayers	817 835 839 842 847 850 854
MOS devices	603
MOS structure	603 621
MOSFET	570
Mueller matrices	45 59 63 241 435 442 452 480
Mueller matrix	59 64 242 434 443 451 452 456 466 502
Mueller-Jones matrix	243
Mueller-Stokes formalism	241 242
multi-element retarder	312
multi-order quarter-wave plate	310

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
multichannel ellipsometer	481 489 490 494 499 501 502 505 509 511 514 515 519

N

nanobiotechnology	801
nanotechnology	802
native oxide	509 515 170
near-surface region	237
Newton	95 126 130
Nicol	305 327
nicol prism	305
nisin	839 853
non-absorbing Solids	111 119
non-linear curve fitting	268
nonlinear parameter estimation	263 294
nonlinearity of detection system	15
normal dispersion	160
NSC representation	225 244 245 455
null	12 476 485 486 493 647
null ellipsometer	322 333 340 386 850
null ellipsometry	12 647 821 825 830 832 842 848 854
nulling currents	485
nulling ellipsometers	239
nulling schemes	41

O

one-electron Bloch states	147
one-electron energy states	143
one-electron model	143 150 153

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links									
one-electron wavefunctions	138	144								
one-modulator PME	462	473								
ophthalmology	844	854								
optic axis	113	117	119	122	124	249	474	751		
oplical activity	23	27	100	102	119	332	340	343	348	413
	424	426	511	701	703	711				
optical anisotropy	28	111	707	712	747	820	849			
optical axis	28	250	306	308	310	316	347	354	661	676
	683	685	690	747	751	788				
optical conductivity	105	131								
optical effective mass	158									
optical fiber	256	417								
optical model	97	165	579	586	589	591	593	597	605	618
	739	809	816	826						
optical path length	619									
optical rotation	122									
optical thickness	619	653	677	772	808					
optically anisotropic	110	114	116	157	167					
optically biaxial	117	121	661	715						
optically uniaxial	117	119	121	440						
ordinary ray	302	304	308	316	347					
organic material	800									
orthogonal polarization states	37	40	57	65	78					
orthogonal polarizations	9	46	58	60	75					
oscillating phase retardation	433									
oscillating retarder	434									
oscillator model	170	192	195	203	205	210	211	739	741	742
oscillator strength	141	147	150	179						
oxidation	254	482								

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links									
oxidation of Si	569	572	574	576	578	584	587	593	595	626
oxide overlayers	223									
oxide thickness	283	508	583	601	611	826	829			
P										
parabolic growth	581	618								
parallel acquisition	497									
parallel data acquisition	505									
parallel-band transitions	221	222								
partial polarization	9	16	19	53	55	58	61	66	388	
partial polarizer	61	372								
partially polarized	3	6	9	17	19	45	58	76		
partially polarized light	3	6	17	19	45	59	76	367		
partially polarizing	12									
passivation	569	603	798							
PCSA	9	12	16							
peeling	578									
pellicles	741	744								
PEM	434	455	461	469	472	475	480	763	766	
PEM static retardation	455									
PEM-polarizer pair	445	456								
peptide	820	839								
permanent dipoles	129	170								
permittivity	5	19	650	103						
perovskite	749									
phase modulated ellipsometers	307	433	447	475	479	766				
phase modulation	93	331	335	339	343	428	433	474	483	491
	562									
phase modulation ellipsometers	336	433								

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links								
phase modulator	299	331	336	340					
phase shifters	301								
phase speeds	110	116	120						
phase velocity	71	107							
phase-function	398	400	511	513	520	523			
phase-modulation	491	497	501						
phase-sensitive detection	461	463	465	486					
phospholipid	831	852							
photoconductive	6	164	180						
photocurrent	447								
photodetectors	94	229							
photodiode array	496	498	500	501	502	505			
Photodiode detectors	723	724							
photoelastic modulator	434	436	480	491					
photolithography	726	740							
photomasks	741	743	744						
photometric data	15								
photometric ellipsometer	498								
photometric ellipsometric systems	43								
Photometric ellipsometry	13	330	763						
photometric instrument	491								
photomultiplier tube	333	414	446	484	488	492	494	498	727
photon counting	265								
photon enhanced oxidation	608								
photonic enhancements	608								
photoresist	725	741	743						
photovoltaic	6	188	229						
photovoltaic cells	94								
piezo-birefringence model	696	699							
piezoelectric transducer	316	436	438						

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
Pillings-Bedworth ratio	578
pixel grouping	489 490 496 501 502 509
plane wave	4 19 29 67 70 84 99 119 640 641 642 643 652 655
plane waves	6 27 29 67 71 80 106 641 654 660 714
plane-wave	112 116 644
plane-wave propagation	121
plasma energy	134 160 163 172 175 176 181 223
plasma frequency	26 129 132 135 161 163 166 174 652
plasmon	132 136
plasmon frequency	132
plasmons	132 175 651 661
PME	435 436 442 446 447 452 455 456 461 465 466 469 470
PMT	447 448 449 465
Pockels cells	335
point-by-point fit	683 688 689
Polarimetry	13 17 90 844 854
polarization ellipse	6 33 47 519 520
polarization modulation ellipsometers	433
polarization modulation ellipsometry	284 295 433 435 437 439 441 443 445 449 451 453 455 457 459 461 463 465 467 469 471 473 475 477 479
polarization sensitivity	14 488 489 492 497 511
polarization state	9 13 16 18 35 38 40 42 44 47 53 57 60 64 66 71 76 241 242 433 434 436 442 443 446 453 454 512 641 647 648 703 723 727
polarization state analyzer	433

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
polarization state detector	240 266 300 307 367 375
polarization state generator	240 266 299 434
polarization state modulation	433
polarizer tracking	392 646
polarizing beam splitting prism	307
polyclonal antibodies	831
polycrystalline	95 101 159 178 205 254 638 710
polycrystalline silicon	287
polyethylene	742
polypeptide	839
polysaccharide	830
polysiloxanes	743
polyvinylchloride	839
porous silicon	684 713
Poynting vector	5 40 50 112 116 121
precision	3 13 16 18 95 470 471 473 476 482 485 486 490 492 497 499 638 685
principal crystallographic axis	304
principal indices of refraction	113 115
principal section	305
prism spectrograph	487 490 501
propagation	3 5 12 28 33 45 53 55 59 80 85 99 102 105 110 114 116 119 132 135 227 246 247 480 638 639 642 643 647 650 652 659 661 668 692 703 713
propagation of light	99 102

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
protein	801 805 810 812 814 817 825 828 830 835 837 840 845 848 850
protein monolayer	806
proteoheparan sulfate	843
protrusions	614 617 627
PSA	10 16 433 442 443 445 450 454 458 460 462 471 478
PSCA	12 14
pseudo dielectric function	591
pseudo-dielectric function	282 285 509 756
pseudodielectric function	753
PSG	241 433 434 442 443 445 450 454 458 460 462 472 478
Ptolemy	321
putidaredoxin	841 850
pyroelectric detector	775
pyrolytic carbon	819 850

Q

quantum mechanical concepts	53 178
quantum mechanical description	5 53 100 168
quantum mechanical features	100
quantum mechanical model	100 103 178 207 227
quantum mechanical principles	97 101 137 159 234
quantum mechanical theories	137
quantum mechanics	9 91 232 234 254 164

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
quantum nature	6
quantum theory of solids	138 157
quarter-wave plate	299 308 315 322
quartz	117 121 438 439 511 727 728

R

RAE	15 469 638 639 646 723 726 727 729
random errors	237 265 268 270 276 278 284 474
read-time errors	499
real time analysis	481 482
real time control	482
real time ellipsometry	352 356 358 386 482 492 578 580 596 623
real time feedback	482
real time monitoring	482
real time spectroscopic ellipsometry	164 171 333 335 346 357 387 389 429 483 487 510 562
reflectance	12 16 77 81 85 89 101 136 146 159 162 167 172 174 175 177 180 186 187 188 189 222 227 229 452 504 505 508 509 517 520 644 656 675 715 724 725 726 732 734 735 737
reflection grating	320
Reinberg algorithm	784 790
residual function	511 512 514 515 521 522
residual-function	511 513 514 520
resonance energy	160 166 168 179 195 196 200 201
resonance frequencies	129 164
resonance quality factor	436
response functions	19 125

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
retardation	13 43 63 433 438 454 455 457 462 463 469 471 477 480 647 648 649 682
retardation plate	309 311 371 374
retarder	8 10 17 43 52 66 440 729 730 745 747
right handed	702
right-handed	33 54 640 642 664 697 112
RMS roughness	625
Rochon	116 495 511 723 727
Rochon polarizer	347 373 394 411 413 416 418 511 723 727
rotating analyzer	17 93 239 245 264 285 293 314 323 325 331 336 338 341 343 349 360 365 367 374 378 385 389 394 397 401 411 414 418 424 426 469 488 523 712 724 726 763 765 821 826
rotating analyzer ellipsometry	265 285 350 401 712
rotating compensator	326 426 433 469 497 523 528 535 542 546 550 552 563 826
rotating element ellipsometer	434
rotating element ellipsometers	336 343 433 449 452 462 463 469 470 475
rotating filter	494
rotating polarizer	239 245 314 324 329 331 335 339 349 353 355 357 361 367 369 371 373 375 377 381 387 389 391 401 403 405 407 409 415 417 419 421 423 425 431 469 488 497 502 504 509 519 563 646 723 727
rotating polarizer ellipsometer	342 378 381 396 398 402

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
rotating-analyzer multichannel ellipsometer	523
rotating-compensator	483 502 513 523 529 531 539 542 545 553 558 562 647 712
rotating-polarizer multichannel ellipsometry	498 514 520 523
roughness measurements	625
RPE	15 469 646 727 729
rutile	684 685 692 714 716 751
S	
sample depolarization	240 289 292 478
sapphire	678 684 689 690 694 703 714 716 727 728 751
self-assembled hexadecanethiol	847
self-calibration	534 545 554
self-compensating	483 485 486 487 489 490 493 494 496 497
self-compensating instrument	339 340 486 489 494
Sellmeier	128 255 738 739
Sellmeier approximation	254 270 283
semiconductor alloys	98 259 260 695 715
semiconductor analysis	94
semiconductor critical point	151
semiconductor or dielectric criterion	96

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links									
semiconductors	100	104	125	128	131	138	144	150	157	159
	164	170	178	179	181	182	194	200	201	211
	220	228	229	231	253	259	479	481	514	661
	684	696	712	714	721	732	756	760		
Short Wavelength Ellipsometry	722									
shot noise	494									
Si oxidation kinetics	581	594	601	603	608	612				
Si-SiO_2 interface	569	581	601	603	609	612	615	619	634	636
SiC	740	751								
signal averaging	486	501	502							
silicon	6	80	88	100	146	159	164	167	180	184
	186	190	191	193	194	201	224	254	259	462
	478	480	509	701	710	717	724	735	743	745
silicon carbide	743	818								
silicon dioxide	717	738	809	815	825	831	839	848		
silicon nitride	271									
silicon oxynitride	849									
silicone rubber	782	787	789							
silver	89	100	159	164	175	223	509			
simple quarter-wave plate	308	310	312							
single crystal silicon	412									
single-crystalline bulk materials	95									
single-plate compensator	525									
single-wavelength ellipsometry	309	324	483	721						
SiO_2	87	254	478	735	748	755				
SiO_4 tetrahedra	572	599								
Snellius	778									
sodium dodecyl sulfate	839									

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
source polarization	488 489 492 497 506 519 723
spatial dispersion	22 124 129 155
spectroscopic immersion ellipsometry	614
static retardation	439 442 455 457 462 469 477 480
statistical analysis	205 220 238
steam oxidation	584
steepest descent method	268
stepper motor	322
Stokes vector	56 61 63 242 442 445 452 502 520
Stokes vector spectroscopy	502
Stokes vectors	12 45 56 59 64 435 442 458
Stokes-Mueller formalism	766
strain rate	612
strain-induced retardation	439 471
stray light sources	344 494
stress	32 254 438 439 479 480 696
stress-optic coefficients	253
strong oscillator	780 787 791
sub-stoichiometric	743
supercholesteric	650
Surface cleaning	95 98
surface electronic state	570
surface mass density	805 809 816 829 846 854
surface oxides	254 689
surface plasmons	132
surface reconstructions	285
surface roughness	95 190 237 254 478 685 689
surface wave	787 794
SWE	309 317 324 585 593 605 625

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
Sylvester-Lagrange theorem	652
synchrotron	721 725 726
synchrotron radiation	723 725 775
systematic errors	238 461 474 475 646

T

tabulated data	253 254
Tauc energy gap	258
Tauc expression	180 257
Tauc formulation	257
Tauc gap	180 182 205
Tauc Law	146
Tauc-Lorentz	180 181 229 258 738 739 740
Tauc-Lorentz amorphous model	273
Tauc-Lorentz formulation	257
Tauc-Lorentz model	258 272 275 279 287
Teflon	407 784 823
temperature	94 98 113 130 153 156 164 167 169 178 180 186 192 200 201 206 220 221 223 224 228 230 436 462 477 482 497 498 509 702 713
tetradecyltrimethylammonium bromide	839
thermal oxidation	581 597 599 606 608 611 620 623 626 629 849
thermal oxidation rate	608
thick SiO_2 films	585 588
thickness period	735
thin film materials	94 164 179 253

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
thin SiO_2 films	576 579 588 599 619 623
thin-film silicon	286 295
thiol chemistry	827 831
tilting-compensator	326
tilting-plate compensator	314
time-varying birefringence	316
TiO_2	685 713 714 716 740 745 751
titanium	743
titanium nitride	743
total internal reflection	305 313 347 824 837 850 852
tourmaline	301
tracking	646
trajectories	576 588 619
transmission	12 41 49 60 63 67 72 84 94 109 188 435 452 495 504 513 639 644 653 654 656 657 678 679 691 692 694 695 698 700 704 708 713 714 721 725 726 729 740 741 744 745 751
transmission coefficients	72 85 334 366 537 539 680
transmission grating	319
transverse electric (TE) mode	30
transverse electromagnetic wave	4
transverse magnetic (TM) mode	30
triglycerides	830 838
trypsin	838
tunneling current	748
twisted nematic liquid crystals	664

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
two-modulator PMEs	467 470 472
two-modulator systems	467 470 472 476 477
two-zone measurement	461
Tycho Brahe	321
U	
ultra thin film regime	602
unbiased estimator	263 597
uncorrelated errors	475
uniaxial	29 113 117 122 124 249 250 646 650 651 661 664 672 677 683 685 688 690 691 693 694 696 701 703 704 706 713
Urbach tail	180 184 192
V	
valence band states	146 696
valine gramicidin A	812
van Hove singularities	98 151
Vernier	321 328
viscous flow model	578 598
voltage waveform	465
volume plasmons	133
VUV ellipsometry	721 729 731 735 737 739 743 747 751 753 755 759 761
W	
wave equation	27
wave plate	309 316 799
waveform analysis	461 466 467
waveform digitizer	448
wavelength calibration	267 477 511

This page has been reformatted by Knovel to provide easier navigation.

Index terms	Links
wavelength errors	266
wavelength spread	266 477
wavenumber	763 769 774 779 781 784 788 792 794
weak oscillator	779 787 790
window	492 500 705 727
wire grid	768
WKB approximation	252
Wollaston prism	449 450

Z

zero order beam	319
zero-order quarter-wave plate	310
zone-averaging	646

This page has been reformatted by Knovel to provide easier navigation.