Michael Quinten

A Practical Guide to Optical Metrology for Thin Films

Related Titles

Quinten, M.

Optical Properties of Nanoparticle Systems

Mie and Beyond

2011

ISBN: 978-3-527-41043-9

Osten, W., Reingand, N. (eds.)

Optical Imaging and Metrology

Advanced Technologies

2012

ISBN: 978-3-527-41064-4

Li, F.M., Nathan, A., Wu, Y., Ong, B. S.

Organic Thin Film Transistor Integration

A Hybrid Approach

2011

ISBN: 978-3-527-40959-4

Abou-Ras, D., Kirchartz, T., Rau, U. (eds.)

Advanced Characterization Techniques for Thin Film Solar Cells

2011

ISBN: 978-3-527-41003-3

Haus, J.

Optical Sensors

Basics and Applications

2009

ISBN: 978-3-527-40860-3

Wehrspohn, R. B., Kitzerow, H.-S., Busch, K. (eds.)

Nanophotonic Materials

Photonic Crystals, Plasmonics, and Metamaterials

2008

ISBN: 978-3-527-40858-0

Fujiwara, H.

Spectroscopic Ellipsometry

Principles and Applications

2007

ISBN: 978-0-470-01608-4

Michael Quinten

A Practical Guide to Optical Metrology for Thin Films

WILEY-VCH Verlag GmbH & Co. KGaA

The Authors

Dr. Michael Quinten
Mauritiusstr. 7
52457 Aldenhoven

Cover
An artifically oxidized bismuth crystal superimposed on a background of medieval glass. After a long period in soil alteration formed a thin nanometric film of layer silicates with locally different thickness on the glass. The thickness of the nanometric oxide film on the crystal is locally different too. (Photograph of crystal by U. Quinten; photograph of glass by G.Müller, www.mueller-mineralien.de).

All books published by **Wiley-VCH** are carefully produced. Nevertheless, authors, editors, and publisher do not warrant the information contained in these books, including this book, to be free of errors. Readers are advised to keep in mind that statements, data, illustrations, procedural details or other items may inadvertently be inaccurate.

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de.

© 2013 Wiley-VCH Verlag & Co. KGaA,
Boschstr. 12, 69469 Weinheim, Germany

All rights reserved (including those of translation into other languages). No part of this book may be reproduced in any form – by photoprinting, microfilm, or any other means – nor transmitted or translated into a machine language without written permission from the publishers. Registered names, trademarks, etc. used in this book, even when not specifically marked as such, are not to be considered unprotected by law.

Print ISBN: 978-3-527-41167-2
ePDF ISBN: 978-3-527-66437-5
ePub ISBN: 978-3-527-66435-1
mobi ISBN: 978-3-527-66436-8
oBook ISBN: 978-3-527-66434-4

Cover Design Adam-Design, Weinheim; Germany
Typesetting Thomson Digital, Noida, India
Printing and Binding Markono Print Media Pte Ltd, Singapore

Printed in Singapore
Printed on acid-free paper

To my family

Contents

Preface *XI*

1 Introduction *1*

2 Propagation of Light and Other Electromagnetic Waves *7*
2.1 Properties of Electromagnetic Waves *7*
2.2 Huygens–Fresnel Principle *14*
2.3 Interference of Electromagnetic Waves *15*
2.4 Reflection and Refraction *16*
2.5 Diffraction *21*
2.5.1 Transmission Gratings *27*
2.5.1.1 Lamellar Transmission Gratings *27*
2.5.1.2 Holographic Transmission Gratings *29*
2.5.2 Reflection Gratings *31*
2.5.2.1 Lamellar Reflection Gratings *31*
2.5.2.2 Blazed Gratings *33*
2.5.2.3 Holographic Gratings *34*
2.6 Scattering *34*
2.7 Dielectric Function and Refractive Index *35*
2.7.1 Models for the Dielectric Function *35*
2.7.2 Kramers–Kronig Analysis of Dielectric Functions *49*
2.7.3 Empiric Formulas for the Refractive Index *50*
2.7.4 EMA Models *53*

3 Spectral Reflectance and Transmittance of a Layer Stack *59*
3.1 Reflectance and Transmittance of a Single Layer *59*
3.1.1 Coherent Superposition of Reflected Light *59*
3.1.2 Influence of Absorption on the Layer *65*
3.1.3 Partial Incoherence due to Thick Substrates *69*
3.1.4 Partial Incoherence due to Roughness *72*
3.1.5 Coherent Superposition of Transmitted Light *74*
3.2 Propagating Wave Model for a Layer Stack *75*
3.2.1 Coherent Reflectance and Transmittance of a Layer Stack *76*

3.2.2 Consideration of Incoherent Substrates *78*
3.2.3 Consideration of Surface Roughness *78*
3.2.4 r-t-ϕ Model for a Layer Stack *79*

4 The Optical Measurement *81*
4.1 Spectral Reflectance and Transmittance Measurement *81*
4.2 Ellipsometric Measurement *85*
4.3 Other Optical Methods *88*
4.3.1 Prism Coupling *88*
4.3.2 Chromatic Thickness Determination *92*
4.4 Components for the Optical Measurement *94*
4.4.1 Light Sources *94*
4.4.1.1 Halogen Lamps *94*
4.4.1.2 White Light LED *95*
4.4.1.3 Superluminescence Diodes *96*
4.4.1.4 Xenon High-Pressure Arc Lamps *97*
4.4.1.5 Deuterium Lamps *97*
4.4.2 Optical Components *99*
4.4.2.1 Lenses and Mirrors *99*
4.4.2.2 Polarizers and Analyzers *101*
4.4.2.3 Optical Retarders *102*
4.4.3 Optical Fibers *103*
4.4.4 Miniaturized Spectrometers *107*
4.4.4.1 Gratings *107*
4.4.4.2 Detectors *110*
4.4.4.3 System Properties *115*

5 Thin-Film Thickness Determination *121*
5.1 Fast Fourier Transform *122*
5.1.1 Single Layer *122*
5.1.2 Layer Stack *129*
5.1.3 Accuracy, Resolution, Repeatability, and Reproducibility *130*
5.2 Regression Analysis with χ^2-Test *131*
5.2.1 Method of Thickness Determination *131*
5.2.2 Accuracy, Resolution, Repeatability, and Reproducibility *137*

6 The Color of Thin Films *141*

7 Applications *149*
7.1 High-Reflection and Antireflection Coatings *150*
7.1.1 HR Coatings on Metallic Mirrors *151*
7.1.2 AR Coatings on Glass *152*
7.1.3 AR Coatings on Solar Wafers *153*
7.2 Thin Single- and Double-Layer Coatings *156*
7.2.1 SiO_2 on Silicon Wafers *157*

7.2.2 Si_3N_4 Hardcoat *157*
7.2.3 Double-Layer System *158*
7.2.4 Porous Silicon on Silicon *158*
7.3 Photoresists and Photolithographic Structuring *160*
7.4 Thickness of Wafers and Transparent Plastic Films *163*
7.4.1 Thickness of Semiconductor, Glass, and Sapphire Wafers *163*
7.4.2 Thickness of Transparent Plastic Films *167*
7.4.3 Thickness of Doped Silicon *169*
7.5 Silicon on Insulator *174*
7.6 Thin-Film Photovoltaics *177*
7.6.1 Inorganic Thin-Film Solar Cells *177*
7.6.2 Organic Thin-Film Solar Cells *180*
7.7 Measurement of Critical Dimensions *182*

Numerics with Complex Numbers *187*

Fourier Transform *191*

Levenberg–Marquardt Algorithm *197*

Downhill Simplex Algorithm *199*

References *201*

Index *209*

Preface

The optical response of a thin film is determined by several parameters: its thickness, its optical properties, and the surrounding (other layers, substrates). Among them, its thickness d is the most important. Compared to the vacuum wavelength of light λ, it must have a certain value to establish characteristic features in reflectance, transmittance, or ellipsometric parameters by interference. The size can be reduced by a factor n, with n being the refractive index of the film. The reason is that the optical thickness $t = n \cdot d$ is the intrinsic parameter that must be compared with λ.

The measurement of reflectance, transmittance, or ellipsometric parameters has become a major tool for in-line inspection, process control, and quality control of thin films since it is fast, contactless, nondestructive, and even cheap compared to other methods.

During my successful stint with industry from 2001 till date, I have become acquainted with several aspects of optical thin-film metrology. It is a very fascinating subject since it connects electrodynamics with solid-state physics. The input parameters of any evaluation algorithm are never constant but may vary from one measurement task to the next because the optical material functions strongly depend on film manufacturing, composition, and stochiometry. Film thickness determination then becomes also a question of refractive index determination.

The purpose of this book is to introduce in optical metrology for thin film thickness determination. It provides information on the electrodynamic basics and methods of measurement and evaluation. Hence, it is directed at all people who are involved in measuring film thickness by optical means, whether as manufacturer, in process and quality control, or in research and development. Hopefully, university lecturers and students of natural sciences and engineering will also find this book beneficial.

To write this book required reading and evaluating many monographs and a still larger number of publications on this subject. To my surprise, a lot of work has been done in ellipsometry, but spectral reflectance measurement for film thickness determination is sparsely described in literature although it is a well-established method. The total amount of published work is, however, too immense to consider them all in such a book. Therefore, I hope to have included the most relevant up to date, and apologize for all the contributions not considered here.

Last but not least, I want to greatly acknowledge all the people who supported me with data material, helpful information, and measurements, namely, Thomas Fries and Juergen Koglin from FRT GmbH, Anke Orth, Faiza Houta, and Bjoern Lewald from FRT GmbH, Alexei Maznev from the Massachussetts Institute of Technology, and Leif J. Hoglund from Semilab AMS. Many thanks to Gerhard Mueller for the picture of the glass on the cover, to my wife Ulrike for the picture of the oxidized Bi crystal on the cover, and finally to my family for their support and patience with me during writing this book.

Aldenhoven,
February 19, 2012

Michael Quinten

1
Introduction

Thin films of transparent or semitransparent materials play an important role in our life. A variety of colors in nature are caused by the interference of light reflected at thin transparent layers. Examples are the iridescent colors of a peacock feather, the impressive colors of lustrous butterfly wings, or simply the play of colors of thin oil films on water.

Much more demonstrative is, however, the use of thin films in technical applications. Films with maximum thickness of a few hundred nanometers are used as protective layers, hard coatings, antireflection coatings, adhesion and antiadhesion coatings, decorative coatings, transparent conductive layers, absorbing layers, in biosensors, and for tinted and annealed architectural glass. The combination of many thin films in multilayer stacks even lead to optical filters with sharp edges in reflection and transmission and almost 100% reflectivity in certain desired spectral ranges. The highest commercial impact these films have in microelectronics. Most microelectronic parts (processors, RAMs, flat screens, CDs/DVDs, hard disks, and some more) are manufactured with the help of thin-film technology. Thicker films of mainly transparent plastics are almost everywhere present as food packaging, wrapping, foils, membranes, lamination, and in display technology and solar cells, to give some examples.

Hence, it is our attempt to get as much information as possible on the properties and composition of surfaces and surface coatings. The two main classes of thin-film measurements are optical and stylus-based techniques. When measuring with a (mechanical) stylus, the thickness and roughness are obtained by monitoring the deflections of the fine-tipped stylus as it is dragged along the surface of the film. Stylus instruments, however, require a step in the film to measure thickness, even when using comparable optical sensors such as chromatic white light sensors. They are often the preferred method when measuring opaque films, such as metals.

Optical techniques determine the thin-film properties by measuring how the films interact with light. They can measure the thickness, roughness, and optical constants of a film. Optical techniques are usually the preferred method for measuring thin films because they are accurate, nondestructive, and require little or no sample preparation. The two most common optical measurement types are the *spectral reflectance measurement* and the *ellipsometry*. They form the main subject of this book. Besides, there exist other nondestructive methods for film thickness determination

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

with more or lower capabilities. Among them we find magnetoinductive and capacitive methods and the eddy current method, as well as the indirect measurement by a vibrating quartz or the measurement with ultrasound. Optical methods comprise light section, X-ray total reflection, photothermal deflection, and confocal chromatic measurement.

Spectral reflectance measurement or *reflectometry* uses the intensity of the light and measures the amount of light reflected from a thin film or a multilayer stack over a range of wavelengths, with the incident light normal (perpendicular) to the sample surface. Spectral reflectance can also measure the thickness, roughness, and optical constants of a broad range of thin films. However, if the film is very thin so that there is less than one reflectance oscillation, there is insufficient information available to determine the film parameters. Therefore, the number of film properties that may be determined decreases for very thin films. If on the other hand one attempts to solve for too many parameters, a unique solution cannot be found, but more than one possible combination of parameter values may result in a calculated reflectance that matches the measured reflectance. Depending upon the film material and the wavelength range of the measurement, the minimum single-film thickness that can be measured using spectral reflectance is in the 20–100 nm range. Additional determination of optical constants increases this minimum thickness. Nevertheless, as spectral reflectance is much simpler and less expensive than the second most common optical measurement – the ellipsometry – it is often used for quick and easy offline and in-line thickness determination in laboratories, production, and process control. To our knowledge, no comprehensive book on reflectometry as it is being practiced exists except for the one by Tompkins and McGahan [1], published in 1999. Therefore, one intention of this book is to bring the reflectometry closer to the practitioner.

In the late 1800s, Paul Drude [2] used the phase shift induced between the perpendicular components of polarized light to measure film thickness down to a few nanometers. This was the first study on film thickness measurement with a method that was later called ellipsometry. When the perpendicular components of polarized light are out of phase, the light is said to be elliptically polarized, for which this technique came to be called ellipsometry. Ellipsometry measures reflectance at nonnormal incidence (typically around 75° from normal) and is rather sensitive to very thin layers. The two different polarization measurements provide twice as much information for analysis. Variable-angle ellipsometry can be used to take reflectance measurements at many different incidence angles, thereby increasing the amount of information available for analysis. In 1977, Azzam and Bashara [3] authored the book *Ellipsometry and Polarized Light*, which has been the key source to be cited in most technical writing on the subject. Later on, several handbooks were published [4–6] that cover the theory of ellipsometry, instrumentation, applications, and emerging areas, in which experts in the field contributed to various aspects of ellipsometry. Fundamental principles and applications of spectroscopic ellipsometry are to be found in the recently published work of Fujiwara [7].

This book starts with Chapter 2 with an introduction to the basics of the propagation of light and other electromagnetic radiation in space and matter. Beyond the general properties of electromagnetic waves, we consider mainly the deviations

from the straightforward propagation by reflection, refraction, and diffraction since they are important for understanding the optical layer thickness determination and the functioning of the optical measuring devices. Interference of electromagnetic waves is a key effect not only for the diffraction of light but also for the optical layer thickness determination as it causes characteristic deviations in the reflectance spectrum of a thin film. From this characteristic interference pattern, all the film parameters are finally deduced.

Optical thickness determination is not only a question of electrodynamics but also a question of solid-state physics. The reason is that propagation in matter also means interaction of the electromagnetic wave with the matter. This interaction can be described with the complex dielectric function, while when discussing wave propagation in and through media the complex refractive index is appropriate. Both are connected via Maxwell's relation. In Chapter 2, we discuss physical models for the dielectric function and present empiric formulas for the refractive index.

The main topics of this book, the determination of the thickness of a layer in a layer stack from measurement of the spectral reflectance or transmittance, is treated in Chapters 3–5. The first step is taken in Chapter 3 with the modeling of the spectral reflectance R and transmittance T of a layer stack. Giving the thicknesses and complex refractive indices of all layers and substrates of the layer stack as input parameters, two common models – the propagating wave model and the r-t-ϕ model – can be used to calculate R and T of the stack (see Figure 1.1). The models are introduced in

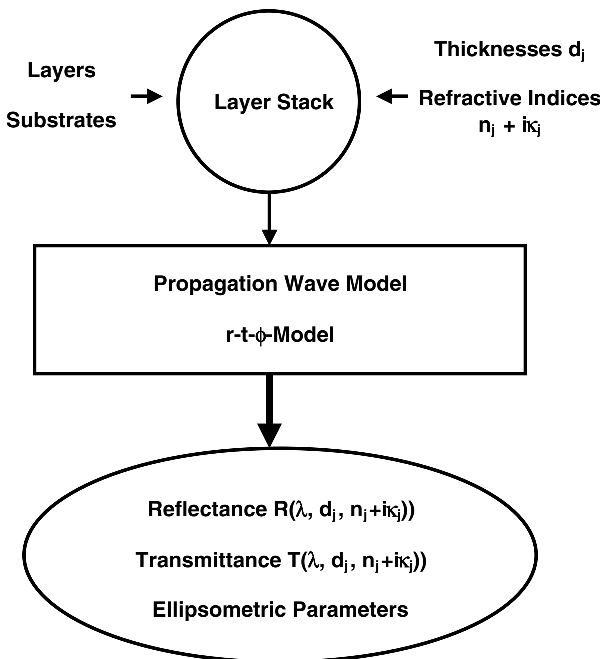


Figure 1.1 Modeling the reflectance R and transmittance T or ellipsometric data of a layer stack.

Chapter 3 and extensions on surface roughness and incoherent substrates are discussed. Absorption of light in the layer restricts the measurability of the thickness to a material-dependent maximum thickness.

In Chapter 4, we introduce the reflectometric and ellipsometric measurement and further optical methods, and discuss the optical components needed for the measurements. In all setups for optical thickness determination, the sample gets illuminated. Hence, light sources and their spectral distribution play a key role in the layer thickness determination, as well as the second key component, spectrometers. With the spectrometer, the reflected light modulated by the thickness interference gets spectrally resolved and analyzed.

Reflectometric and ellipsometric measurements do not measure the physical properties themselves but the optical response of the system caused by the physical properties. Hence, one needs to solve an inverse problem in order to find the value of actual physical properties of interest, such as thicknesses of the layers and optical properties of the materials. This inverse problem is solved numerically by finding the best fit between measured and calculated data, and physical properties are inferred from the model that gives the best fit (see Figure 1.2). To get reliable results, it is important to check the validity of the used model and to understand the sensitivity of the measured data to parameters of interest. In Chapter 5, we present and discuss numerical methods for determination of layer thickness and determination of optical constants of the layer material.

Chapter 6 is devoted to the apparent color of thin films. As the photographs on the cover of this book demonstrate, the interference in thin films leads to various colors depending on the thickness and refractive index of the film. However, not all colors

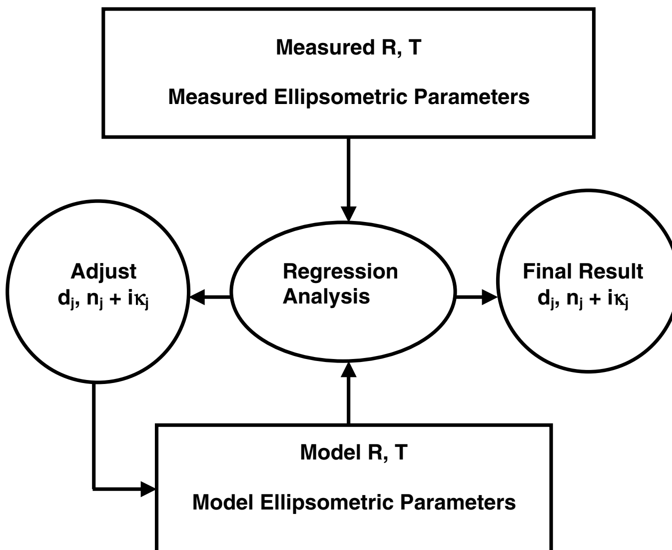


Figure 1.2 Fit procedure when analyzing measured R, T, or ellipsometric data for film thickness.

are available from one single layer. Instead, multilayer systems are needed to cover a certain color gamut.

Finally, in Chapter 7 we present several technical applications where film thickness measurement is important. They are accompanied by corresponding measuring results. The applications can be classified into the following:

- Applications with a single unsupported layer, for example, glass, sapphire, or semiconductor wafers, and transparent polymer films.
- Applications with one layer on a substrate, for example, protective layers (hard coats), broadband antireflection coatings, photoresists, and transparent conductive layers (TCF and TCO).
- Applications with two layers on a substrate. Examples of two layers on a substrate are photoresists on silica on a wafer, bonded wafers, and SOI wafers (SOI, silicon on insulator).
- Multilayer applications, for example, high reflective (HR) and antireflective (AR) coatings, beam splitter coatings, dielectric mirrors, optical filters, thin-film solar cells, and OLEDs (organic light emitting diodes).

We want to point out that all calculations of reflectance and transmittance spectra, the evaluation of thickness parameters and color, and the determination of optical constants were carried out with self-made software packages, MQLayer, MQNandK, and MQColor [8].

2
Propagation of Light and Other Electromagnetic Waves

This chapter introduces the basics of the propagation of light and other electromagnetic radiation in space and matter. Beyond the general properties of electromagnetic waves, we consider mainly the deviations from the straightforward propagation by reflection, refraction, and diffraction since they are important for understanding the optical layer thickness determination and the functioning of the optical measuring devices. Last but not least, propagation in matter also means interaction of the electromagnetic wave with the matter for what we also discuss the dielectric function and the refractive index in this chapter.

2.1
Properties of Electromagnetic Waves

When discussing the properties of electromagnetic waves, it seems appropriate to give first a definition of a wave. A wave generally is a process that is periodic in time and space. That means there exists a periodicity T in time after that the wave looks the same as at a certain time point t, and a periodicity $\mathbf{R}$ in space where the wave looks the same as at a certain point $\mathbf{r}$:

$$\mathbf{A}(\mathbf{r}+\mathbf{R}, t+T) = \mathbf{A}(\mathbf{r}, t). \quad (2.1)$$

Mathematically, $\mathbf{A}(\mathbf{r},t)$ fulfills the *wave equation* (in Cartesian coordinates):

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\mathbf{A}(\mathbf{r}, t) = 0, \quad (2.2)$$

with c being the propagation velocity. That means, in general, we search for a vector with its second derivative in time being proportional to its second derivative in space. The actual solution, however, is additionally determined by the boundary conditions of this differential equation.

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

When talking about electromagnetic waves, we often find, though it is not mandatory, that solutions of the wave equation are harmonic functions in time and space like

$$\mathbf{A}(\mathbf{r}, t) = \mathbf{A}_0 \cdot \exp(i(\mathbf{kr} - \omega\, t + \phi)), \tag{2.3}$$

with $k = |\mathbf{k}| = 2\pi/|\mathbf{R}|$, $\omega = 2\pi/T$, and ϕ an arbitrary constant. The length $|\mathbf{R}|$ is called *wavelength* λ and describes the distance between two successive identical phases of the wave in space, for example, the distance between two maxima or two minima. The propagation velocity corresponds to the vacuum velocity of light $c = 299\,792\,458$ m/s. k and ω fulfill the *dispersion relation*

$$k^2 = \frac{\omega^2}{c^2}. \tag{2.4}$$

Note that in (2.3) we used the notation with complex numbers, with $i = \sqrt{-1}$ being the complex unit. This notation will be used throughout the book. For an introduction to the numerics with complex numbers, we refer to Appendix A.

For electromagnetic waves, we have to consider an electric field $\mathbf{E}(\mathbf{r},t)$ and a magnetic field $\mathbf{H}(\mathbf{r},t)$ that must fulfill on the one hand the above conditions for $\mathbf{A}(\mathbf{r},t)$ and on the other hand Maxwell's equations:

$$\mathrm{div}(\mathbf{D}) = \varrho, \tag{2.5}$$

$$\mathrm{div}(\mathbf{B}) = 0, \tag{2.6}$$

$$\mathrm{curl}\, \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{2.7}$$

$$\mathrm{curl}\, \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}. \tag{2.8}$$

These equations, however, contain three more vectors, the current density $\mathbf{J}$, the electrical flux density or displacement $\mathbf{D}$, and the magnetic induction $\mathbf{B}$. To resolve Maxwell's equations for $\mathbf{E}$ and $\mathbf{H}$, it is therefore necessary to supplement them by relations that connect $\mathbf{J}$, $\mathbf{D}$, and $\mathbf{B}$ with $\mathbf{E}$ and $\mathbf{H}$.

When applying an electric field $\mathbf{E}$ on any material, the electric field forces unbound charge carriers to move, resulting in a current with density $\mathbf{J}$. The current is proportional to the applied field, with the *conductivity* σ being the proportionality constant:

$$\mathbf{J} = \sigma \mathbf{E}. \tag{2.9}$$

For bound charge carriers, the situation is different. They cannot move but are displaced. Inside the body of condensed matter the electric field usually displaces only the electrons while the ions are too inert as to follow the electric field. Thereby, each atom becomes an electric dipole with dipole moment $\mathbf{p}$. These dipole moments add up to a macroscopic net polarization $\mathbf{P}$ of the material that is related to the electric

field **E** by the general equation

$$\mathbf{P} = \varepsilon_0 \chi \mathbf{E}. \tag{2.10}$$

The factor χ is the macroscopic *susceptibility* of the matter. The polarization **P** contributes to the electrical flux density or displacement **D**:

$$\mathbf{D} = \varepsilon_0 \mathbf{E} + \mathbf{P} = \varepsilon_0 (1 + \chi) \mathbf{E} \tag{2.11}$$

defining the dielectric constant or permittivity ε as

$$\varepsilon = 1 + \chi. \tag{2.12}$$

Susceptibility χ and dielectric constant ε are the optical material functions. In the framework of Maxwell's theory, they enter the field relations as constants that are valid for the bulk material under consideration.

Applying a time harmonic electric field, the dipoles oscillate with the same frequency as the applied field. That means the center of gravity of the displaced electrons changes from one side to the other side, resulting in a displacement current $\partial \mathbf{D}/\partial t$.

Similar to the electric field **E**, the magnetic field **H** also causes two reactions of the material on the applied field, a magnetization **M**, and a current displacement $\partial \mathbf{D}/\partial t$. The latter is the result of the Lorentz force on bound and free electrons. The magnetization **M** contributes to the magnetic induction **B** in the material

$$\mathbf{B} = \mu_0 \mathbf{H} + \mu_0 \mathbf{M} = \mu \mu_0 \mathbf{H} \tag{2.13}$$

due to reorientation of permanent magnetic dipoles in the applied field. Looking at frequencies of electromagnetic waves, permanent magnetic dipoles are too inert as to follow a rapidly oscillating magnetic field. This holds true for frequencies ranging from the far infrared to infinity. Therefore, the *relative permeability* μ can be assumed to be 1 throughout the above frequency range, even for magnetic materials.

Finally, we point to the fact that when dealing with electromagnetic waves, static charges are absent, that is, $\varrho = 0$. If we further restrict only to time harmonic fields for the sake of simplification, the time dependence of the fields can be separated with the ansatz

$$\mathbf{E} = \mathbf{E}(\mathbf{r}) \cdot \exp(-i\omega t) \quad \text{and} \quad \mathbf{H} = \mathbf{H}(\mathbf{r}) \cdot \exp(-i\omega t) \tag{2.14}$$

and the corresponding Maxwell equations for the unknown parts **E(r)** and **H(r)** now read

$$\operatorname{div}(\mathbf{E}) = 0, \tag{2.15}$$

$$\operatorname{div}(\mathbf{H}) = 0, \tag{2.16}$$

$$\operatorname{curl}(\mathbf{E}) = i\omega\mu_0 \mathbf{H}, \tag{2.17}$$

$$\operatorname{curl}(\mathbf{H}) = (-i\varepsilon\varepsilon_0\omega + \sigma)\mathbf{E}. \tag{2.18}$$

Since we always assume homogeneity in time, it is usual, and done here and in the following, to omit the time dependence $\exp(-i\omega t)$ in the formulas, being assumed to be unaffected by matter (this is not always the case, we remind about, for example, inelastic scattering).

For solving the above Maxwell equations simultaneously for **E** and **H**, Equations 2.17 and 2.18 must be decoupled. This can be achieved by taking the curl of one equation and inserting the result into the other equation, resulting in the *vector wave equation* or *Helmholtz equation*

$$\mathrm{curl}(\mathrm{curl}(\mathbf{Z}))-k^2\mathbf{Z}=0, \tag{2.19}$$

where **Z** represents either **E** or **H**. The wavenumber k satisfies the *dispersion relation*

$$k^2=\frac{\omega^2}{c^2}\left(\varepsilon(\omega)+i\frac{\sigma(\omega)}{\omega\varepsilon_0}\right)=\frac{\omega^2}{c^2}\tilde{n}^2(\omega). \tag{2.20}$$

The quantity $\tilde{n}(\omega)$ is the complex refractive index with $\tilde{n}(\omega)=n(\omega)+i\kappa(\omega)$. The term in parentheses combines the permittivity ε (polarization) with the conductivity σ (absorption) to the complex *dielectric function* $\varepsilon(\omega)=\varepsilon_1(\omega)+i\varepsilon_2(\omega)$. For vacuum with $\tilde{n}(\omega)=1$, (2.20) corresponds to (2.4). The term *dielectric function* is used here instead of *dielectric constant* to take into account the dependence upon the frequency ω.

Equations 2.17 and 2.18 also define two separate solutions, a transverse magnetic solution (TM) and a transverse electric solution (TE).

This becomes obvious if we use the so-called Hertz vectors $\mathbf{\Pi_e}$ and $\mathbf{\Pi_h}$. It was shown by Hertz [9] and more generally by Righi [10] that it is possible under ordinary conditions to define an electromagnetic field in terms of a properly chosen single vector function $\mathbf{\Pi}$.

Assuming that the magnetic field can be derived from a Hertz vector $\mathbf{\Pi_e}=(0, 0, \Pi_e)^T\cdot\exp(-i\omega t)$ by

$$\mathbf{H}=(-i\omega\varepsilon_0\varepsilon+\sigma)\cdot\mathrm{curl}(\mathbf{\Pi_e}). \tag{2.21}$$

Then, the corresponding electric field **E** follows from (2.18)

$$\mathbf{E}=\mathrm{curl}(\mathrm{curl}(\mathbf{\Pi_e})). \tag{2.22}$$

It is simple to show that the z-component H_z is zero in Cartesian coordinates, that is, its axial or longitudinal component is absent. Thus, we have derived from a scalar function Π_e an electromagnetic field characterized by the absence of the longitudinal component of the magnetic vector. This field is called electric or more properly transverse magnetic (TM) mode.

Assuming now that the electric field can be derived, Hertz vector $\mathbf{\Pi_h}=(0, 0, \Pi_h)^T\cdot\exp(-i\omega t)$ by

$$\mathbf{E}=i\omega\mu_0\cdot\mathrm{curl}(\mathbf{\Pi_h}). \tag{2.23}$$

Then, the magnetic field follows from (2.17) as

$$\mathbf{H}=\mathrm{curl}(\mathrm{curl}(\mathbf{\Pi_h})). \tag{2.24}$$

It is simple to show that in this case the longitudinal component E_z of the electric vector is zero in Cartesian coordinates, defining the magnetic or more properly transverse electric (TE) mode.

In curvilinear coordinate systems, complete solutions of the vector wave equation in a form directly applicable to the solution of boundary value problems are known only for the separable systems of cylindrical, spheroidal, and spherical coordinates.

In curvilinear coordinate systems, it is also common practice to introduce *vector harmonics* **M** and **N** instead of using Hertz vectors. The vector harmonics are defined by

$$\mathbf{M} = \mathrm{curl}(\Phi\mathbf{a}), \tag{2.25}$$

$$\mathbf{N} = k^{-1} \cdot \mathrm{curl}(\mathbf{M}), \tag{2.26}$$

where the quantity $\mathbf{\Phi}$ is a scalar potential satisfying the *scalar wave equation*

$$\left(\nabla^2 + \mathrm{k}^2\right) \cdot \Phi = 0 \tag{2.27}$$

and $\mathbf{a}$ is an arbitrary constant vector. The scalar potential $\mathbf{\Phi}$ and the vector $\mathbf{a}$ are connected with the Hertz vectors via

$$k \cdot \Pi_\mathbf{e} = \Phi \cdot \mathbf{a}, \tag{2.28}$$

$$\mathrm{i}\omega\mu_0\Pi_\mathbf{h} = \Phi \cdot \mathbf{a}. \tag{2.29}$$

If both the electric and the magnetic field vector do not have a longitudinal component, the solution of Maxwell's equations and the Helmholtz equation is called TEM wave. Generally, **E** and **H** are given as a linear superposition of TM and TE modes. Notice that in several electrodynamic problems like wave propagation in waveguides it is not possible to obtain pure TE or TM solutions, except for the ground mode.

It is worth to mention that a third, irrotational solution of Maxwell's equations and the vector wave equation exists as

$$\mathbf{L} = -\mathrm{grad}(\phi_\mathrm{L}) \tag{2.30}$$

that satifies the Laplace equation

$$\mathrm{div}(\mathbf{L}) = -\nabla^2\phi_\mathrm{L} = 0. \tag{2.31}$$

L describes longitudinal waves with wavenumber k_L. These waves cannot be described with the dielectric function defined in (2.20), but are described with the dielectric function $\varepsilon_\mathrm{L}(\omega,k_\mathrm{L})$ that satisfies

$$\varepsilon_\mathrm{L}(\omega, k_\mathrm{L}) = 0. \tag{2.32}$$

If longitudinal waves can propagate in the material in a certain frequency range, the function **L** describes longitudinal electron density fluctuations in the material, the *longitudinal bulk plasmon modes*. At these frequencies, the interior TM fields are coupled to the bulk plasmons, while the TE fields are not affected.

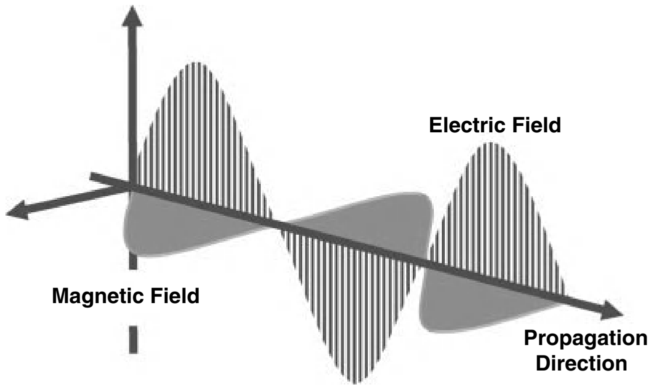


Figure 2.1 Sketch of a time harmonic transverse electromagnetic wave.

In the following, we consider a plane monochromatic TEM wave. It is described by the equation

$$\mathbf{Z}(\mathbf{r}, t) = \mathbf{Z_0} \cdot \exp(i(\mathbf{kr} - \omega t + \phi)), \tag{2.33}$$

with $\mathbf{Z}$ representing either $\mathbf{E}$ or $\mathbf{H}$.

Applying Maxwell's equations (2.15) and (2.16) to this wave leads to

$$\mathbf{k} \cdot \mathbf{E} = 0 \text{ and } \mathbf{k} \cdot \mathbf{B} = 0, \tag{2.34}$$

meaning that the electric field and the magnetic field are perpendicular to the propagation wave vector $\mathbf{k}$. This is also illustrated in Figure 2.1. Equation 2.34 describes *transverse waves*, a characteristic property of electromagnetic waves.

Considering the phase of this plane electromagnetic harmonic wave, we can recognize that the surfaces of constant phase and constant amplitude coincide and are planes normal to the direction of propagation. The surfaces of constant phase satisfy the equation

$$\mathbf{k} \cdot \mathbf{r}\text{-}\omega t = \text{constant}, \tag{2.35}$$

which is the representation of planes perpendicular to $\mathbf{k}$. According to the dispersion equation (2.20), the plane wave propagates with the velocity of light c when propagating in vacuum, and with reduced velocity $c/\hat{n}(\omega)$ in a medium with complex refractive index $\hat{n}(\omega) = n(\omega) + i \cdot \kappa(\omega)$. If the refractive index is complex, then also $\mathbf{k}$ is a complex-valued vector $\mathbf{k} = \mathbf{k_1} + i\mathbf{k_2}$ and the wave becomes inhomogeneous because its amplitude decreases with $\exp(-\mathbf{k_2 r})$. If $\mathbf{k_1}$ and $\mathbf{k_2}$ are collinear, the wave evanesces *along* the propagation direction. The surfaces of constant phase and constant amplitude still coincide in this special case. If in general $\mathbf{k_1}$ and $\mathbf{k_2}$ include an angle Ω, the surfaces of constant phase and constant amplitude never coincide but are oblique.

The magnitude of the electric and magnetic field of an electromagnetic wave that propagates through an absorbing medium along, for example, the x-direction decreases by the factor $\exp(-\beta(\omega)\cdot x)$ with

$$\beta(\omega) = \frac{\omega}{c}\kappa(\omega). \tag{2.36}$$

As a consequence, the intensity that is proportional to the square of the magnitude decreases by the factor $\exp(-2\beta(\omega)\cdot x)$. The difference $1 - \exp(-2\beta(\omega)\cdot x)$ is lost in the medium by exciting vibrations of molecules or lattice vibrations (phonons), which is called *absorption* with the absorption constant $\beta(\omega)$.

Electromagnetic waves exhibit another property that is characteristic of transverse waves: *polarization*. As we know from (2.34), the electromagnetic fields $\mathbf{E}$ and $\mathbf{H}$ of a TEM wave are perpendicular to the wave vector $\mathbf{k}$. They oscillate in a fixed plane spanned either by $(\mathbf{E}, \mathbf{k})$ or $(\mathbf{H}, \mathbf{k})$. We denote the electromagnetic wave as *p-polarized* if the electric field oscillates in the plane of incidence, and as *s-polarized* if the electric field oscillates perpendicular to the plane of incidence. In general, we can write the electric field as the sum of a p- and an s-polarized component $\mathbf{E_p}$ and $\mathbf{E_s}$ that are orthogonal and lie in the plane $\mathbf{k}\cdot\mathbf{r} - \omega t = \text{constant}$ on which $\mathbf{k}$ is perpendicular. Moreover, $\mathbf{E_p}$ and $\mathbf{E_s}$ may have different magnitudes E_{p0} and E_{s0} and different constant phase shifts ϕ_p and ϕ_s, that is,

$$\mathbf{E_p} = \mathbf{E_{p0}} \exp\Big(i(\mathbf{k}\cdot\mathbf{r} - \omega t + \phi_p)\Big), \tag{2.37}$$

$$\mathbf{E_s} = \mathbf{E_{s0}} \exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t + \phi_s)). \tag{2.38}$$

If we investigate now the locus of $|\mathbf{E}| = \sqrt{E_p^2 + E_s^2}$ in the plane $\mathbf{k}\cdot\mathbf{r} - \omega t = \text{constant}$, we find after some mathematics that E_p and E_s must satisfy the relation

$$\left(\frac{E_p}{E_{p0}}\right)^2 - 2\frac{E_p}{E_{p0}}\frac{E_s}{E_{s0}}\cos\delta + \left(\frac{E_s}{E_{s0}}\right)^2 = \sin^2\delta, \tag{2.39}$$

with δ being the difference in the constant phase shifts ϕ_p and ϕ_s, $\delta = \phi_p - \phi_s$. In general, (2.39) describes an ellipse in this plane and $\mathbf{E}$ rotates along this ellipse. Depending on δ we can distinguish the following three cases:

- Linear polarization with $\delta = 0$.
- Circular polarization with $\delta = \pi/2$.
- Elliptical polarization with δ being arbitrary, but not 0 and not $\pi/2$.

For linear polarization, the field vector does not rotate. It is customary to describe the rotation in the case of circular polarization as right handed if the vector rotates clockwise when viewed in direction opposite to the propagation direction. Accordingly, it is called left handed if the vector rotates counterclockwise. The three cases of polarization are illustrated in Figure 2.2.

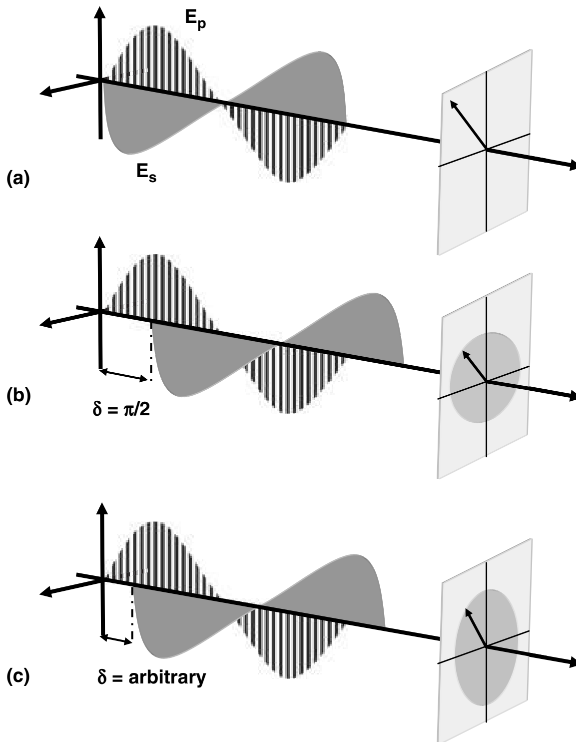


Figure 2.2 Illustration of (a) linear polarization, (b) circular polarization, and (c) elliptical polarization of the electromagnetic wave.

2.2 Huygens–Fresnel Principle

It was Christian Huygens (1629–1695) who first formulated how a wave propagates in a medium (till this time, it was not proven that electromagnetic waves are transverse waves). With his proposal he arrived at an explanation of the reflection and refraction at an interface between two media and at an explanation of dichroism.

The main statements of Huygens are as follows:

- Each point of an existing wavefront of a plane wave at the site $\mathbf{r}$ and at time point t_0 is origin of a new secondary circular wave with the same wavelength λ, the same time period T, and the same polarization as the original plane wave.
- The envelope of these secondary waves determines the new wavefront and the stimulus in a point P.

In Figure 2.3, we illustrate the principle for a plane wave. The black lines indicate the planes of constant phase where the magnitude of the wave is maximum. The dashed lines are the secondary circular waves. As it becomes obvious from the figure, the envelope of the secondary waves forms a new plane of constant phase not only at

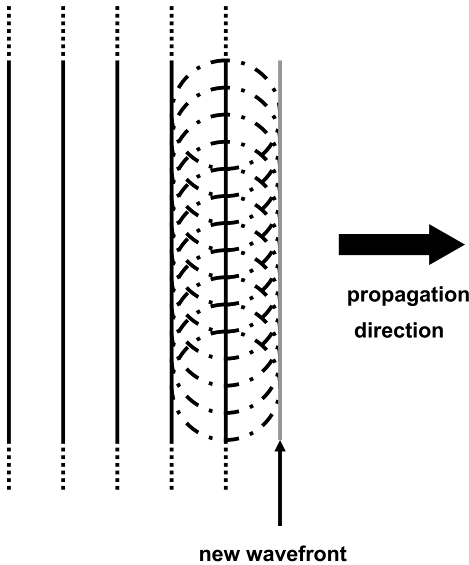


Figure 2.3 Huygens' principle applied on a plane wave.

$\mathbf{r} + \lambda\mathbf{e}$ but also at $\mathbf{r} - \lambda\mathbf{e}$, with $\mathbf{e}$ being the unit vector in propagation direction. The latter shows a gap in Huygens' explanation because it cannot be excluded that the incident wavefront gets weakened by the secondary waves.

Auguste-Jean Fresnel (1788–1827) closed this gap when combining the ideas of Huygens with Young's principle of interference by linear superposition (Thomas Young, 1773–1829). From this, it follows that the envelope of the secondary waves results in a new wavefront only in propagation direction. One important result of this combination is the following statement:

> The propagation direction of the reconstructed wave is perpendicular to the wavefront.

With the *Huygens–Fresnel principle* can be described not only the straight propagation of waves but also all deviations from it that are known as diffraction. We will apply it in the following to give a description of reflection and refraction (Section 2.4) and a description of diffraction (Section 2.5). As interference plays an important role in the discussion of diffraction, we will introduce it first in Section 2.3.

2.3
Interference of Electromagnetic Waves

Thomas Young established in 1801 the principle of interference of waves as linear superposition of waves with different phases. We introduce this principle for the interference of two waves.

Let $\mathbf{A}(\mathbf{r}, t) = \mathbf{A_0}\exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t))$ be one wave and $\mathbf{B}(\mathbf{r}, t) = \mathbf{B_0}\exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t + \phi))$ the second wave with different magnitude having a phase difference ϕ. The linear superposition of $\mathbf{A}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$ means that a new wave with $\mathbf{C}(\mathbf{r}, t)$ results from the sum of $\mathbf{A}(\mathbf{r}, t)$ and $\mathbf{B}(\mathbf{r}, t)$:

$$\begin{aligned}\mathbf{C}(\mathbf{r},t) &= \mathbf{A}(\mathbf{r},t) + \mathbf{B}(\mathbf{r},t)\\ &= (\mathbf{A}_0 - \mathbf{B}_0)\exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t)) + \mathbf{B}_0\cdot\exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t))\cdot(1+\exp(i\phi))\\ &= \exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t))\cdot(\mathbf{A}_0 - \mathbf{B}_0 + 2\mathbf{B}_0\exp(i\phi/2)\cdot\cos(\phi/2)).\end{aligned} \tag{2.40}$$

The magnitude $C = |\mathbf{C}(\mathbf{r},t)|$ of the resulting wave strongly depends on the value of the phase difference ϕ: if $\phi = (2m + 1)\pi$, with m being an integer number, the magnitude is minimum and even completely vanishes for $\mathbf{A_0} = \mathbf{B_0}$. On the other hand, it becomes maximum for $\phi = 2m\pi$. For identical magnitudes of the two superposed waves, we therefore can expect complete extinction, called *destructive interference*, and a doubling of the magnitude, called *constructive interference*. For all other phase differences, ϕ, an intermediate state occurs.

For illustration, we assume two one-dimensional waves with same magnitude at a fixed time point t_0. Then, the phase difference is defined by a difference δ in the path along the x-axis and is $\phi = (2\pi/\lambda)\delta$. $C(x, t_0)$ follows as

$$\begin{aligned}C(x,t_0) &= A\sin\left(\frac{2\pi}{\lambda}x + \omega t_0\right) + A\sin\left(\frac{2\pi}{\lambda}(x+\delta) + \omega t_0\right)\\ &= 2A\cos\left(\frac{\pi\delta}{\lambda}\right)\cdot\sin\left(\frac{2\pi}{\lambda}\left(x + \frac{\delta}{2}\right) + \omega t_0\right).\end{aligned} \tag{2.41}$$

As can be recognized from Figure 2.4a, the superposition of the two waves result in a new wave with magnitude $A < C < 2A$ that is phase shifted compared to the two primary waves by $\pi\delta/\lambda$ if the shift δ is arbitrary. For the two special cases $\delta = (2m + 1)\lambda/2$ and $\delta = 2m\lambda$, the destructive interference, Figure 2.4b, with $C = 0$, and the constructive interference, Figure 2.4c, with $C = 2A$, are obtained.

2.4 Reflection and Refraction

In this section, we introduce a plane interface as boundary between two different materials and consider the resulting wave propagation in each of the two materials. The interface strongly disturbs the free propagation of the wave and we get new waves by reflection and refraction.

Consider a plane interface between two media 1 and 2 with refractive indices n_1 and n_2. For the moment, we assume that n_2 is real valued, that is, medium 2 is nonabsorbing. Medium 1 is always assumed to be nonabsorbing in the following. The discussion of absorbing medium 2 follows after the derivation of reflection and refraction. A plane wave in medium 1 with wave vector $\mathbf{k_{inc}} = k_{\mathrm{inc}}\,(\sin\alpha, -\cos\alpha, 0)^{\mathrm{T}}$

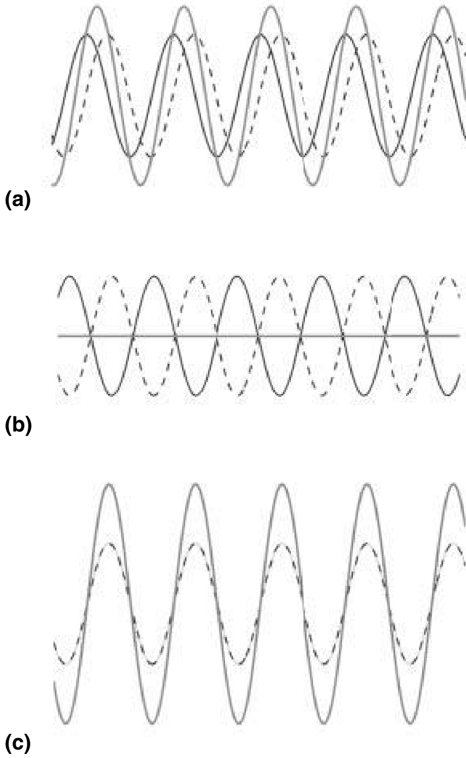


Figure 2.4 Interference of two one-dimensional waves with the same magnitude and phase difference ϕ: (a) ϕ arbitrary, (b) $\phi = (2m + 1)\pi$ or $\delta = (2m + 1)\lambda/2$, and (c) $\phi = 2m\pi$ or $\delta = 2m\lambda$.

hits this interface so that its propagation direction encloses an angle α with the normal to the interface. Applying now Huygens' principle, we can construct the wavefronts in medium 1 and medium 2. This is illustrated in Figure 2.5.

When constructing the wavefronts in medium 1, we find as a result a second wave with wave vector $\mathbf{k_{ref}} = k_{inc}\,(\sin\alpha, \cos\alpha, 0)^T$. It is *reflected* at the interface between the two media. It also includes the angle α to the normal on the interface. From this we can deduce the *reflection law*:

$$\text{Angle of reflection } \alpha_R = \text{angle of incidence } \alpha, \tag{2.42}$$

which is well known since the ancient times (Euclid, 300 B.C.).

When constructing the wavefronts in medium 2, we have to take into account that the wavelength changes from λ/n_1 to λ/n_2. If $n_2 > n_1$, the resulting wave with $\mathbf{k_{refr}}$ in medium 2 encloses an angle β with the normal that is smaller than α because the distance of planes of constant phase – the wavelength – becomes smaller than in medium 1. Vice versa, for $n_2 < n_1$, the angle β becomes larger than α because now the wavelength becomes larger than in medium 1. The relation between β and α is not at

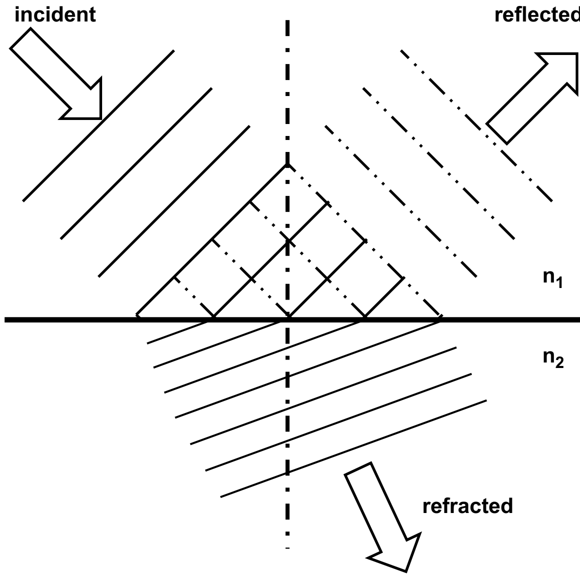


Figure 2.5 Reflection and refraction on the interface between medium 1 with refractive index n_1 and medium 2 with refractive index n_2, with $n_1 < n_2$.

all as obvious as the reflection law. Already, Ptolemy (approx. 90–168 A.D.) studied the relation between α and β, but did not arrive at a formula. It was the Dutch Willebrord van Roijen Snell who was first to describe mathematically this relation in 1618, which is well known as *Snell's law of refraction,*

$$n_1(\lambda) \cdot \sin(\alpha) = n_2(\lambda) \cdot \sin(\beta). \tag{2.43}$$

Unfortunately, Snell died before he could publish his results, but he reported them to his scientific colleagues, and also to the French René Descartes who published it later in 1637 in his *Discours de la Méthode Pour Bien Conduire sa Raison et Chercher la Verité dans les Sciences* (Discourse on the Method of Rightly Conducting One's Reason and of Seeking Truth in the Sciences).

Both the reflection law and Snell's law can also be derived from the boundary conditions for the electromagnetic fields at a surface S

$$(\mathbf{E}_{\mathbf{inc}} + \mathbf{E}_{\mathbf{refl}}) \times \mathbf{n}|_{\text{surface}} = \mathbf{E}_{\mathbf{refr}} \times \mathbf{n}|_{\text{surface S}}, \tag{2.44}$$

$$(\mathbf{H}_{\mathbf{inc}} + \mathbf{H}_{\mathbf{refl}}) \times \mathbf{n}|_{\text{surface}} = \mathbf{H}_{\mathbf{refr}} \times \mathbf{n}|_{\text{surface S}}. \tag{2.45}$$

They relate the electromagnetic fields in medium 1 ($\mathbf{E}_{\mathbf{inc}}$, $\mathbf{H}_{\mathbf{inc}}$, $\mathbf{E}_{\mathbf{refl}}$, $\mathbf{H}_{\mathbf{refl}}$) to the electromagnetic fields ($\mathbf{E}_{\mathbf{refr}}$, $\mathbf{H}_{\mathbf{refr}}$) in medium 2. $\mathbf{n}$ is the vector normal to the surface S. The reflection law then corresponds to the statement that the components of $\mathbf{k}_{\mathbf{inc}}$ and $\mathbf{k}_{\mathbf{ref}}$ parallel to the plane surface are identical, and Snell's law corresponds to the statement that the components of $\mathbf{k}_{\mathbf{inc}}$ and $\mathbf{k}_{\mathbf{refr}}$ parallel to the plane surface are identical.

Table 2.1 Index of refraction n of various transparent materials.

Material	Refractive index n
Vacuum	1.0
Air	1.00029
Water	1.334
Benzene C_6H_6	1.501
Carbon tetrachloride CCl_4	1.46
Carbon disulfide CS_2	1.628
Various crown glasses	$1.5 \leq n_M \leq 1.55$
Various flint glasses	$1.6 \leq n_M \leq 1.9$
Fused silica	1.46
Polycarbonate	1.59
CaF_2	1.43
MgF_2	1.38
Si_3N_4	2.03
Diamond	2.417
MgO	1.74
Al_2O_3	1.77
TiO_2	2.03
CeO_2	2.47
ZnO	2.0

All data are at wavelength $\lambda = 589$ nm.

For a fixed refractive index n_1 of medium 1, the larger is the refraction, the larger the refractive index n_2 of medium 2 becomes. In Table 2.1, we give exemplaric values for the refractive index of various transparent materials at wavelength 589 nm.

Up to this point, we only considered the propagation direction of the waves. Now, we want to consider the relation of magnitudes of the reflected and refracted electromagnetic fields with the magnitude of the incident wave. These relations – called *Fresnel equations* – also follow from applying the boundary conditions. We pass on the derivation but give here only the results as *Fresnel coefficients* in the following four equations. Note that the results depend upon the polarization of the incident wave and are different for s-polarization and p-polarization:

$$r_s = \frac{E_{s,ref}}{E_{s,inc}} = \frac{n_1 \cos(\alpha) - n_2 \cos(\beta)}{n_1 \cos(\alpha) + n_2 \cos(\beta)}, \tag{2.46}$$

$$t_s = \frac{E_{s,refr}}{E_{s,inc}} = \frac{2 n_1 \cos(\alpha)}{n_1 \cos(\alpha) + n_2 \cos(\beta)}, \tag{2.47}$$

$$r_p = \frac{E_{p,ref}}{E_{p,inc}} = -\frac{n_1 \cos(\beta) - n_2 \cos(\alpha)}{n_1 \cos(\beta) + n_2 \cos(\alpha)}, \tag{2.48}$$

$$t_p = \frac{E_{p,refr}}{E_{p,inc}} = \frac{2\, n_1 \cos(\alpha)}{n_1 \cos(\beta) + n_2 \cos(\alpha)}. \tag{2.49}$$

The corresponding ratios for the intensity R_s, T_s, R_p, and T_p are obtained by considering the incident, reflected, and refracted energy flux density. It follows that

$$R_{s,p} = r_{s,p} \cdot r^*_{s,p}, \tag{2.50}$$

$$T_{s,p} = \frac{n_2 \cos(\beta)}{n_1 \cos(\alpha)} \cdot t_{s,p} \cdot t^*_{s,p}, \tag{2.51}$$

where the asterisk denotes the complex conjugate. As long as the refractive index n_2 is real valued, the multiplication of $r_{s,p}$ or $t_{s,p}$ with its conjugate complex simply means to build the square of $r_{s,p}$ or $t_{s,p}$, respectively.

It is not necessary but often helpful to replace $\cos(\beta)$ in the above Fresnel equations by

$$\cos(\beta) = \sqrt{\frac{n_2^2 - n_1^2 \sin^2(\alpha)}{n_2^2}} \tag{2.52}$$

to reduce the formula on the always real angle of incidence α.
What happens with the reflection at the surface S and the transmission through the surface including Snell's law if the medium 2 becomes absorbing, that is, if the refractive index of medium 2 becomes complex valued $\tilde{n}_2(\lambda) = n_2(\lambda) + i\kappa_2(\lambda)$?
The first obvious consequence is that $\sin(\beta)$ in (2.43) and hence the angle of refraction β must also be complex. It follows that

$$\mathrm{Re}(\sin(\beta)) = \frac{n_2}{n_2^2 + \kappa_2^2} n_1 \sin(\alpha), \tag{2.53}$$

$$\mathrm{Im}(\sin(\beta)) = -\frac{\kappa_2}{n_2^2 + \kappa_2^2} n_1 \sin(\alpha). \tag{2.54}$$

In that case also, the planes of constant phase do not further coincide with the planes of constant amplitude. We will see this when we discuss the reflection and transmission of a single layer in Section 3.1.

The next consequence is that the Fresnel coefficients r_s, t_s, r_p, and t_p become complex numbers. Then, the intensity coefficients R_s, T_s, R_p, and T_p must be calculated from the product of a Fresnel coefficient with its conjugated complex value, as in (2.50) and (2.51).

At the end of this section, we consider the reverse situation: the wave is coming from the medium with higher refractive index n_2 (optically thick) and is going to the medium with lower refractive index n_1 (optically thin). From Snell's law (2.43), it follows that for an angle of incidence $\alpha_C = \sin^{-1}(n_1/n_2)$ – this is called *critical angle of total reflection* – the maximum angle of refraction of $\beta = 90°$ is reached. The corresponding refracted beam travels along with the interface in the optically thinner medium. For $\alpha > \alpha_C$ the incident light gets totally reflected at the interface. However, in the optically thinner medium, there still exists a wave propagating along the interface, but with the amplitude of the electromagnetic fields decreasing exponentially with $\exp(-\gamma z)$ in the direction perpendicular to the interface as illustrated in

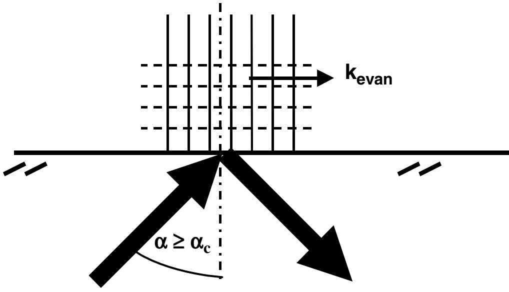


Figure 2.6 Evanescent wave from total internal reflection at the interface between an optically thicker medium and an optically thinner medium.

Figure 2.6. That means the planes of constant phase and the planes of constant amplitude are perpendicular to each other.

This kind of waves is called *evanescent wave*. The attenuation constant γ depends upon the angle of incidence α and the refractive indices n_1 and n_2:

$$\gamma = \frac{2\pi}{\lambda}\left(n_2^2 \sin^2 \alpha - n_1^2\right)^{1/2}. \tag{2.55}$$

For $\alpha = \alpha_C$, it is $\gamma = 0$, and for $0 \leq \alpha < \alpha_C$, it becomes purely imaginary, so that the wave is not attenuated in z-direction but corresponds to a radiating wave, as expected from Snell's law.

These evanescent waves from total internal reflection exhibit a peculiarity with respect to the polarization. Due to the complex phase shift when total reflection occurs, the electric field of the p-polarized evanescent wave rotates in time. This means that the p-polarized evanescent wave has almost always a component in propagation direction of the evanescent wave. This part corresponds to a *longitudinal* electromagnetic wave.

2.5 Diffraction

Unlike reflection and refraction, diffraction not only results in a change in the propagation direction of the wave but is also accompanied by the linear superposition of partial waves from different sites that have the same propagation direction but a different phase. That means we must consider two effects: the redirection of light, described by the Huygens–Fresnel principle, and the interference of waves according to Young.

For derivation of the diffraction by a grating, we first consider a single slit in an opaque plate. The wavefront in the plane of the slit is origin of new circular waves with the same wavelength because the matter in front of the slit and behind the slit is the same. However, as the slit has a finite size, a finite number of circular waves are also generated. This is shown in Figure 2.7. The finiteness of the slit causes the superposition of the circular waves to result in plane wavefront only in the middle

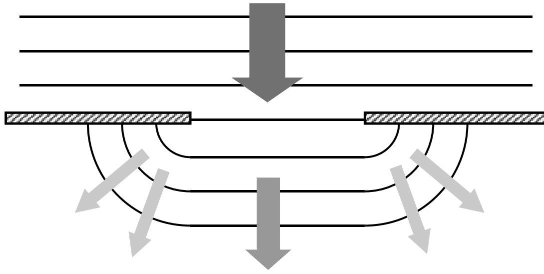


Figure 2.7 Diffraction by a slit – Huygens–Fresnel principle.

but is curved at the edges. Accordingly, the propagation vector of the new reconstructed wave has not only components in the propagation direction of the incident wave but also spreads into the complete half space behind the slit.

The next step is to consider a light beam propagating in a certain direction enclosing the angle β with respect to the normal on the slit. We can divide this beam into either an even ($2p$) or an odd ($2p + 1$) number of partial waves going in the same direction but from different points along the slit as shown in Figure 2.8.

Two neighboring partial waves then have a phase difference $\phi = (2\pi b/2p\lambda)\sin\beta$ for an even number of partial waves and $\phi = (2\pi b/(2p+1)\lambda)\sin\beta$ for an odd number of partial waves.

For $2p$ partial waves, we always find pair-wise two beams that interfere destructively. That means in the directions where

$$\sin\beta_m = m\frac{\lambda}{b}, \qquad m = \pm 1, \pm 2, \pm 3, \ldots \tag{2.56}$$

we will find destructive interference and hence no intensity. The number m is, in principle, the order of the appearing minimum.

Vice versa, for an odd number of partial waves, we find pair-wise beams that interfere destructively, but there remains finally a bundle of light beams that cannot interfere destructively because there are no other beams available for this interference. Hence, maximum intensity should be obtained in the directions with

$$\sin\beta_m = \frac{(2m+1)}{2}\frac{\lambda}{b}, \qquad m = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{2.57}$$

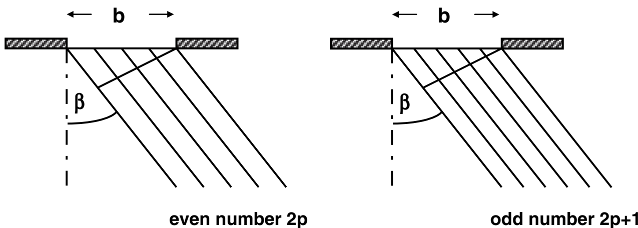


Figure 2.8 Diffraction by a slit – interference of an even ($2p$) or an odd ($2p + 1$) number of partial waves.

In a more elaborated derivation, we calculate the radiant flux through the slit into the half space behind the slit. It is proportional to the square of the electric field distribution in the slit

$$J_{\text{slit}}(\beta) \propto \left| \int_{-b/2}^{b/2} E_0 \cdot \exp(ikx)\mathrm{d}x \right|^2 \qquad (2.58)$$

with $k = (2\pi/\lambda)\sin\beta$. The result is

$$J_{\text{slit}}(\beta) \propto b^2 \frac{\sin^2((\pi b/\lambda)\sin\beta)}{((\pi b/\lambda)\sin\beta)^2}. \qquad (2.59)$$

The energy of the wave in the slit is distributed in the half space behind the slit and can be described by a squared sinc function $\text{sinc}^2(\sin(\beta))$.

Only the minima of this always positive definite function are obvious. They are obtained in the directions where $\sin(\beta)$ fulfills (2.56). For the maxima, we find the condition

$$\tan\left(\frac{\pi \mathrm{b}}{\lambda}\sin\beta\right) = \frac{\pi \mathrm{b}}{\lambda}\sin\beta. \qquad (2.60)$$

This condition is approximately fulfilled if $\sin(\beta)$ satisfies (2.57). The higher the order n is, the smaller the deviation is. For example, for $m = 1$ the maximum is obtained at 1.43π (instead of 1.5π), for $m = 2$ it is obtained at 2.46π (instead of 2.5π), and for $m = 3$ it is obtained at 3.47π (instead of 3.5π). A special case is the maximum for $\sin(\beta) = 0$. It follows directly from (2.60), but for calculation of the intensity of this maximum, we must apply L'Hospital's rule on the sinc function in (2.61).

Up to now, we considered the slit only in one dimension, having a width b. In fact, the slit has also a height h and we can find also diffraction in the vertical direction if the height h is sufficiently small. Figure 2.9a shows an example for the diffraction by a rectangular slit of width b and height h and Figure 2.9b shows the diffraction by a quadratic slit. The pictures are taken from Ref. [11].

We turn now to the grating that consists of a thin plane plate with periodically arranged slits (see Figure 2.10) illuminated by a plane wave. The periodicity of the slits is the *grating period*, g. The reciprocal of the grating period is called *groove density*, $L = 1/g$.

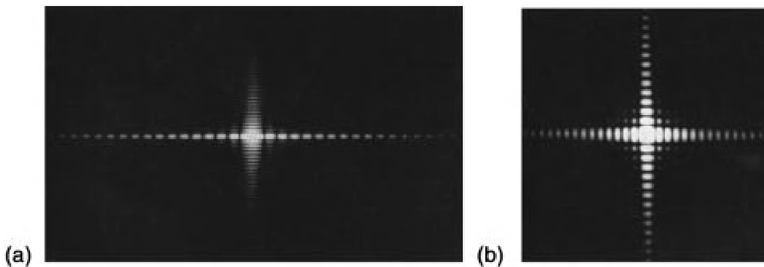


Figure 2.9 (a) Diffraction by a rectangular slit of width b and height h. (b) Diffraction by a quadratic slit. Reproduced from Ref. [11], with kind permission of Springer Science + Business Media.

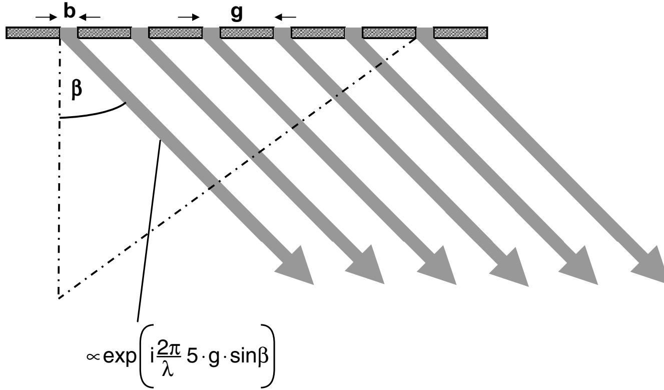


Figure 2.10 Diffraction by a grating – interference of waves from periodically arranged slits of width b and periodicity g.

From each slit a wave propagates in the direction given by the diffraction angle β. From slit to slit they have the phase difference $\varphi = (2\pi/\lambda)g \cdot \sin\beta$ and a phase factor $\exp(i\varphi)$. These waves superpose in the far field (far from the grating surface). Then, the radiant flux through the slits of the grating into the half space behind the grating is

$$J_{\text{grating}}(\beta) \propto J_{\text{slit}}(\beta) \cdot \left| \sum_{n=0}^{N-1} (\exp(i\varphi))^n \right|^2 \propto \frac{\sin^2((\pi b/\lambda)\sin\beta)}{((\pi b/\lambda)\sin\beta)^2} \cdot \frac{\sin^2((\pi N g/\lambda)\sin\beta)}{\sin^2((\pi g/\lambda)\sin\beta)}. \tag{2.61}$$

The intensity diffracted by the grating into direction of the angle β consists of two factors: the first $J_{\text{slit}}(\beta)$ describes the diffraction by a single slit and the second results from the multiple interference of the waves coming from all N slits of the grating. This second factor determines the diffracted light in a distinct manner, as it introduces additional minima and maxima. If the single slit diffraction leads to zero intensity in direction of the angle β, this minimum is kept in the intensity diffracted by the grating. However, between two minima of $J_{\text{slit}}(\beta)$ the second factor – the *grating function*, GF – can be zero with maxima of GF in between. Another effect of the multiple interference is that the intensity of these new maxima from the GF strongly increases with the number of illuminated slits N proportional to N^2 so that these maxima dominate the intensity distribution in the half space behind the grating.

The maxima of GF are found after some mathematics. The first derivation of GF yields the conditions for extrema, and the second derivative finally shows which of the extrema are the maxima. The maxima are given by

$$\sin\beta = m\frac{\lambda}{g}, \qquad m = 0, \pm 1, \pm 2, \pm 3, \ldots, \tag{2.62}$$

where m is the order of diffraction, and by the solution of the equation

$$N \cdot \cot\left(\frac{\pi N g}{\lambda}\sin\beta\right) - \cot\left(\frac{\pi g}{\lambda}\sin\beta\right) = 0 \tag{2.63}$$

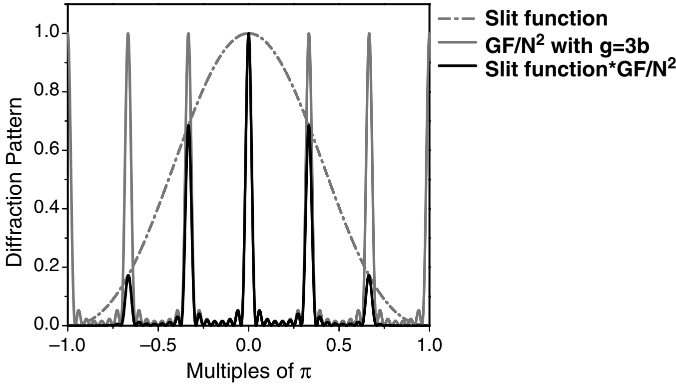


Figure 2.11 Diffraction pattern of a grating with $N = 8$ slits.

with $\cot(x) = \cos(x)/\sin(x)$. Equation 2.62 is the equation for the prominent maxima and is called *grating equation*. Equation 2.63 gives all less prominent maxima in between the prominent maxima. An example is given in Figure 2.11 for a grating with $N = 8$ slits and $g = 3b$. The prominent maxima are at $m \cdot \pi/3$, as expected for $g = 3b$.

Up to now we considered the grating illuminated by a plane wave with its propagation direction perpendicular to the grating, that is, with an angle of incidence $\varepsilon = 0$. When turning to the case that the incident light encloses an angle of incidence ε to the normal on the grating (see Figure 2.12), the above diffraction theory must be adjusted to take into account the angle of incidence.

For $\varepsilon \neq 0$ the partial waves have an additional phase shift given by $g \cdot \sin(-\varepsilon)$. Then, the maxima of the GF obey the grating equation

$$\sin\beta - \sin\varepsilon = m\frac{\lambda}{g}, \qquad m = 0, \pm 1, \pm 2, \pm 3, \ldots \tag{2.64}$$

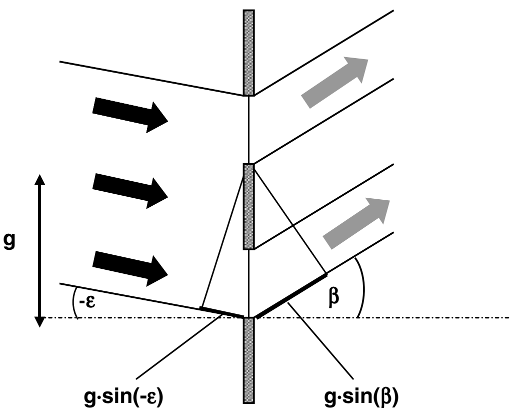


Figure 2.12 Diffraction by a grating with angle of incidence $\varepsilon \neq 0$.

The negative sign of the angle of incidence results from the sign convention for angles: right-handed angles have a negative sign and left-handed angles have a positive sign.

All components of light corresponding to $m = 0$ follow a straight line given by $\beta = \varepsilon$. It is not possible to separate the wavelengths for this order.

Please note that for a reflection grating, the sign of the angle of incidence ε and the sign of the angle of diffraction β are identical by convention. Therefore, for a reflection grating the grating equation reads

$$\sin\beta + \sin\varepsilon = m\frac{\lambda}{g}, \qquad m = 0, \pm1, \pm2, \pm3, \ldots \tag{2.65}$$

For a given set of angles of incidence and diffraction, the grating equation is valid for more than one wavelength, giving rise to several orders of diffracted radiation. For example, if the wavelength $\lambda = 600$ nm gets diffracted in the first order $m = \pm1$, then the wavelength $\lambda = 300$ nm also gets diffracted in the second order $m = \pm2$ at the same angle of diffraction. The number of orders is limited by the grating period g and the angle of incidence. The maximum spectral bandwidth that can be obtained in a specified order without spectral overlap from adjacent orders is called *free spectral range* $F_\lambda = \lambda/m$. An illustration of the free spectral range for the orders $m = \pm1$, ±2, and ±3 and a grating with $g = 2500$ nm and wavelength range 400–700 nm is given in Figure 2.13. The free spectral range becomes smaller for higher orders m. If the grating period decreases, the free spectral range increases.

Depending on the grating period and the angle of incidence ε, it may not be possible to obtain diffracted light. For example, for $g = 500$ nm (groove density $L = 1/g = 2000$ grooves/mm), $\varepsilon = 0°$, and $\lambda = 500$ nm, one obtains $\sin(\beta) = 1$ for the first order $m = \pm1$. Then, all wavelengths $\lambda > 500$ nm cannot be diffracted.

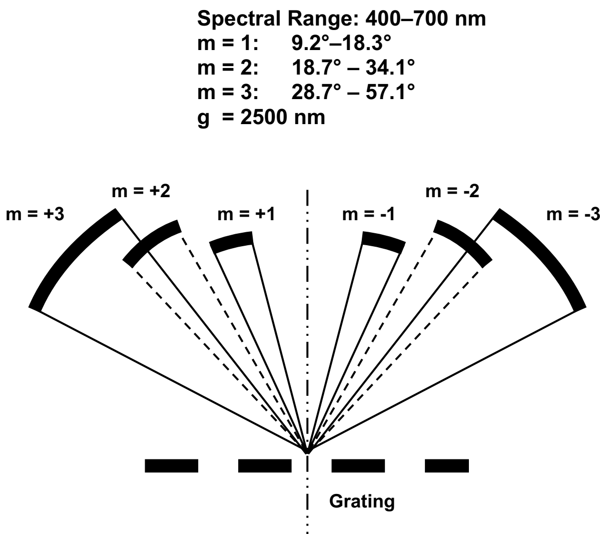


Figure 2.13 Free spectral ranges of the orders $m = \pm1$, ±2, and ±3 for a grating with $g = 2500$ nm.

One of the most important properties of a grating is the *grating efficiency*. The absolute efficiency of a grating is the percentage of incident monochromatic radiation on a grating that is diffracted into a desired order. Diffraction with a grating as described above (slits in a nontransparent thin plate) is symmetric with respect to the energy diffracted into the order $+m$ and $-m$. However, this homogeneous distribution over all diffraction orders can be modified. For that, other kinds of gratings are used with grooves instead of slits. The energy distribution can then be modified by the groove profile, the groove dimensions, the groove materials, and other groove and grating properties such as the reflectivity as in the case of reflection gratings. The result is that the diffracted energy is redirected into a certain desired order or certain orders do not further contribute to the energy distribution of the diffracted light.

These grooved gratings can roughly be divided into reflection gratings and transmission gratings. The difference becomes clear from the name: for reflection gratings the incident light gets reflected by a mirror but gets diffracted by the periodic structure of or on the surface of the mirror, while for transmission gratings the light passes a transparent plate with a periodic transparent structure either on one surface or inside the plate where the light gets diffracted. In the following, we discuss five technically relevant cases, starting with transmission gratings.

2.5.1 Transmission Gratings

The two most important transmission gratings are lamellar transmission gratings with a surface relief profile and holographic transmission gratings. They are discussed now.

2.5.1.1 Lamellar Transmission Gratings

Figure 2.14 shows a schematic drawing of a transmission grating with surface relief profile. The grating consists of grooves of width b and depth d in a bulk transparent

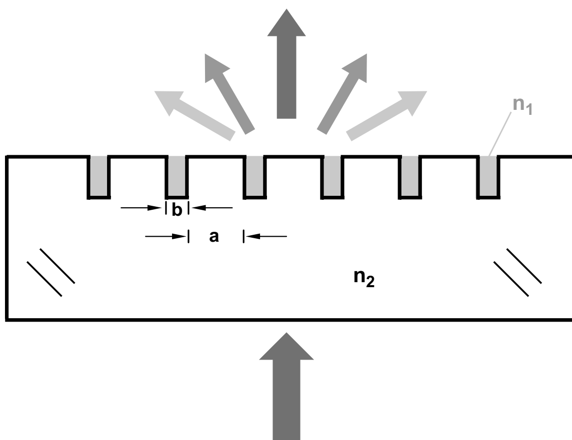


Figure 2.14 Scheme of a lamellar transmission grating with surface relief profile.

material with refractive index n_2. The grooves are separated with a distance a. The grating period is then $g = a + b$. The grooves may be filled with a transparent medium with refractive index n_1. Usually, it is ambient air with $n_1 = 1$.

Light passing through this grating will be diffracted by the periodic structure but the partial waves will have additional phase shifts $\exp(ikn_1 d)$ and $\exp(ikn_2 d)$ due to the fact that they pass either the grooves or the bulk material.
Therefore, we have for the magnitude of electric field of the wave diffracted into the diffraction angle β:

$$E(\beta) = E_0 \cdot \mathrm{GF} \cdot \left\{ \exp(ikn_1 d) \int_0^b \exp(ikx\sin(\beta))\mathrm{d}x + \exp(ikn_2 d) \int_b^{a+b} \exp(ikx\sin(\beta))\mathrm{d}x \right\}, \tag{2.66}$$

where the abbreviation GF stands again for the grating function. As the grating function describes the interference by the periodic arrangements of the grooves with the periodicity g, it is the same as in (2.61). Therefore, we only have to calculate anew the diffraction by a groove. The two integrals in the brackets of (2.66) describe the diffraction at the single groove with width b and the diffraction at the single block between the grooves with width a.

After some mathematics, the intensity $J(\beta)$ diffracted into the diffraction angle β follows as

$$J(\beta) \propto \mathrm{GF}^2 \cdot \left\{ \begin{array}{l} a^2 \dfrac{\sin^2((\pi a/\lambda)\sin(\beta))}{((\pi a/\lambda)\sin(\beta))^2} + b^2 \dfrac{\sin^2((\pi b/\lambda)\sin(\beta))}{((\pi b/\lambda)\sin(\beta))^2} + \\ \quad + 2ab \dfrac{\sin((\pi a/\lambda)\sin(\beta))}{((\pi a/\lambda)\sin(\beta))} \dfrac{\sin((\pi b/\lambda)\sin(\beta))}{((\pi b/\lambda)\sin(\beta))} \cos\left(\dfrac{2\pi d}{\lambda}(n_2 - n_1) + \dfrac{\pi g}{\lambda}\sin(\beta)\right) \end{array} \right\}. \tag{2.67}$$

In the following, we consider diffraction maxima of order m. Furthermore, the grating is illuminated by a parallel beam perpendicular to the grating, that is, the angle of incidence is $\varepsilon = 0$. Then, we can rewrite (2.67) into a dependence upon the diffraction order m for the diffraction maxima

$$J(m) \propto \mathrm{GF}^2 \cdot \left\{ \begin{array}{l} a^2 \dfrac{\sin^2(m\pi(a/g))}{(m\pi(a/g))^2} + b^2 \dfrac{\sin^2(m\pi(b/g))}{(m\pi(b/g))^2} + \\ \quad + 2ab \dfrac{\sin(m\pi(a/g))}{(m\pi(a/g))} \dfrac{\sin(m\pi(b/g))}{(m\pi(b/g))} \cos\left(\dfrac{2\pi d}{\lambda}(n_2 - n_1) + m\pi\right) \end{array} \right\}. \tag{2.68}$$

For each diffraction order $m = \pm 1, \pm 2, \ldots$, it becomes maximum if the cosine term becomes $+1$. This results in a condition for the depth of the grooves d as

$$d = \frac{(2p - m)}{(n_2 - n_1)} \frac{\lambda}{2}, \tag{2.69}$$

with p being an integer number $p=1, 2, 3, \ldots$. The minimum thickness is obtained for $(2p-m)=1$. A further improvement is obtained if $a=b=g/2$, that is, for a symmetric profile. Then, the intensity in a diffraction order m becomes

$$J(m)=2\cdot\left(\frac{E_0\cdot \mathrm{GF}\cdot g}{m\cdot\pi}\right)^2\cdot\left\{1+\cos\left(\frac{2\pi d}{\lambda}(n_2-n_1)+m\pi\right)\right\}. \tag{2.70}$$

An interesting case is the zeroth order $m=0$. Please note that in this case we must use (2.67) and L'Hospital's rule to derive the intensity $J(0)$ to

$$J(0)=2\cdot\left(\frac{E_0\cdot \mathrm{GF}\cdot g}{\pi}\right)^2\cdot\left\{1+\cos\left(\frac{2\pi d}{\lambda}(n_2-n_1)\right)\right\}. \tag{2.71}$$

For the simple grating with slits in a nontransparent plate, it contains the maximum energy. But in the present case, the situation is different due to the introduction of the phase shifts, resulting in the cosine term. For a groove depth

$$d=\frac{\lambda}{2(n_2-n_1)}, \tag{2.72}$$

the intensity $J(0)$ vanishes as the cosine term then becomes -1. Then, also all even orders $m=2q$ vanish completely for this groove depth, as can be easily derived from (2.70). The odd orders have a maximum with their intensity $I(m)$ being proportional to $1/m^2$.

2.5.1.2 **Holographic Transmission Gratings**

Holographic transmission gratings are formed by illuminating a photosensitive material with the interference pattern of two crossed laser beams of same wavelength λ_0 that include an angle γ. The periodicity of the interference pattern is given by

$$G=\frac{\lambda_0}{2\sin(\gamma/2)}. \tag{2.73}$$

Inserting a photosensitive material with refractive index n_{PR} into this interference pattern, the photosensitive material changes its constitution and refractive index in the illuminated regions with a resulting periodicity

$$G_{\mathrm{PR}}=\frac{\lambda_0}{2\cdot n_{\mathrm{PR}}\cdot\sin(\gamma_{\mathrm{PR}}/2)}. \tag{2.74}$$

The angle changes due to refraction into the medium. By this way the sinusoidal profile of the interference pattern is copied into the photoresist. Compared to the surface relief grating, the holographic grating is rather similar, except that the profile is sinusoidal. Figure 2.15 depicts a schematic drawing of a holographic grating.

The calculation of the diffracted intensity is, however, more complicated than for the surface relief grating due to the sinusoidal profile. One has to further distinguish between *amplitude gratings* and *phase gratings*.

For a pure phase grating, the absorption is zero over the whole hologram. The illumination with the interference pattern results in a variation of the optical path,

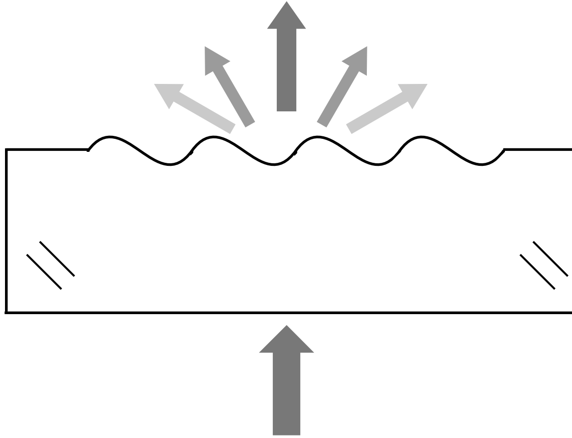


Figure 2.15 Scheme of a holographic transmission grating with sinusoidal profile.

either by a change in the refractive index or by a change in the thickness of the material (when the photoresist gets developed). The latter is similar to the lamellar gratings. Hence, the magnitude of the electric field of the wave diffracted into the diffraction angle β is given by

$$E(\beta) = E_0 \cdot \mathrm{GF} \cdot \int_0^g \exp\left(ikn_{\mathrm{PR}} \cos\left(\frac{2\pi}{g}x\right) \cdot x\right) \exp(ikx \sin(\beta))\,\mathrm{d}x. \qquad (2.75)$$

Despite this more complicated calculation, the principal results are comparable to those of the surface relief transmission grating.

In contrast, the amplitude grating exhibits a change in the absorption of the photoresist in dependence on the illumination. The interference pattern generates a harmonic variation in the transmission through the photoresist. If the thickness of the photoresist film is larger than the grating period g, the grating is called *volume grating* or *Bragg grating*. The interference pattern then causes interference planes (lamellar grating) and one must treat the grating three dimensionally. One possibility to describe the diffraction by such a volume grating is the coupled wave theory of Kogelnik [12]. The great advantage of using this coupled two-wave theory is its computational simplicity and the wide range of variation in the parameters it allows. The disadvantage of this theory is that it considers only the orders $m = 0$ and $m = \pm 1$, and the grating must fulfill the Bragg condition. Gerritsen *et al.* [13] showed that Kogelnik's two-wave coupled theory is quite applicable in detail to deep, narrowly spaced photoresist gratings.

A more complex but exact theory has been developed [14–16] for various grating profiles, but so far only the maximum efficiency of these gratings has been calculated, not the angular or wavelength dependence of the efficiency. Volume gratings have been analyzed by Chang and George [17] using a Raman–Nath formalism modified to incorporate losses. For computation of zero-, first-, and second-order diffracted

beams for a multitude of practical cases, four second-order coupled wave equations are used. Significant differences are found in comparison with computations where only two coupled waves are used. A general expression for the average intensity in each diffracted order is given in Equation 2.12.

Alferness [18, 19] developed in 1975 a model that considers an arbitrary number of diffraction orders. However, it could not be applied for a long time because it needs high computation capacities for numerical calculation. The main idea is simple: the volume grating is divided into many thin gratings. Each grating is considered exactly. The result for intensity diffracted by the volume grating is obtained by the successive calculation of the intensity diffracted by the nth thin grating under consideration of the result for the intensity after $n-1$ preceding gratings.

2.5.2 Reflection Gratings

The simplest reflection grating consists of a periodic arrangement of grooves in a nontransparent highly reflective material, a surface relief grating or lamellar grating. Of more practical and technical interest, however, are gratings with a sawtooth profile, also called blazed gratings. Usually, they are ruled into the support by special ruling machines. Nonruled reflection gratings are made as holographic gratings. These three types are discussed in the following sections.

2.5.2.1 Lamellar Reflection Gratings

Figure 2.16 shows a schematic drawing of a reflection grating with surface relief profile. The grating consists of grooves of width b and depth d in a bulk nontransparent highly reflective material. The grooves are separated with distance a. The grating period is then $g = a + b$. The refractive index of the medium in front of the grating may have the refractive index n. Usually, it is ambient air with $n = 1$.

Light incident perpendicular on this grating will be reflected and diffracted by the periodic structure, but the partial waves reflected in the grooves will have an

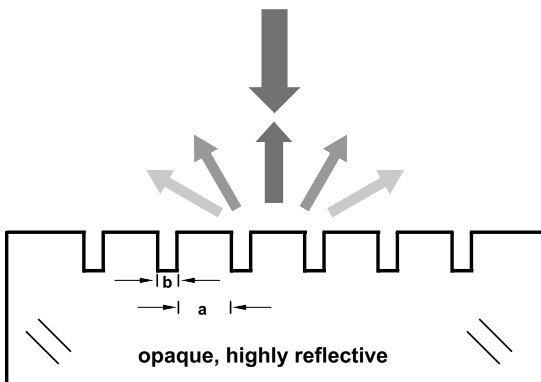


Figure 2.16 Scheme of a reflection grating with surface relief profile (lamellar grating).

additional phase shift $\exp(i2knd)$ due to the fact that they pass the grooves twice compared to the waves reflected at the top surface of the grating.

Therefore, we have for the magnitude of electric field of the wave diffracted into the diffraction angle β

$$E(\beta) = E_0 \cdot \mathrm{GF} \cdot \left\{ \exp(i2knd) \int_0^b \exp(ikx\sin(\beta))\mathrm{d}x + \int_b^{a+b} \exp(ikx\sin(\beta))\mathrm{d}x \right\}, \tag{2.76}$$

where the abbreviation GF is again the grating function. The two integrals in the brackets of (2.76) describe the diffraction at the single groove and the diffraction at the single block, respectively, between the grooves.

After some mathematics, the intensity $J(\beta)$ diffracted into the diffraction angle β follows as

$$J(\beta) \propto \mathrm{GF}^2 \cdot \left\{ \begin{array}{l} a^2 \dfrac{\sin^2((\pi a/\lambda)\sin(\beta))}{((\pi a/\lambda)\sin(\beta))^2} + b^2 \dfrac{\sin^2((\pi b/\lambda)\sin(\beta))}{((\pi b/\lambda)\sin(\beta))^2} + \\ + 2ab \dfrac{\sin((\pi a/\lambda)\sin(\beta))}{((\pi a/\lambda)\sin(\beta))} \dfrac{\sin((\pi b/\lambda)\sin(\beta))}{((\pi b/\lambda)\sin(\beta))} \cos\left(\dfrac{4\pi nd}{\lambda} + \dfrac{\pi g}{\lambda}\sin(\beta)\right) \end{array} \right\} \tag{2.77}$$

This result is very similar to that of the transmission grating with surface relief profile. The difference is in the cosine term where the difference between the refractive indices of the bulk material and the material in the grooves $(n_2 - n_1)$ is now replaced by the term $2n$.

In the following, we consider again diffraction maxima of order m for the grating being illuminated by a parallel beam perpendicular to the grating, that is, the angle of incidence ε is $\varepsilon = 0$, and in the symmetric case $a = b = g/2$. Then, we obtain

$$J(m) = 2 \cdot \left(\frac{E_0 \cdot \mathrm{GF} \cdot g}{m \cdot \pi}\right)^2 \cdot \left\{1 + \cos\left(\frac{4\pi nd}{\lambda} + m\pi\right)\right\}. \tag{2.78}$$

For each diffraction order m, it becomes maximum if the cosine term becomes $+1$. This results in a condition for the depth of the grooves d as

$$d = \frac{(2p - m)\lambda}{4n} \tag{2.79}$$

with p being an integer number $p = 1, 2, 3, \ldots$

An interesting case is again the zeroth order $m = 0$. The intensity $J(0)$ is obtained as

$$J(0) = 2 \cdot \left(\frac{E_0 \cdot \mathrm{GF} \cdot g}{\pi}\right)^2 \cdot \left\{1 + \cos\left(\frac{4\pi nd}{\lambda}\right)\right\}. \tag{2.80}$$

Due to the phase shift $\exp(i2knd)$, it is possible to suppress the zeroth order completely if the groove depth d satisfies the condition

$$d = \frac{\lambda}{4n}. \tag{2.81}$$

Also, all even orders $m = 2q$ vanish completely for this groove depth, as can be easily derived from (2.78). The odd orders have a maximum with their intensity $J(m)$ being proportional to $1/m^2$.

2.5.2.2 Blazed Gratings

Gratings with grooves that have a sawtooth profile with one side longer than the other are called *blazed gratings*. The angle between the long side and the plane of the grating is called *blaze angle* θ_B. The scheme of a blazed grating is shown in Figure 2.17.

A blazed grating combines diffraction with reflection in order to put most of the light into one diffracted order.

Considering the case that the diffraction direction for an order m is the same as the direction of specular reflection at the facet normal, we have $\varepsilon - \theta_B = \theta_B - \beta$ or $\varepsilon + \beta = 2\theta_B$. Then, the intensity in this order is maximum, and the regular reflection (the zeroth order) is also in this direction. The grating equation reads in this case as follows:

$$\sin(2\theta_B - \varepsilon) + \sin\varepsilon = m\frac{\lambda}{g}. \tag{2.82}$$

Blazed grating groove profiles are calculated for the Littrow configuration for the diffraction order $m = 1$. In the Littrow configuration, incident and reflected angle are equal. This is fulfilled if $\varepsilon = \beta = \theta_B$. Inserting this into (2.82) leads to the blaze condition

$$2\sin\theta_B = \frac{\lambda_B}{g}. \tag{2.83}$$

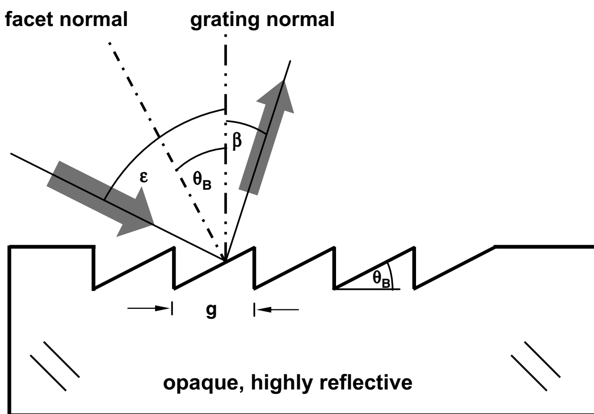


Figure 2.17 Scheme of a blazed grating.

By comparing the blazed grating with a lamellar grating, we find that if g is the grating constant of the blazed grating, the corresponding grating constant g_{lamellar} is $g_{\text{lamellar}} = g/\tan(\theta_B)$.

2.5.2.3 **Holographic Gratings**

For holographic reflection gratings, a polished highly reflecting support gets covered with a photoresist. The grating is formed by illuminating the photosensitive material with the interference pattern of two crossed laser beams of same wavelength. Then, the photoresist changes its constitution and refractive index in the illuminated regions. By this way, the sinusoidal profile of the interference pattern is copied into the photoresist. This process leads to strictly parallel and equally spaced lines with sinusoidal profile. The sinusoidal profile causes that the brightness achieved by the conventional holographic reflection grating is not as high as that of ruled gratings. To improve brightness, this sinusoidal profile is machined by ion bombardment to achieve an echelle profile. Such blazed holographic gratings are much brighter than conventional holographic gratings.

2.6 Scattering

Reflection, refraction, and diffraction of light and other electromagnetic radiation represent a redirection of the light from its original path into distinct directions given by the reflection law, the Snell's law, or the grating equation.

Unlike these processes, scattering of radiation distributes the light in all directions. Thereby, the ratio (size of the scatterer)/wavelength plays an important role for the spatial distribution of light. A scatterer very small compared to the wavelength (Rayleigh limit) scatters the light symmetrically in forward and backward direction. With increasing size, however, forward scattering dominates the spatial distribution. Not only the size of the scatterer but also its shape, constitution, and refractive index affect light scattering. The size has the main effect. Finally, the concentration of scatterers in a volume influences the spatial distribution of scattered light. Densely packed scatterers lead to an almost homogeneous distribution of scattered light like, for example in clouds, fog, dispersion colors, or paper. In each case, the scattering diminishes the radiation flux in propagation direction of the reflected, refracted, or diffracted light by the redistribution of the incident light into all solid angles.

For this book, the scattering caused by surface imperfections is of interest as opposed to scatter from individual molecules and particles as scatterer. The latter can be read in more detail in several monographs [20–25]. Light scattering from optical coatings has its origin in the structural irregularities at interfaces and in the bulk of materials. As we will see in Section 3.1.4, the scattering by these imperfections can be taken into account by modified Fresnel equations that consider the structural irregularities by an rms roughness parameter.

2.7 Dielectric Function and Refractive Index

For describing interaction of light with matter, the photon energy $\hbar\omega$ and the complex dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ are the physically relevant quantities, while when discussing wave propagation in and through media the wavelength λ and the complex refractive index $n + i\kappa$ are much proper.
Dielectric function and refractive index are connected by Maxwell's relation

$$n + i\kappa = \sqrt{\varepsilon_1 + i\varepsilon_2}. \tag{2.84}$$

This relation can be rewritten either for the dielectric function as

$$\varepsilon_1 = 2 \cdot n \cdot \kappa, \tag{2.85}$$

$$\varepsilon_2 = n^2 - \kappa^2, \tag{2.86}$$

or for the refractive index as

$$n = \sqrt{\frac{\varepsilon_1}{2} + \frac{1}{2}\sqrt{\varepsilon_1^2 + \varepsilon_2^2}}, \tag{2.87}$$

$$\kappa = \sqrt{-\frac{\varepsilon_1}{2} + \frac{1}{2}\sqrt{\varepsilon_1^2 + \varepsilon_2^2}}. \tag{2.88}$$

We will introduce physical models for the dielectric function in Section 2.7.1. In Section 2.7.2, we discuss Kramers–Kronig relations for the real and imaginary part of the dielectric function. Finally, in Section 2.7.3 we introduce empiric formulas for the refractive index. All models for the dielectric function and the refractive index are based on a parameterization of the corresponding quantity.

2.7.1 Models for the Dielectric Function

Interaction of electromagnetic fields with matter is dominated by the forces exerted by the incident electric (and magnetic) field on the electric charges in the matter. The electric field **E** inside the body of condensed matter usually displaces only the electrons while the ions are too inert as to follow the electric field with the same frequency. Thereby, each atom becomes an electric dipole with dipole moment **p**. At low frequencies in the far infrared, the incident light can also couple to the ions via TO phonons and induces dipole moments by displacing the negative and positive charged ions in different directions.

The dipole moments add up to a macroscopic net polarization **P** of the sample that is related to the electric field **E** by the general equation

$$\mathbf{P} = \varepsilon_0 \chi \mathbf{E}. \tag{2.89}$$

The factor χ is the macroscopic *susceptibility* of the matter. In general, χ depends on $\mathbf{E}$, that is, $\chi(\mathbf{E})$. This field dependence can be described by a series expansion of $\mathbf{P}$ in powers of $\mathbf{E}$, where susceptibilities $\chi_{ijk}\ldots$ are introduced, which are now a constant in relation to $\mathbf{E}$

$$P_i = \underbrace{\varepsilon_0 \sum_j \chi_{ij} E_j}_{\substack{\text{linear term } \chi^{(1)} \\ \text{(anisotropic medium)}}} + \underbrace{\varepsilon_0 \sum_j \sum_k \chi_{ijk} E_j E_k}_{\text{quadratic term } \chi^{(2)}} + \underbrace{\varepsilon_0 \sum_j \sum_k \sum_l \chi_{ijkl} E_j E_k E_l + \ldots}_{\text{cubic term } \chi^{(3)}} \tag{2.90}$$

for each component $i = x, y, z$ of $\mathbf{P}$. As we are interested in optical thickness determination, we can restrict our discussion in the following to the linear term. Nonlinear effects are negligible or even do not occur. Then, it might be necessary to consider anisotropy in the linear term. For simplicity, we assume optical isotropy for which the susceptibility simply is a scalar quantity.

The linearity means that charges displaced by the applied field from their position of equilibrium are retreated by forces that are proportional to the distance from the position of equilibrium. In consequence, a charge q_j with mass m_j executes forced oscillations in a time periodic electric field $\mathbf{E} = \mathbf{E}_0(\mathbf{r})\exp(-i\omega t)$. This is the *harmonic oscillator model* because the retreating force is linear in the displacement $\mathbf{r}_j$. It was developed by H. A. Lorentz in the beginning of the twentieth century [26].

From the force balance, we obtain the following relation for the displacement $\mathbf{r}_j$

$$\sum_j m_j \frac{\partial^2 \mathbf{r}_j}{\partial t^2} + m_j \gamma_j \frac{\partial \mathbf{r}_j}{\partial t} + D_j \mathbf{r}_j = \sum_j q_j \mathbf{E}(t). \tag{2.91}$$

The second term on the left side of (2.91) accounts for the perturbation of the movement of charge q_j by interactions with other charges and lattice defects. Their contributions are assumed to lead to a damping of the oscillation with damping constant γ_j. For a time harmonic field $\mathbf{E} = \mathbf{E}_0(\mathbf{r})\exp(-i\omega t)$, the solution of (2.91) is given as

$$\mathbf{r}_j = \frac{q_j}{m_j} \frac{\mathbf{E}_0}{(\omega_j^2 - \omega^2 - i\omega\gamma_j)}, \tag{2.92}$$

with the resonance frequency $\omega_j = D_j/m_j$. The macroscopic net polarizability $\mathbf{P}$ follows from all dipole moments $\mathbf{p}_j = q_j \mathbf{r}_j$ in the sample volume V via

$$\mathbf{P} = \frac{1}{V} \sum_j N_j \mathbf{p}_j = \frac{1}{V} \sum_j N_j q_j \mathbf{r}_j, \tag{2.93}$$

from which we obtain the susceptibility χ_j for all N_j charges q_j as

$$\chi_j(\omega) = \frac{1}{V} \sum_j \frac{N_j q_j^2}{\varepsilon_0 m_j} \frac{1}{(\omega_j^2 - \omega^2 - i\omega\gamma_j)}. \tag{2.94}$$

Here, N_j/V is the total number density of charges q_j with mass m_j and eigenfrequency ω_j. In this respect, χ in (2.94) is also valid for ionic crystals (e.g., NaCl and

MgO), with the corresponding ionic charges and masses. The assumption of linearity of (2.90) allows to add all various contributions χ_j to a total susceptibility χ. Then, the dielectric function $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ is

$$\varepsilon(\omega) = 1 + \chi = 1 + \sum_j \frac{\omega_{Pj}^2}{(\omega_j^2 - \omega^2 - i\omega\gamma_j)} \tag{2.95}$$

with the abbreviation

$$\omega_{Pj}^2 = \frac{N_j q_j^2}{V\varepsilon_0 m_j} \tag{2.96}$$

being called *plasma frequency* of the *j*th oscillator.

Note that the ansatz $\exp(-i\omega t)$ for the time dependence of the electric field leads to the definition of $\varepsilon(\omega)$ as $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ and to $\tilde{n}(\omega) = n(\omega) + i\kappa(\omega)$. Alternatively, the ansatz $\exp(i\omega t)$ leads to the definition of $\varepsilon(\omega)$ as $\varepsilon(\omega) = \varepsilon_1(\omega) - i\varepsilon_2(\omega)$ and to $\tilde{n}(\omega) = n(\omega) - i\kappa(\omega)$, which can also be found in common literature.

At low frequencies, that is, for $\omega \to 0$, the real part $\varepsilon_1(0)$ of the dielectric function becomes constant

$$\varepsilon_1(0) = 1 + \sum_j \frac{\omega_{Pj}^2}{\omega_j^2}, \tag{2.97}$$

while the imaginary part $\varepsilon_2(0)$ vanishes. $\varepsilon_1(0)$ represents the static dielectric constant of the material. The constant ratio ω_{Pj}^2/ω_j^2 defines a new quantity, the *oscillator strength* S_j of the *j*th harmonic oscillator, so that ω_{Pj}^2 in (2.95) can be replaced by $S_j\omega_j^2$.

Figure 2.18 exemplarily shows the dielectric function and the corresponding refractive index of a harmonic oscillator with resonance frequency $3.5 \times 10^{15}\,\mathrm{s}^{-1}$,

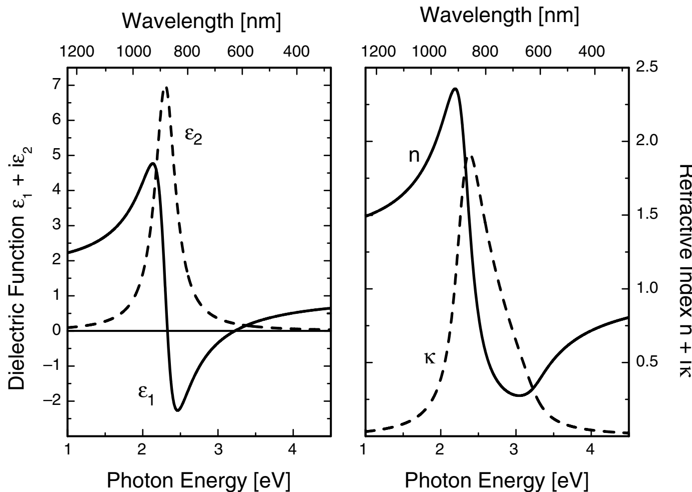


Figure 2.18 Dielectric function and refractive index of a harmonic oscillator with resonance frequency $3.5 \times 10^{15}\,\mathrm{s}^{-1}$, damping constant $3.5 \times 10^{14}\,\mathrm{s}^{-1}$, and oscillator strength $S = 1$.

damping constant $3.5 \times 10^{14}\,\mathrm{s}^{-1}$, and oscillator strength $S = 1$. At the resonance frequency, the imaginary part exhibits a maximum and rapidly decreases to the right and the left of the resonance frequency. Far from the resonance frequency, ε_2 vanishes. The real part decreases at high frequencies when approaching the resonance frequency and even becomes negative. In the vicinity of the resonance frequency, the real part changes very rapidly to high positive values from which it continuously decreases again when going to low frequencies. The decrease with decreasing frequency is called *normal dispersion*, while in the region of rapid increase with decreasing frequency it is called *anomal dispersion*.

For *statistically perturbed* or *amorphous materials*, it seems appropriate to use an extension of the harmonic oscillator model on the so-called *Brendel oscillators*. A Brendel oscillator is a harmonic oscillator with eigenfrequency ω_k and width γ_k that is inhomogeneously broadened by an infinite sum over sharp harmonic oscillators with eigenfrequency x and width γ_k [27]. These oscillators are Gaussian distributed around the harmonic oscillator with eigenfrequency ω_k with a standard deviation σ_k:

$$\chi_{\mathrm{Brendel}}(\omega) = \frac{\omega_{\mathrm{P}k}^2}{\sigma_k\sqrt{2\pi}} \int_0^\infty \mathrm{d}x \frac{\exp\left(-0.5((x-\omega_k/\sigma_k))^2\right)}{x^2-\omega^2-i\omega\gamma_k}. \tag{2.98}$$

The advantage of a Brendel oscillator compared to a harmonic oscillator simply broadened by an increased γ is that the contours in the real and imaginary part of the dielectric function become smoother due to the Gaussian distribution. An example for a Brendel oscillator with resonance frequency $3.5 \times 10^{15}\,\mathrm{s}^{-1}$, damping constant $3.5 \times 10^{14}\,\mathrm{s}^{-1}$, oscillator strength $S = 1$, and width parameter $0.05 = 5\%$ is given in Figure 2.19.

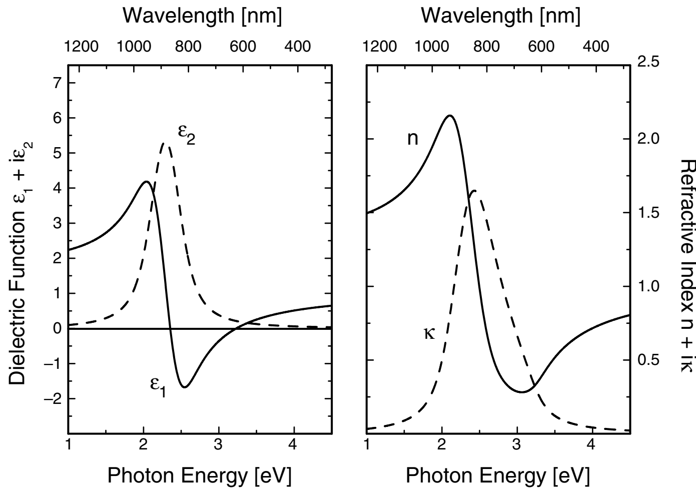


Figure 2.19 Dielectric function and refractive index of a Brendel oscillator with resonance frequency $3.5 \times 10^{15}\,\mathrm{s}^{-1}$, damping constant $3.5 \times 10^{14}\,\mathrm{s}^{-1}$, oscillator strength $S = 1$, and width parameter 0.05.

In the simpler model of Kim [28], the damping γ_k of a harmonic oscillator ω_k is assumed to be frequency dependent:

$$\chi_{\mathrm{Kim}}(\omega) = \frac{\omega_{\mathrm{P},k}^2}{\omega_k^2 - \omega^2 - i\omega\gamma_k(\omega)} \tag{2.99}$$

with

$$\gamma_k(\omega) = \gamma_k \exp\left(-\frac{1}{1+\sigma_k^2}\left(\frac{\omega-\omega_k}{\gamma_k}\right)^2\right). \tag{2.100}$$

The parameter $\sigma_k \geq 0$ is used to switch between a Gaussian or a Lorentzian shape of $\gamma_k(\omega)$. For $\sigma_k = 0$, a pure Gaussian shape is obtained, and for $\sigma_k > 5$ a Lorentzian shape is obtained.

In metals, semimetals, and semiconductors, an important contribution to the dielectric function stems from unbound charge carriers, the so-called *free electrons*. Within the harmonic oscillator model, their contribution is obtained when assuming the eigenfrequency $\omega_{\mathrm{fe}} = 0$, corresponding to moving in a potential $V(\mathbf{r}) = 0$. Then, the susceptibility of the free electrons – the *Drude susceptibility* [29, 30] – reads

$$\chi_{\mathrm{fe}}(\omega) = -\frac{\omega_{\mathrm{P}}^2}{\omega^2 + i\omega\gamma_{\mathrm{fe}}} \tag{2.101}$$

with the abbreviation

$$\omega_{\mathrm{P}}^2 = \frac{Ne_0^2}{Vm_{\mathrm{eff}}\,\varepsilon_0} \tag{2.102}$$

being the *plasma frequency* of the free electrons assuming them as a plasma. In a parabolic band structure, the *effective mass* m_{eff} of the electrons is identical to the electron mass m_{e}, but in nonparabolic band structures m_{eff} may differ from m_{e}.

Figure 2.20 displays the dielectric function $\varepsilon_{\mathrm{Drude}} = 1 + \chi_{\mathrm{fe}}$ and the corresponding refractive index of a pure Drude metal with plasma frequency $\omega_{\mathrm{P}} = 5 \times 10^{15}\,\mathrm{s}^{-1}$ and damping constant $\gamma_{\mathrm{fe}} = 5 \times 10^{14}\,\mathrm{s}^{-1}$. The real part of $\varepsilon_{\mathrm{Drude}}$ becomes zero at $\omega = \omega_{\mathrm{P}}$ that corresponds to a photon energy of 3.275 eV. At this frequency, it is possible to excite longitudinal oscillations of the electron plasma. They can be described with the curl-free solution $\mathbf{L}$ of the Maxwell equations and the Helmholtz equation (see Section 2.1). For realistic metals, the contribution of the interband transitions shifts this position to lower photon energies (longer wavelengths). At $\omega > \omega_{\mathrm{P}}$ the real part is positive and the imaginary part tends to zero. In this spectral region, the Drude metal behaves like a nonabsorbing dielectrics. Indeed, for a few simple metals, this transparency was proven at very high frequencies [31].

For $\omega < \omega_{\mathrm{P}}$, the real part ε_1 becomes negative and rapidly decreases with decreasing frequency. The imaginary part ε_2 is positive and increases with decreasing frequency, but with a smaller slope than that of the real part. For the corresponding refractive index in Figure 2.20, the real part n tends to zero while κ strongly increases with decreasing frequency. The consequences of this behavior are as follows:

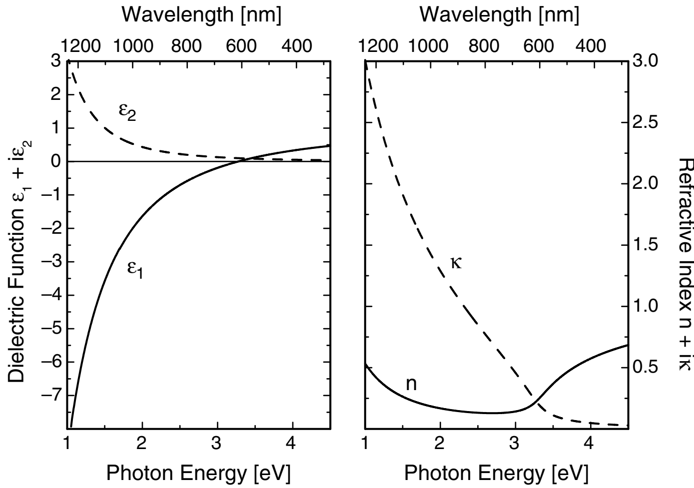


Figure 2.20 Dielectric function and refractive index of a pure Drude metal with plasma frequency $5 \times 10^{15}\,\mathrm{s}^{-1}$ and damping constant $5 \times 10^{14}\,\mathrm{s}^{-1}$.

1) Strong absorption of electromagnetic waves in a metal. The electromagnetic wave cannot penetrate too far into the metal. Typical penetration depths of realistic metals lie on the order of 20–30 nm, depending on the metal.
2) The reflectivity of a metal approaches 100%. For negligible n and large κ values, the reflectivity at normal incidence becomes

$$R = \frac{(n-1)^2 + \kappa^2}{(n+1)^2 + \kappa^2} \to \frac{1+\kappa^2}{1+\kappa^2} = 1. \tag{2.103}$$

For small frequencies ω, the imaginary part increases like $1/\omega$ with decreasing ω and even diverges at $\omega = 0$. However, the DC conductivity remains finite, approaching the value

$$\sigma_{\mathrm{DC}} = \varepsilon_0 \frac{\omega_{\mathrm{p}}^2}{\gamma_{\mathrm{fe}}}. \tag{2.104}$$

The classical Drude model works with a damping constant that does not depend upon frequency. This is a good approximation in most cases. However, there are situations where the damping of the free carriers exhibits a characteristic dependence on frequency, for example, in the case of scattering at charged impurities. The *extended Drude* model [32] uses a rather simple but successful choice of the frequency dependence of the damping term:

$$\gamma_{\mathrm{fe}}(\omega) = \Gamma_{\mathrm{L}} - \frac{\Gamma_{\mathrm{L}} - \Gamma_{\mathrm{H}}}{\pi}\left[\tan^{-1}\left(\frac{\omega - \Omega_{\Gamma}}{\Gamma_{\Gamma\omega}}\right) + \frac{\pi}{2}\right]. \tag{2.105}$$

The function for the damping constant is chosen to change smoothly from a constant Γ_L at low frequencies to another constant level Γ_H at high frequency. The transition region is defined by the crossover frequency Ω_Γ – the center of the transition region – and the crossover width parameter $\Gamma_{\Gamma\omega}$.

Amorphous semiconductor and oxide materials often have optical functions that depend upon deposition conditions and do not have so sharp features like a harmonic oscillator. The first expression for the imaginary part of their dielectric function above the band edge was suggested by Tauc *et al.* [33]. Jellison and Modine [34] derived a model based on a combination of the Tauc band egde and the Lorentz oscillator formulation. With this *Tauc–Lorentz model*, the imaginary part of the complex dielectric function of amorphous materials with band gap (mainly semiconductor materials) can be modeled as

$$\varepsilon_{2\mathrm{TL}}(\omega) = \frac{\omega_\mathrm{P}^2}{\omega} \frac{\omega_\mathrm{res} \cdot \gamma \cdot \left(\omega - \omega_\mathrm{gap}\right)^2}{\left(\omega^2 - \omega_\mathrm{res}^2\right)^2 + \omega^2\gamma^2}. \tag{2.106}$$

Here, the oscillator has a resonance frequency ω_res and a damping constant γ. ω_gap is the frequency corresponding to the band gap energy $E_\mathrm{gap} = \hbar\omega_\mathrm{gap}$. The plasma frequency is proportional to the momentum transition matrix element. Note that in the original paper, the formula is expressed in terms of photon energies.

The real part $\varepsilon_{1,\mathrm{TL}}$ is obtained from the imaginary part by Kramers–Kronig relation (see Section 2.7.2). The dielectric function of a Tauc–Lorentz oscillator with resonance frequency $3.5 \times 10^{15}\,\mathrm{s}^{-1}$, damping constant $3.5 \times 10^{14}\,\mathrm{s}^{-1}$, oscillator strength $S = 1$, and band gap 1.5 eV is depicted in Figure 2.21.

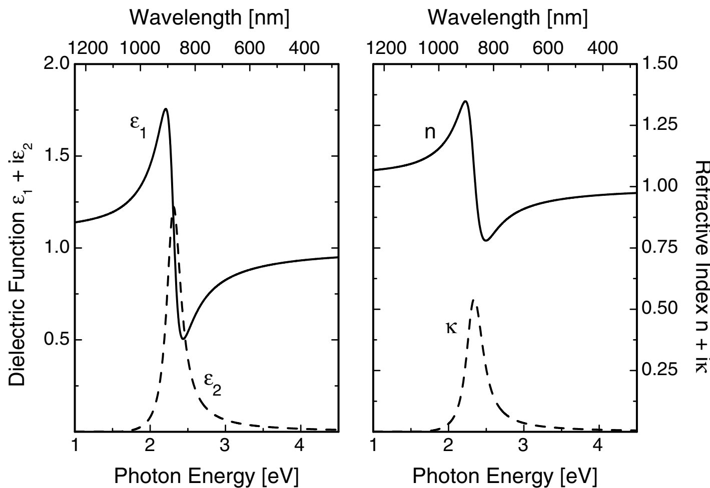


Figure 2.21 Dielectric function of a Tauc–Lorentz oscillator with resonance frequency $3.5 \times 10^{15}\,\mathrm{s}^{-1}$, damping constant $3.5 \times 10^{14}\,\mathrm{s}^{-1}$, oscillator strength $S = 1$, and band gap 1.5 eV.

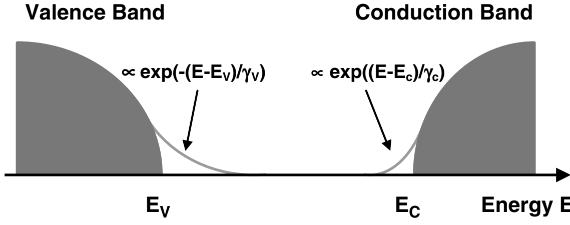


Figure 2.22 Density of states $N(E)$ in the model of O' Leary, Johnson, and Lim [35] for an amorphous semiconductor.

Another model for the dielectric function of amorphous semiconductor materials stems from O' Leary, Johnson, and Lim [35] and is well known as *OJL model*. In a defect-free crystalline semiconductor, the absorption spectrum terminates abruptly at the energy gap. In contrast, in an amorphous semiconductor the absorption spectrum reaches into the gap region. The reason is that the electronic states in the valence band and conduction band can be divided into localized states and states that are randomly distributed through these amorphous semiconductors.

While the distribution of localized states follow a square root functional dependence in the band region, the distribution of randomly distributed states shows an exponential functional dependence in the tail region. Introducing breadths γ_C and γ_V (both energies) for the conduction and valence band tails, O' Leary, Johnson, and Lim modeled the density of states for the conduction and valence band of an amorphous semiconductor. Figure 2.22 shows the density of states $N(E)$ in this model.

O' Leary, Johnson, and Lim derived the optical absorption coefficient $\alpha_{\mathrm{OJL}}(E)$ with $E = \hbar\omega$ as

$$\alpha_{\mathrm{OJL}}(E) = D^2(E)\frac{\sqrt{2}}{\pi^2\hbar^3}m_C^{3/2}\frac{\sqrt{2}}{\pi^2\hbar^3}m_V^{3/2}J(E). \tag{2.107}$$

Here, m_C and m_V are the effective mass of electrons in the conduction and valence band. The optical transition matrix element $D^2(E)$ is proportional to $1/E$ but its exact functional dependence remains unknown and must be adjusted accordingly.

The normalized joint density of states (JDOS) $J(E)$ in the OJL model differs for the two cases $E \leq E_{\mathrm{gap}} + (\gamma_C + \gamma_V)/2$ and $E \geq E_{\mathrm{gap}} + (\gamma_C + \gamma_V)/2$:

1) $E \leq E_{\mathrm{gap}} + (\gamma_C + \gamma_V)/2$

$$J(E) = \frac{\gamma_C^2}{\sqrt{2e}}\cdot\exp\left(\frac{E-E_{\mathrm{gap}}}{\gamma_C}\right)\cdot Y\left(\frac{\gamma_V}{2\gamma_C}\right) + \frac{\gamma_V^2}{\sqrt{2e}}\cdot\exp\left(\frac{E-E_{\mathrm{gap}}}{\gamma_V}\right)\cdot Y\left(\frac{\gamma_C}{2\gamma_V}\right)$$
$$+\frac{1}{2\sqrt{e}}\frac{(\gamma_C\gamma_V)^{3/2}}{\gamma_V-\gamma_C}\cdot\left[\exp\left(\frac{E-E_{\mathrm{gap}}-\gamma_C/2}{\gamma_V}\right)-\exp\left(\frac{E-E_{\mathrm{gap}}-\gamma_V/2}{\gamma_C}\right)\right] \tag{2.108}$$

with $Y(z) = \sqrt{z}\cdot\exp(-z) + 0.5\sqrt{\pi}\cdot\mathrm{erfc}(\sqrt{z})$ and erfc = error function.

2) $E \geq E_{gap} + (\gamma_C + \gamma_V)/2$

$$J(E) = \frac{\gamma_C^2}{\sqrt{2e}} \cdot \exp\left(\frac{E - E_{gap}}{\gamma_C}\right) \cdot Y\left(\frac{E - E_{gap}}{\gamma_C} - \frac{1}{2}\right)$$
$$+ \frac{\gamma_V^2}{\sqrt{2e}} \cdot \exp\left(\frac{E - E_{gap}}{\gamma_V}\right) \cdot Y\left(\frac{E - E_{gap}}{\gamma_C} - \frac{1}{2}\right) +$$
$$+ \frac{1}{2\sqrt{e}} \left(E - E_{gap}\right)^2 \cdot L\left(\frac{\gamma_C}{2(E - E_{gap})}, \frac{\gamma_V}{2(E - E_{gap})}\right) \tag{2.109}$$

with $L(x, y) = \int_x^{1-y} \sqrt{z}\sqrt{1-z}\,dz$.

The imaginary part of the complex dielectric function of amorphous semiconductor materials with band gap E_{gap} finally follows from the optical absorption coefficient $\alpha_{OJL}(E)$ as

$$\varepsilon_{2,OJL}(E) = \frac{\hbar c \cdot \alpha_{OJL}(E)}{E}. \tag{2.110}$$

The real part $\varepsilon_{1,OJL}$ is obtained from the imaginary part by Kramers–Kronig relations (see Section 2.7.2). An example for the dielectric function according to the OJL model with band gap 2.0 eV, valence band damping 0.03 eV, conduction band damping 0.03 eV, and decay parameter 0.2 eV is given in Figure 2.23.

A controversial model is the model of Forouhi and Bloomer [36] for amorphous materials. It gives a relation for the absorption index κ in dependence on the photon

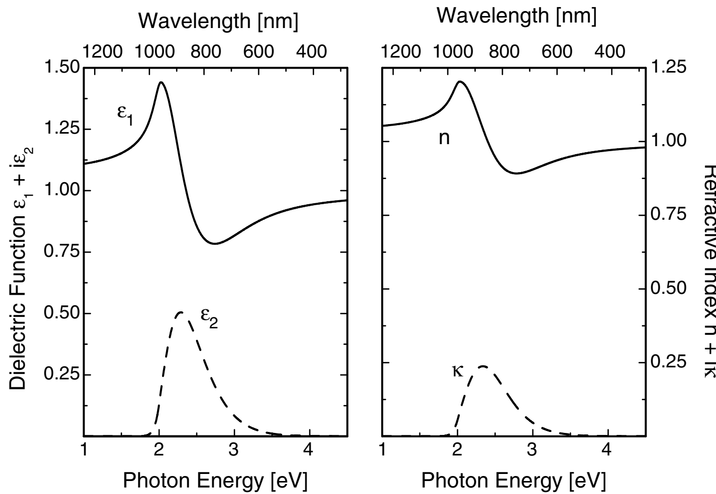


Figure 2.23 Dielectric function according to the OJL model: band gap 2.0 eV, valence band damping 0.03 eV, conduction band damping 0.03 eV, and decay parameter 0.2 eV.

energy E rather than on the imaginary part ε_2 of the dielectric function

$$\kappa_{\mathrm{FB}}(E) = \sum_{j=1}^{N} \frac{A_j\left(E - E_{\mathrm{gap}}\right)^2}{E^2 - B_j E + C_j}, \tag{2.111}$$

with A_j, B_j, and C_j being positive, nonzero constants characteristic of the medium, so that $4C_j - B_j^2 > 0$. E_{gap} is the band gap energy. The corresponding refractive index n_{FB} is obtained by a Hilbert transform

$$n_{\mathrm{FB}}(E) = n_{\mathrm{FB}}(\infty) + \frac{1}{\pi}\wp \int_{-\infty}^{\infty} \frac{k_{\mathrm{FB}}(\Omega) - k_{\mathrm{FB}}(E)}{\Omega - E}\,\mathrm{d}\Omega, \tag{2.112}$$

where $\wp$ is the principal value of the integral, to

$$n_{\mathrm{FB}}(E) = n_{\mathrm{FB}}(\infty) + \sum_{j=1}^{N} \frac{B_{0j}E + C_{0j}}{E^2 - B_j E + C_j}. \tag{2.113}$$

B_{0j} and C_{0j} are constants that depend on A_j, B_j, C_j, and $n_{\mathrm{FB}}(\infty)$ is a constant greater than unity.

There are some problems with these formulas:

- For certain parameters A, B, and C, the absorption index may be less than zero, $k_{\mathrm{FB}}(E) < 0$, for $E < E_{\mathrm{gap}}$. This is unphysical.
- $k_{\mathrm{FB}}(E)$ becomes constant for $E \to \infty$. In experiments, $k(E) \to 0$ for $E \to \infty$ proportional to $1/E^3$.
- The time reversal symmetry requires that $k(-E) = -k(E)$, which is not satisfied.
- As $k_{\mathrm{FB}}(\infty) \neq 0$, Kramers–Kronig relations are not fulfilled.

At the end of this section, we give five selected examples for the dielectric function of real materials and the corresponding complex refractive index following from Maxwell's equation (2.84), and discuss also the reflectance and transmittance of a thick layer with $d = 1000\,\mu\mathrm{m}$ according to these dielectric functions.

We start with the semiconducting material silicon Si in Figure 2.24. The optical constants are taken from Ref. [37]. The dielectric function is plotted in the wavelength range 0.2–1.6 μm in Figure 2.24a. It is conditioned by two strong harmonic oscillators in the wavelength range 0.2–0.4 μm. The oscillator strengths are as high as the real part ε_1 becomes negative and alters from -20 to $+40$ within a wavelength band of only 100 nm width. At wavelengths longer than 400 nm, ε_1 exhibits normal dispersion and ε_2 rapidly decreases. As a consequence of the high value of ε_1, the refractive index n of silicon is also high and amounts to values between 5 and 3.4 in the spectral region above 400 nm wavelength as can be recognized from Figure 2.24b. Looking at the reflectance and transmittance of a layer of $d = 1000\,\mu\mathrm{m}$ in Figure 2.24c, we can recognize that the two harmonic oscillators lead to two maxima in the reflectance. In this spectral region, the reflectance spectrum is rather sensitive to changes caused by thin transparent films, either formed as natural oxide or deposited by chemical or physical methods. The absorption in silicon is as high as a layer of $d = 1000\,\mu\mathrm{m}$ is not transparent below 1000 nm wavelength, but decreases from short wavelengths to

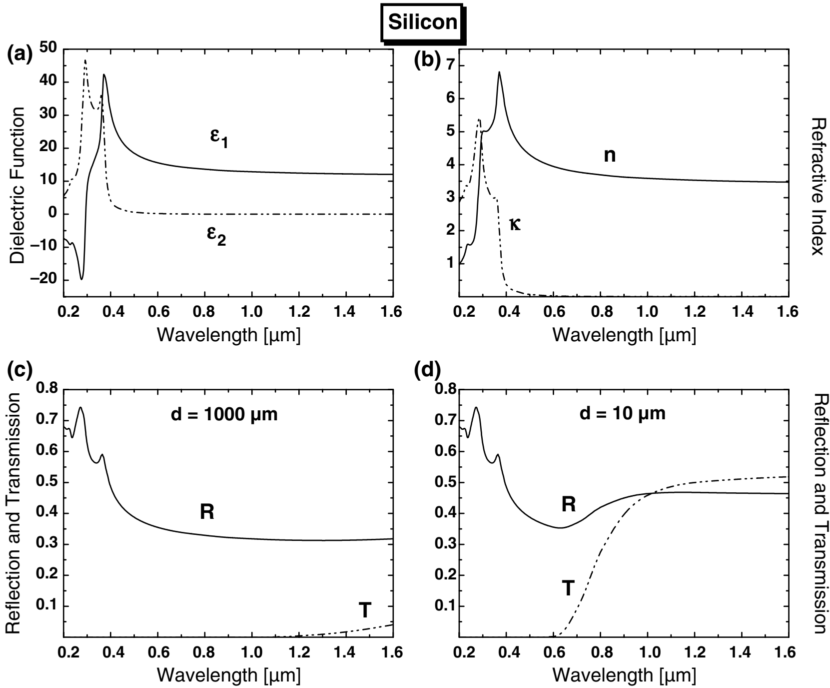


Figure 2.24 (a) Dielectric function of Si. (b) Complex refractive index of Si. (c) Reflectance and transmittance of a plane silicon wafer with $d = 1000\,\mu m$. (d) Reflectance and transmittance of a silicon diaphragm with $d = 10\,\mu m$. This layer was treated incoherently, for which no thickness interference is observable. Optical constants from Ref. [37].

long wavelengths. At wavelengths longer than 1000 nm, this thick wafer slowly becomes transparent. For discussion of the transmittance, it seems therefore appropriate to study thinner wafers. The graph in Figure 2.24d shows the reflectance and transmittance of a layer with only $d = 10\,\mu m$. The differences compared to the thick layer are obvious. The layer increasingly becomes transparent for wavelengths longer than 600 nm. This also influences the reflectance as now light reflected at the rear side of the layer contributes to the reflectance, making it higher than for the thick layer. Note that for calculation of R and T of this thin layer, we assumed the layer to be incoherent, that is, no thickness interference occurs.

The next example is (stochiometric) silicon nitride Si_3N_4 in the wavelength range 0.3–1.2 μm in Figure 2.25. The optical constants are taken from Ref. [38]. As the imaginary part ε_2 of the dielectric function is zero and hence also $\kappa = 0$ for all wavelengths in this wavelength range, this material is transparent. The high refractive index of $n > 2$ results from strong interband transitions in the UV spectral range that can be described by harmonic oscillators (Brendel oscillators). Si_3N_4 is often used as antireflective coating for silicon wafer in photovoltaics or as hard coat for protection of surfaces. The high refractive index also causes a high reflectance of a wafer of 1 mm thickness of $R > 20\%$ (including rear side reflection).

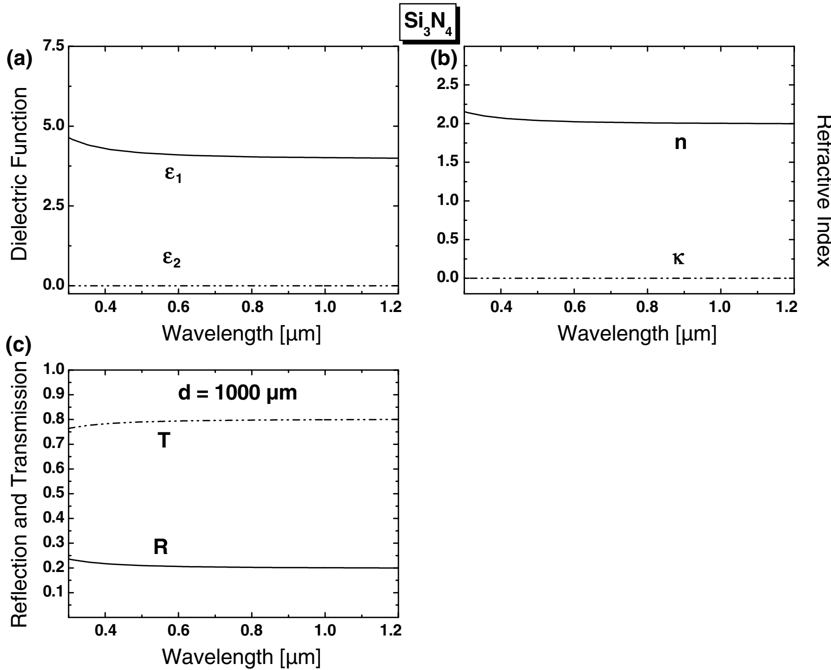


Figure 2.25 (a) Dielectric function of Si_3N_4. (b) Complex refractive index of Si_3N_4. (c) Reflectance and transmittance of a plane silicon nitride wafer with $d = 1000\,\mu m$. Optical constants from Ref. [38].

Tin-doped indium oxide (ITO, $In_2O_3{:}SnO_2$) is of distinct interest because in this material free electrons contribute to a certain electrical conductivity of the material, but thin layers of ITO keep transparent at wavelengths in the visible spectral range. Hence, ITO films are often used in applications where an electrical current is used for switching a process and where in addition the film must be transparent, for example, in flat panel displays or OLEDs (organic light emitting diodes) or as TCO (transparent conductive oxide) in thin-film photovoltaics. Owing to the increased prices of indium, today alternatives to ITO have been developed with aluminum-doped zinc oxide (AZO), fluorine-doped tin oxide (FTO), and antimony-doped tin oxide (ATO). The highest optical transparency and the highest electrical conductivity, however, are still obtained for ITO. The dielectric function of ITO in Figure 2.26a can be approximated in the wavelength range from 0.2 to 1.6 μm by a sum of a harmonic oscillator in the UV and a Drude susceptibility with plasma frequency ω_P lying in the near infrared [39]. In between, in the visible spectral range, the contributions of both to the imaginary part ε_2 of the dielectric constant are small, so that absorption in the visible becomes negligible for thin films below 1 μm thickness. On the other hand, the close lying harmonic oscillator in the UV leads to a high refractive index in the visible spectral range of approximately $n \approx 1.8$–1.9, as can be recognized from Figure 2.26b. For a wafer of $d = 1000\,\mu m$ the absorption in the layer is too large to show transmittance. But for very thin films, as in the graph of Figure 2.26d, the transmittance is in

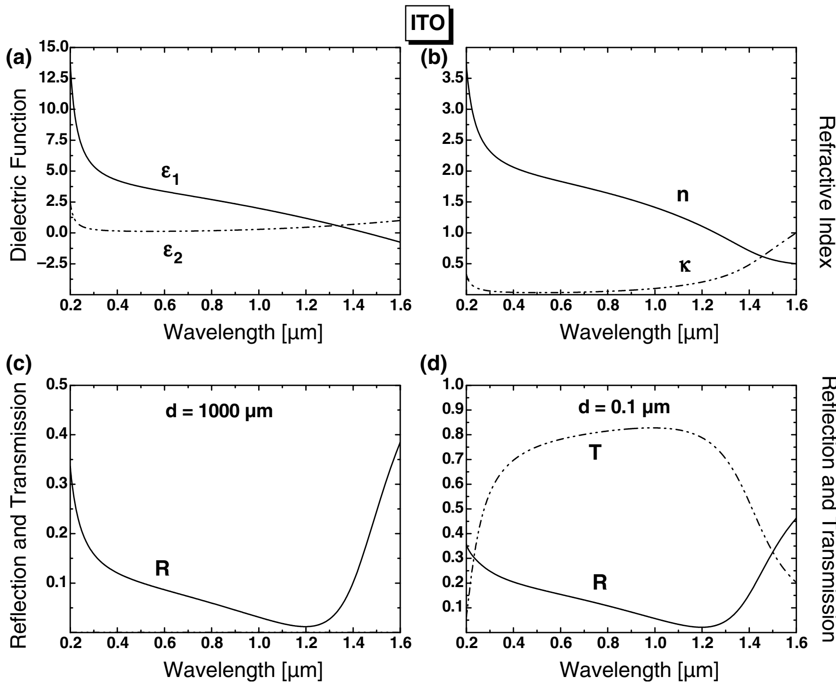


Figure 2.26 (a) Dielectric function of ITO. (b) Complex refractive index of ITO. (c) Reflectance and transmittance of a plane ITO wafer with $d = 1000\,\mu m$. (d) Reflectance and transmittance of a plane ITO wafer with $d = 0.1\,\mu m$. This layer was treated incoherently, for which no thickness interference is observable. Optical constants from Ref. [39].

the order of 70–80%. Note that for calculation of R and T of this thin film, we assumed the layer to be incoherent, that is, no thickness interference occurs.

Another selected example is the metal Al. An outstanding property of this metal is that its dielectric function can be approximated rather well by a Drude dielectric constant for free electrons, and it also exhibits an interband transition around 800 nm wavelength. This can clearly be recognized from Figure 2.27a. The corresponding complex refractive index in Figure 2.27b exhibits a maximum in the real part n and a dip in the imaginary part κ. Nevertheless, the low n and high κ leads to an almost constant high reflectance with $R > 90\%$ except a dip around 825 nm caused by the interband transition. Optical constants were taken from Ref. [38].

Similar to aluminum, the dielectric function of gold (Au) in Figure 2.28 is determined by a Drude term of the free electrons in gold, but in fact only at wavelengths longer than approximately 650 nm. Optical constants were taken from Ref. [38]. At lower wavelengths, the contribution of interband transitions from the 5d electrons to the hybridized 6sp band has an enormous influence on the dielectric function. They cause the reflectance of a gold mirror to be only about 40% at wavelengths below 460 nm. Then, a steep increase follows in the range from 460 to 580 nm to values of $R > 90\%$. In the following spectral range, the reflectance still increases to values $R > 97\%$ in the considered wavelength range. Gold is often used

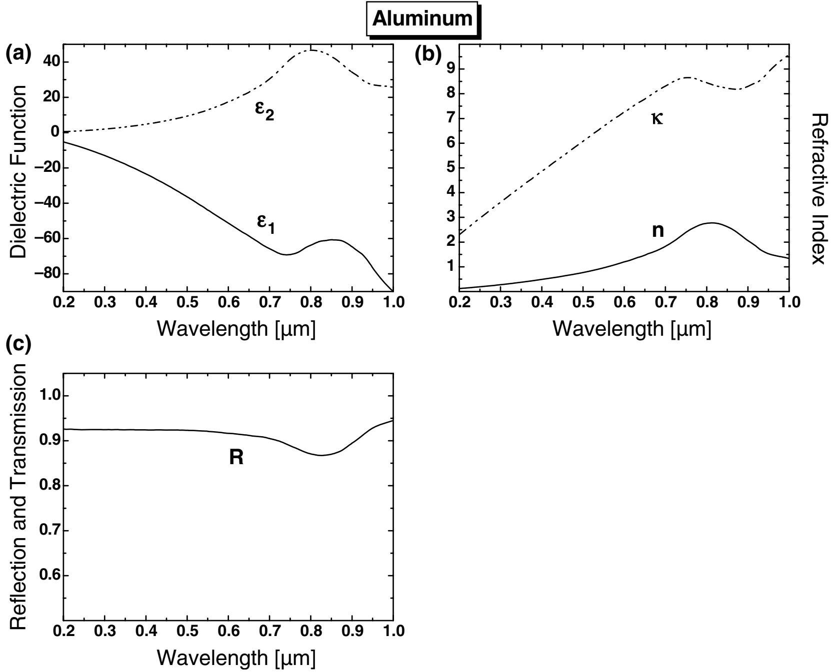


Figure 2.27 (a) Dielectric function of Al. (b) Complex refractive index of Al. (c) Reflectance of an aluminum mirror. Optical constants from Ref. [38].

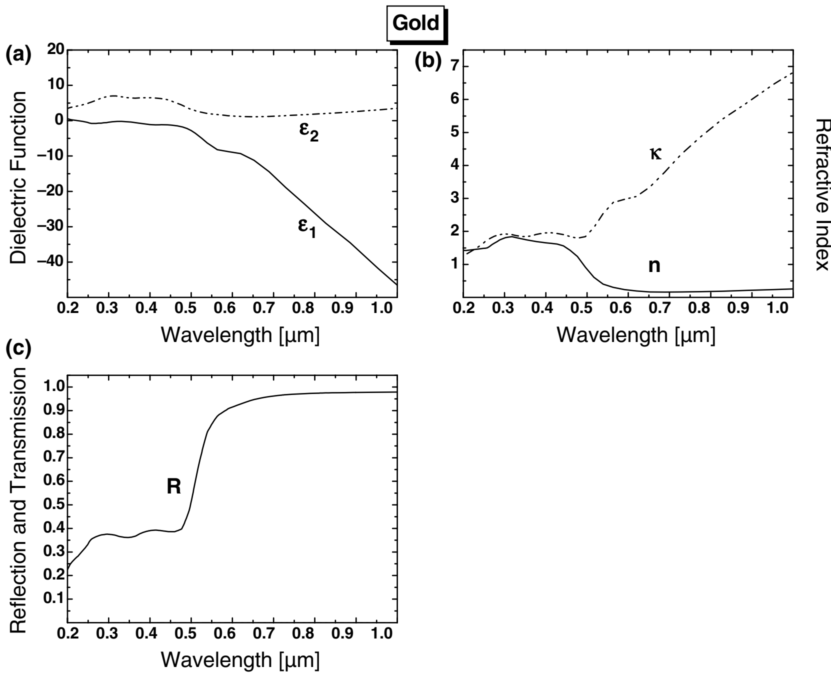


Figure 2.28 (a) Dielectric function of Au. (b) Complex refractive index of Au. (c) Reflectance of a gold mirror. Optical constants from Ref. [38].

as a mirror material in the infrared region because its reflectance amounts always to $R > 97\%$ in the infrared, approaching almost 100%. The steep increase at 520 nm is caused by the interband transitions that reach far into the visible spectral region. The strong difference in the reflectance at right and at left of this band edge leads to the yellow color of a gold mirror. A similar behavior can also be found for titanium nitride TiN. TiN belongs to a larger group of compounds with extreme stability (hardness, chemical inertness, high melting point, and high Young's modulus) caused by the intricate mixture of covalent bonds and ionic contributions. They result in a ceramic behavior with respect to hardness and inertness, and a metallic behavior with high electrical and thermal conductivity, and free-electron-like optical behavior. In these transition metal nitrides, the d-electrons just below the Fermi surface contribute to the free electron gas. Another colored metal is copper where also the dielectric function (not shown here) is determined by a Drude term of the free electrons and the contribution of interband transitions from the 3d electrons to the hybridized 4sp band. The band edge is shifted still further into the visible spectral range (580 nm) for the color of a copper mirror is red.

2.7.2
Kramers–Kronig Analysis of Dielectric Functions

The real and imaginary part of the dielectric function are related to each other by the *Kramers–Kronig relations* [40, 41], sometimes called *dispersion integrals*. These very general relations hold for any frequency-dependent function that connects an output to an input in a linear causal way.

Consider a complex valued function $F(\omega)$ in the plane of complex frequencies $\Omega = \omega + i\delta$. Let ω_0 be some real and positive value. Then, the function

$$G(\omega) = \frac{F(\omega) - F(\infty)}{\omega - \omega_0} \tag{2.114}$$

has to be integrated round a closed contour C in the upper complex half plane (see Figure 2.29). The principal value of the integral along the real axis becomes

$$\wp \int_{-\infty}^{\infty} G(\omega)\mathrm{d}\omega = i\pi(F(\omega_0) - F(\infty)), \tag{2.115}$$

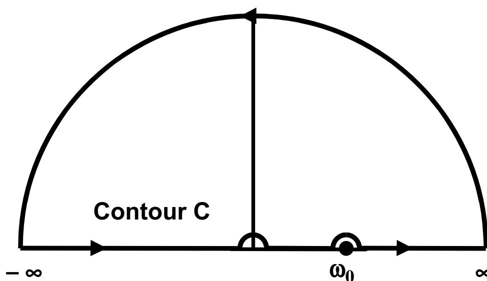


Figure 2.29 Integration path along a closed contour C in the upper complex half plane.

presuming that $F(\omega)$ is regular in the upper half plane. The symbol $\wp$ indicates the principal value of the integral. Decomposition of (2.114) into real and imaginary part yields a connection between the real part and the imaginary part of $F(\omega)$.

Assuming $\lim_{\omega\to\infty}\chi(\omega)=0$, one finds from this for the complex dielectric function

$$\varepsilon_1(\omega)=1+\frac{2}{\pi}\wp\int_0^\infty\frac{\Omega\varepsilon_2(\Omega)}{\Omega^2-\omega^2}\mathrm{d}\Omega, \tag{2.116}$$

$$\varepsilon_2(\omega)=-\frac{2}{\pi}\wp\int_0^\infty\frac{\varepsilon_1(\Omega)-1}{\Omega^2-\omega^2}\mathrm{d}\Omega. \tag{2.117}$$

In principle, the measurement of $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ requires two independent experiments. The Kramers–Kronig relations can replace one of them, making the determination of $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ easier. Many published optical functions result from such a Kramers–Kronig analysis of the reflectance or the absorption coefficient or from combination of electron energy loss experiments with Kramers–Kronig relations. Problems usually arise from the fact that the integrals in (2.116) and (2.117) are extended from 0 to ∞. Experimental values of $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$ are, however, available only for restricted regions. Therefore, the above integrals must be divided into parts, for example, the real part

$$\varepsilon_1(\omega)=1+\frac{2}{\pi}\int_0^{\Omega_1}\frac{\Omega\varepsilon_2(\Omega)}{\Omega^2-\omega^2}\mathrm{d}\Omega+\frac{2}{\pi}\wp\int_{\Omega_1}^{\Omega_2}\frac{\Omega\varepsilon_2(\Omega)}{\Omega^2-\omega^2}\mathrm{d}\Omega+\frac{2}{\pi}\int_{\Omega_2}^{\infty}\frac{\Omega\varepsilon_2(\Omega)}{\Omega^2-\omega^2}\mathrm{d}\Omega. \tag{2.118}$$

While the mid integral can be calculated exactly, the two other integrals must be estimated using reasonable assumptions.

2.7.3 Empiric Formulas for the Refractive Index

Optical constants $n+i\kappa$ or $\varepsilon_1+i\varepsilon_2$ can be modeled using the models presented in Section 2.7.1. Beyond that several empiric formulas for the refractive index $n(\lambda)$ have been developed. Especially glass manufacturers use them still today to parametrize the precise measurements at distinct well-defined wavelengths. In the following, we give an overview of the most common empiric formulas for the refractive index.

One of the most widely used formula is the *Sellmeier formula* [42]:

$$n^2-1=\sum_{j=1}^{N}\frac{A_j\lambda^2}{\lambda^2-B_j}. \tag{2.119}$$

This formula corresponds to the sum over $j=1,\ldots,N$ undamped harmonic oscillators with eigenfrequencies $\omega_j=2\pi c/\sqrt{B_j}$ and oscillator strengths $\omega_{\mathrm{P}j}=\omega_j\sqrt{A_j}$.

This formula is nowadays used by glass manufacturers like SCHOTT AG (since 1992) or OHARA Inc. as three-term Sellmeier formula to approximate the refractive index of their glass. For example, the well-known crown glass N-BK7 from SCHOTT AG has the following parameters: $A_1 = 1.30961212$, $B_1 = 0.0060069867$, $A_2 = 0.231792344$, $B_2 = 0.0200179141$, $A_3 = 1.01046945$, and $B_3 = 103.560653$ for the wavelength measured in microns. The most important modification of this formula is to replace $n^2 - 1$ in (2.119) by $n^2 - {n_0}^2$.

Another widely used formula is the *Schott formula*. It was originally developed by Erich Schott in 1966 at SCHOTT AG. He was the son of the famous Otto Schott who founded the "Schott & Associates Glass Technology Laboratory," together with Ernst Abbe and Carl Zeiss, the owner of an optical workshop, and his son Roderich Zeiss in 1884. This formula was used by SCHOTT AG up to 1992 and then replaced by a three-term Sellmeier formula. The general form is

$$n^2 = \sum_{j=0}^{N} A_j \lambda^{2j} + \sum_{k=1}^{M} B_k \lambda^{-2k}. \tag{2.120}$$

The original Schott formula is obtained for $N = 1$ and $M = 4$. This formula is also known as *Laurent formula* as it is a Laurent series in the wavelength λ. Big glass manufacturers like CORNING Inc., HOYA Inc., HIKARI Inc., or SUMITA Inc. use this formula for their glass. For example, the glass A63-65 (CORNING Inc.) has the parameters $A_0 = 2.12165$, $A_1 = -0.00950349$, $B_1 = 0.008318686$, $B_2 = 0.0003339573$, $B_3 = -2.364166 \times 10^{-5}$, and $B_4 = 9.972179 \times 10^{-7}$ for the wavelength measured in microns.

The third important empiric formula often used for transparent plastics is the *Cauchy formula* developed by the famous mathematician A. L. Cauchy [43, 44]:

$$n = A_n + \frac{B_n}{\lambda^2} + \frac{C_n}{\lambda^4} \qquad \kappa = A_\kappa + \frac{B_\kappa}{\lambda^2} + \frac{C_\kappa}{\lambda^4}. \tag{2.121}$$

The advantage of the Cauchy formula compared to the Sellmeier and Schott formula is that it also considers the imaginary part κ of the complex refractive index. It is therefore suited to also fit the optical constants of absorbing materials. It is often used for photoresists that are absorbing in the UV and at wavelengths in the violet/blue visible spectral region.

In close relation to the Cauchy formula stands an extension of Cauchy's formula, the *exponential Cauchy formula*:

$$n = A_n + \frac{B_n}{\lambda^2} + \frac{C_n}{\lambda^4} \qquad \kappa = A_\kappa \exp\left(B_\kappa \left(\frac{1239.856115}{\lambda} \right) - C_\kappa \right). \tag{2.122}$$

The difference to the Cauchy formula is the use of the Urbach equation [45] to model the absorption index κ by an exponential function. The values of parameters B_κ and C_κ are always such that κ exponentially decreases.

In fact, many optical constants can be approximated by these four empiric formulas. Besides them, many other more or less known empiric formulas and variations of the above formulas exist that are applied to transparent materials. We have all of them summarized in Table 2.2 together with Sellmeier, Schott, Cauchy, and exponential Cauchy.

Table 2.2 Empiric formulas for the complex refractive index.

Name	Equation	Comments
Sellmeier [42]	$n^2-1=\sum_{j=1}^{N}\frac{A_j\lambda^2}{\lambda^2-B_j}$	The original Sellmeier formula is obtained for $N=3$
	$n^2-n_0^2=\sum_{j=1}^{N}\frac{A_j\lambda^2}{\lambda^2-B_j}$	Modification: replace the 1 by a constant refractive index n_0. The most often used with $N=2$
Schott	$n^2=\sum_{j=0}^{N}A_j\lambda^{2j}+\sum_{k=1}^{M}B_k\lambda^{-2k}$	The original Schott formula is obtained for $N=1$ and $M=4$. Formulas with N up to 4 and M up to 5 can be found
Cauchy [43, 43]	$n=A_n+\frac{B_n}{\lambda^2}+\frac{C_n}{\lambda^4},\quad \kappa=A_\kappa+\frac{B_\kappa}{\lambda^2}+\frac{C_\kappa}{\lambda^4}$	Suitable for absorbing materials, as it considers also the imaginary part κ of the complex refractive index
Exponential Cauchy	$n=A_n+\frac{B_n}{\lambda^2}+\frac{C_n}{\lambda^4}$	Suitable for absorbing materials, as it considers also the imaginary part κ of the complex refractive index
	$\kappa=A_\kappa\exp\left(B_\kappa\left(\frac{1.239841875}{\lambda}\right)-C_\kappa\right)$	B_κ and C_κ must be chosen such that κ always decreases exponentially
Conrady [46, 47]	$n=A+\frac{B}{\lambda}+\frac{C}{\lambda^{3.5}}$	
	$n=A+\frac{B}{\lambda^2}+\frac{C}{\lambda^{3.5}}$	The first of these two Conrady formulas is the most commonly used
Kingslake [48]	$n^2=A+\sum_{j=1}^{N}\frac{B_j}{\lambda^2-C_j}$	This formula is known for $N=2$ as Kettler–Drude or Helmholtz–Drude formula
Herzberger [49, 50]	$n=A+\sum_{j=1}^{N}B_j\lambda^{2j}+\sum_{k=1}^{M}\frac{C_k}{\left(\lambda^2-0.028\right)^k}$	The Herzberger formula is intended for use in the infrared. The most commonly used formula are for $N=2$ or 3 and $M=2$. Another modification is obtained by replacing 0.028 with 0.035 or a variable value
Hartmann [51]	$n=A+\frac{B}{(\lambda-C)^N}$	There are three modifications: $N=1$, $N=2$, and $N=1.2$
Modified Sellmeier [52]	$n^2=A+\sum_{j=1}^{N}B_j\lambda^{2j}+\sum_{k=1}^{M}C_k\lambda^{-2k}+\sum_{p=1}^{P}\frac{D_p\lambda^2}{\lambda^2-E_p}+\sum_{q=1}^{Q}\frac{F_p}{\lambda^2-G_p}$	Various combinations of Sellmeier, Schott, and Herzberger type formulas are available, depending on N, M, P, and Q
	$n^2-n_0^2=\frac{A_1\lambda^2}{\lambda^2-B_1}+\frac{A_2}{\lambda^2-B_2}$	Special case: $A=n_0{}^2$, $N=M=0$, $P=Q=1$
Handbook of Optics [52]	$n^2=A+\frac{B}{\lambda^2-C}-D\lambda^2$	These two formulas were introduced in the *Handbook of Optics*. They are obtained from the modified Sellmeier, for $N=1$, $M=P=0$, $Q=1$ and for $N=1$, $M=Q=0$, $P=1$.
	$n^2=A+\frac{B\lambda^2}{\lambda^2-C}-D\lambda^2$	

Wavelength λ in microns.

2.7.4 EMA Models

Consider a composite of at least two nonmixable components: well-separated inclusions with dielectric function ε statistically distributed in a nonabsorbing homogeneous matrix with dielectric function ε_M as illustrated on the left side of Figure 2.30. Its optical response is determined by both the inclusions and the matrix material and is difficult to predict in general. However, if it is possible to replace the inhomogeneous composite by a homogeneous material of one common dielectric function ε_{eff} as illustrated at the right side of Figure 2.30, the reflectance, transmittance, and absorbance of this medium can be calculated as linear response. For that purpose, a model for the dielectric function ε_{eff} of this *effective medium* must be established in dependence on the inclusion properties and the surrounding matrix and the concentration of inclusions in the composite.

The existing effective medium models essentially differ in the way an average is calculated from the dielectric functions ε and ε_M of the embedded inclusions and the matrix, respectively.

The first effective medium concept goes back to Newton (see [53]). It has been modified by Beer [54], Gladstone and Dale [55], Landau and Lifschitz [56], and Looyenga [57]. It is based on simply averaging certain powers of the dielectric functions of the two mixed components, weighted with the filling factor f:

$$\varepsilon_{eff} = f\varepsilon + (1-f)\varepsilon_M \quad \text{Newton} \tag{2.123}$$

$$\varepsilon_{eff}^{1/2} = f\varepsilon^{1/2} + (1-f)\varepsilon_M^{1/2} \quad \text{Beer, Gladstone} \tag{2.124}$$

$$\varepsilon_{eff}^{1/3} = f\varepsilon^{1/3} + (1-f)\varepsilon_M^{1/3} \quad \text{Landau-Lifschitz, Looyenga} \tag{2.125}$$

The topology of the system is here described only by the average volume fraction of the inclusions, often also called *filling factor f*:

$$f = \frac{\sum_k N_k \cdot V_{\text{inclusion}, k}}{V}. \tag{2.126}$$

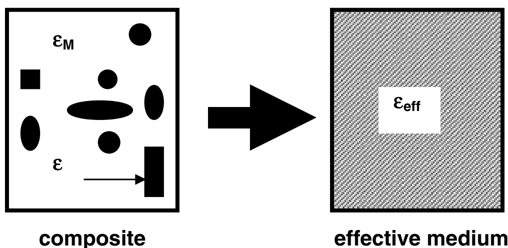


Figure 2.30 Scheme of the effective medium: the realistic composite is replaced by a homogeneous effective medium.

This relation, however, is true only for low concentrations where the inclusions are well separated.
By applying upper and lower bounds, Lichtenecker [53] obtained the *logarithmic mixture rule*.

$$\log(\varepsilon_{\mathrm{eff}}) = f \cdot \log(\varepsilon) + (1-f) \cdot \log(\varepsilon_{\mathrm{M}}). \tag{2.127}$$

The simplest approach to an effective medium that explicitely considers also the shape of the inclusions stemmed from J.C. Maxwell Garnett [58] in 1904 for spherical inclusions

$$\frac{\varepsilon_{\mathrm{eff}} - \varepsilon_{\mathrm{M}}}{\varepsilon_{\mathrm{eff}} + 2\varepsilon_{\mathrm{M}}} = f \frac{\varepsilon - \varepsilon_{\mathrm{M}}}{\varepsilon + 2\varepsilon_{\mathrm{M}}}. \tag{2.128}$$

His approach is well known as Maxwell-Garnett theory, although Maxwell is the third Christian name of James Clerk Maxwell Garnett.

This formula was reformulated to obtain two-dimensional effective medium approaches [59]. Replace the 2 in the denominators by $(D-1)$. $D=3$ is the above result for three-dimensional inclusions of spherical shape. $D=2$ yields the two-dimensional Maxwell-Garnett formula.

In the Bruggeman ansatz [60], the effective dielectric constant is given from

$$f \cdot \frac{\varepsilon - \varepsilon_{\mathrm{eff}}}{\varepsilon + 2\varepsilon_{\mathrm{eff}}} + (1-f) \cdot \frac{\varepsilon_{\mathrm{M}} - \varepsilon_{\mathrm{eff}}}{\varepsilon_{\mathrm{M}} + 2\varepsilon_{\mathrm{eff}}} = 0 \tag{2.129}$$

and in its general form (for more than two media) from

$$\sum_n f_n \cdot \frac{\varepsilon_n - \varepsilon_{\mathrm{eff}}}{\varepsilon_n + 2\varepsilon_{\mathrm{eff}}} = 0. \tag{2.130}$$

Again, for the two-dimensional solution replace the 2 in the denominator of each term by $(D-1)$ with $D=2$ or 3.

Particularly for thin films, Aspnes *et al.* [61, 62] discussed the connection between the microstructure of a heterogeneous thin film and its macroscopic dielectric response ε and developed effective medium theories for two- and three-dimensional isotropic films. The solution is generalized to obtain the Maxwell-Garnett and Bruggeman expressions.

EMA models are used in various applications. Here, we point to three applications where EMA models are successfully used.

1) **Calculation of the refractive index of dielectric particle–dielectric matrix composites**. The similarity of effective medium models to the molar fraction used to calculate the refractive index of mixed glasses makes them attractive for calculation of the refractive index of dielectric nanoparticle–dielectric matrix composites. In Table 2.3, we give results obtained for Al_2O_3, TiO_2, ZrO_2, and SiO_2 nanoparticles embedded in the polymer PMMA using three-dimensional Maxwell-Garnett formula. The idea is to increase or to lower the refractive index of PMMA by oxide nanoparticle inclusions.

Table 2.3 Refractive indices of PMMA with several oxide nanoparticle inclusions.

		n @ 486 nm	**n @ 587 nm**	**n @ 656 nm**
PMMA		1.4977	1.4918	1.4895
Al_2O_3 in PMMA	$f=0.01$	1.5005	1.4945	1.4919
	$f=0.10$	1.5246	1.5185	1.5159
TiO_2 in PMMA	$f=0.01$	1.5094	1.5026	1.4998
	$f=0.10$	1.6152	1.6015	1.5958
ZrO_2 in PMMA	$f=0.01$	1.5042	1.4982	1.4956
	$f=0.10$	1.5629	1.5564	1.5531
SiO_2 in PMMA	$f=0.01$	1.4974	1.4915	1.4889
	$f=0.10$	1.4943	1.4884	1.4859

Actually, the inclusion of the highly refracting Al_2O_3, TiO_2, and ZrO_2 nanoparticles results in an increased refractive index. Vice versa, the inclusion of SiO_2 nanoparticles with a refractive index lower than that of PMMA also lowers the refractive index of the composite. The effects of the inclusions depend on the volume fraction f.

Maxwell-Garnett seems to be helpful here to predict the refractive index of dielectric–dielectric composites. Nevertheless, we want to point to the fact that the nanoparticle inclusions still scatter the light. A well-suited measure for the influence of the scattering is the *haze*. If Al_2O_3, TiO_2, or ZrO_2 nanoparticles with sizes of 10 nm or larger are suspended in a PMMA plate of 1 mm thickness, the mean haze becomes intolerable for filling factors $f \geq 0.01$, according to the ASTM standard D 1003-97 *Test Method for Haze and Luminous Transmittance of Transparent Plastics*.

2) **Rough surfaces**. Roughness of a surface cannot be considered in detail, but only in average. The reason is the stochastic lateral and vertical distribution of the parts of the surface being higher or lower than a constant average level. In surface topography metrology, several roughness parameters have been established that allow characterization of the roughness of a surface. In optical thickness measurements, however, most of these parameters are less useful. Instead, the rough surface is often treated as a homogeneous thin layer with effective refractive index (see Figure 2.31). For the calculation of the effective refractive index EMA models

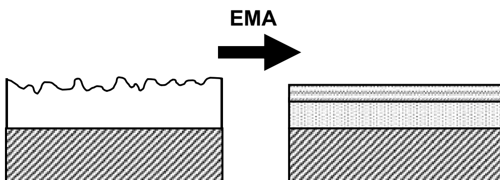


Figure 2.31 Replacement of a film with rough surface by a film plus thin layer with optical constants according to an EMA.

are used with the dielectric functions of the surface material and the material in front of the surface. This method is often used in ellipsometry.

3) **Regular pattern of vias and trenches.** A special case where EMA models are successfully used is the measurement of critical dimensions (depth, width on top and on bottom, and taper angle) of vias and trenches in silicon wafers. For determination of the critical dimensions, infrared reflectometry or ellipsometry (model-based infrared reflectometry = MBIR) is used where the region of the vias or trenches are modeled by one or more homogeneous thin layers with effective refractive index. For calculation of the effective refractive index, two-dimensional EMA models are used. We will discuss this application in more detail in Chapter 7.

At the end of this section, we want to point to general problems arising with EMAs. The most limiting factor of all effective medium approximations is the assumption that scattering by the inclusions can be neglected, for what the inclusions must be very small compared to the wavelength of light. Hence, their application to composites with transparent purely scattering inclusions is questionable. But for strongly absorbing inclusions also, the question may arise whether they can be applied because absorbing inclusions usually scatter the light more strongly than transparent inclusions of the same size. For illustration, we compare exemplarily in Table 2.4 the scattering cross section of small spherical particles with fixed size $2R$ much smaller than the wavelength $\lambda = 514$ nm of the incident light. Compared to a silica sphere, the silver sphere of the same size scatters the light about 25 times stronger. This is the result for a silver sphere off surface plasmon resonance. In resonance, the scattering cross section of the silver sphere is additionally increased by a factor 3–4. As a consequence of this stronger scattering, the failure of effective medium models must become obvious earlier for small metal inclusions than for dielectric inclusions including voids.

Another problem arise with the spectral behavior of the resulting effective dielectric function for high filling factors f. Again, we consider the case of small spherical inclusions.

As long as the filling factor f is less than 0.001, the results for ε_{eff} of the most prominent effective medium approximations from Maxwell-Garnett and Bruggeman and the mixing rule from Looyenga do not differ. This drastically

Table 2.4 Values for the dielectric constant ε and the scattering cross section of very small spherical inclusions at the wavelength $\lambda = 514$ nm.

Material	Dielectric constant ε	$\sigma_{\text{sca}} \propto \left\lvert \frac{\varepsilon - \varepsilon_M}{\varepsilon + 2\varepsilon_M} \right\rvert^2$
Ag	$-10.3 + i \cdot 0.205$	1.853
GaAs	$17.66 + i \cdot 3.207$	0.725
Si	$17.89 + i \cdot 0.525$	0.72
Si_3N_4	$4.15 + i \cdot 0$	0.263
SiO_2	$2.13 + i \cdot 0$	0.075

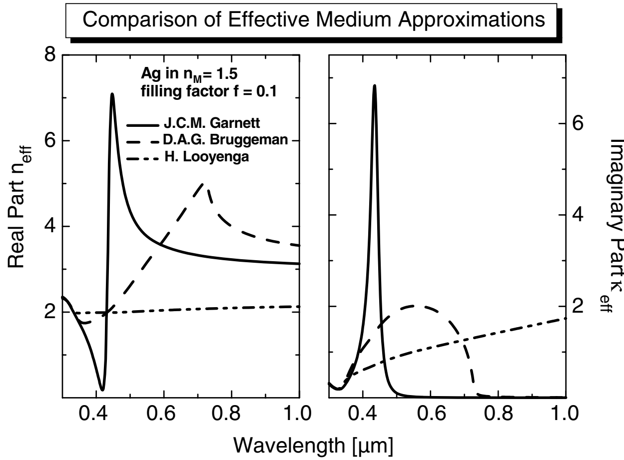


Figure 2.32 Comparison of the refractive index $n_{\mathrm{eff}} + i\kappa_{\mathrm{eff}}$ of Maxwell-Garnett, Bruggeman, and Looyenga for spherical silver inclusions ($f = 0.1$) in a transparent matrix with a constant refractive index $n_{\mathrm{M}} = 1.5$.

changes for higher volume fractions. Figure 2.32 shows a comparison of these three models for spherical silver inclusions ($f = 0.1$) in a transparent matrix with a constant refractive index $n_{\mathrm{M}} = 1.5$ (approximately valid for various crown glasses).

For the Maxwell-Garnett model, the refractive index n_{eff} of this composite exhibits features that indicate the presence of a harmonic oscillator with a resonance wavelength of $\lambda = 435$ nm that is close to the wavelength position of the surface plasmon resonance (SPR) of small Ag-spheres with $2R = 2$ nm in a medium with $n_{\mathrm{M}} = 1.5$ ($\lambda_{\mathrm{SPR}} = 421$ nm). The Bruggeman n_{eff} appears to be composed of a series of close lying harmonic oscillators with different oscillator strengths. However, the curvatures of n_{eff} and κ_{eff} are unexpectedly strange and cannot really be explained by a sum over harmonic oscillators. Completely unexpected is the behavior of the Looyenga n_{eff}. While the real part is almost constant, the imaginary part increases continuously with increasing wavelength. For a more detailed discussion, we refer to Ref. [25].

3
Spectral Reflectance and Transmittance of a Layer Stack

After introducing the basics of electromagnetic wave propagation in Chapter 2, we turn now to the main topics of this book, the determination of the thickness of a single layer or a layer in a layer stack from measurement of the spectral reflectance or transmittance. This is done in three steps. The first step in this chapter consists in the modeling of the spectral reflectance and transmittance of a layer stack. For introduction, we consider in the first section the reflectance and transmittance of a single layer and discuss the influence of a substrate, the absorption of light in the layer, and the partial incoherence caused by thick substrates and roughness of interfaces. Then, we present in Sections 3.2 and 3.3 two models for the wave propagation in a multilayer stack, the *propagating wave model* and the *r-t-ϕ model*. For a review on the theory of the propagation of optical waves in layered media and for learning how to design and to analyze optical devices based on multilayer systems, we refer to Refs [63–65].

The next two steps then follow with the discussion of the commonly used measurement equipment in Chapter 4 and the evaluation methods in Chapter 5.

3.1
Reflectance and Transmittance of a Single Layer

3.1.1
Coherent Superposition of Reflected Light

Consider a single layer of thickness d and complex refractive index $n_1 + i\kappa_1$ on a substrate with complex refractive index $n_2 + i\kappa_2$. A parallel beam of incident light hits the top surface S_1 between the layer and the front medium with real refractive index n_0 under an angle α with respect to the normal to the surface. This is illustrated in Figure 3.1. The actual magnitude and phase of the incident wave at the boundary between the front medium 0 and the layer medium 1 are not of interest and can be set to the value A_0. In the following, we first consider the reflectance of the layer on the substrate in detail. Later in Section 3.1.5, we will also consider the transmittance and

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

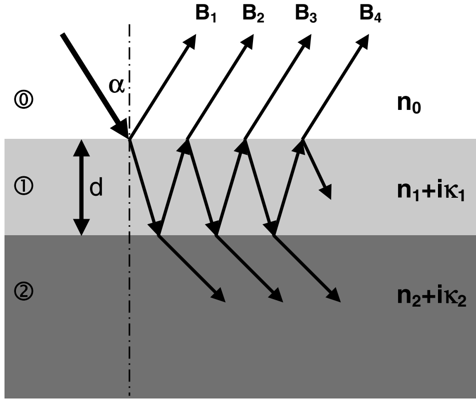


Figure 3.1 Reflection of a single layer on a substrate.

show that it is also possible to get the information on the layer thickness by measuring the transmittance.

A part of the incident wave gets reflected according to reflection law, $B_1 = r_{01}A_0$, with r_{01} being the Fresnel reflection coefficient at surface S_1 when going from medium 0 (front medium) to medium 1 (layer). The remainder $t_{01}A_0$ gets refracted into the layer according to Snell's law. It propagates through the layer and hits the surface S_2 between the layer and the substrate under an angle β with respect to the normal to the surface. Neglecting for the moment the absorption along the path through the layer (i.e., $\kappa_1 = 0$), the phase factor of the wave has changed to $\exp(i(2\pi/\lambda)(n_1(\lambda)\cdot d/\cos(\beta)))$, so that we have at S_2:

$$t_{01}A_0\exp(i(2\pi/\lambda)(n_1(\lambda)\cdot d/\cos(\beta))).$$

Again, a part gets reflected, but with the Fresnel reflection coefficient r_{12} of the interface between layer and substrate (medium 2). This part reaches the surface S_1 after traveling through the layer. On the way from S_2 to S_1, again the phase factor changes by the factor $\exp(i(2\pi/\lambda)(n_1(\lambda)\cdot d/\cos(\beta)))$. At S_1 a part gets reflected with r_{10} and transmitted with t_{10} into the front medium. Note that r_{10} is simply $r_{10} = -r_{01}$, but t_{10} differs from t_{01} because of the traveling direction (either front medium–layer or layer–front medium). The magnitude and phase of the second refracted term is

$$B_2 = t_{10}r_{12}t_{01}A_0\exp\left(i\frac{4\pi}{\lambda}\frac{n_1(\lambda)\cdot d}{\cos(\beta)}\right).$$

If we continue our consideration of the reflected beam, we find that the magnitude and phase of the beams refracted into the front medium can be described by $B_{m+1} = t_{10}r_{12}t_{01}(r_{10}r_{12})^{(m-1)}A_0\exp(i(4\pi/\lambda)(n_1(\lambda)\cdot m\cdot d/\cos(\beta)))$ with m being a positive integer number $m \geq 1$, describing the order of the reflected beam.

In fact, all beams that enter the substrate and get reflected at the rear side of the substrate also contribute to the total amount of reflected light, but they can be neglected here.

We can now deduce the following statements:

1) The way through the layer depends on the angle of incidence because the path through the layer is $d/\cos(\beta)$. Only for $\alpha = 0°$ (normal incidence), the path length is the layer thickness d. For the experimental determination of d with reflectometry, it is therefore recommended to use normal incidence or small apertures with up to maximum $\alpha = 5°$. Then, the polarization dependence of the reflection and transmission can also be neglected. Using ellipsometry this angular dependence must explicitely be considered since ellipsometry works at larger angles of incidence.
2) The magnitudes of the interfering beams B_1 and B_{m+1}, $m \geq 1$ are always different. Even in the case of a free-standing layer (unsupported, $r_{12} = r_{01}$), the magnitude of B_{m+1} differs from the magnitude of B_1 by at least the factor $t_{10}t_{01}$.
3) It is possible to have interferences from B_1 with B_{m+1} with $m \geq 2$. To have a significant signal from these higher orders, the factor $(r_{10} \cdot r_{12})^{(m-1)}$ must be large enough.

In the following, we first consider the two beam interference of B_1 with B_2 for the case of incident angle $\alpha = 0°$. The magnitude of the electric field is the result of the linear superposition of B_1 and B_2. To obtain a measure for the reflectance, we must calculate $|B|^2 = B \times B^*$, where the asterisk denotes the conjugated complex number, and must normalize it to the incident intensity, which is $|A_0|^2$. The result is

$$
\begin{aligned}
R(\lambda, d) = {} & R_{01}(\lambda) + R_{12}(\lambda) \cdot \left(1 - R_{01}^2(\lambda)\right) \\
& + 2\sqrt{R_{01}(\lambda) \cdot R_{12}(\lambda) \cdot \left(1 - R_{01}^2(\lambda)\right)} \cdot \cos\left(\frac{4\pi}{\lambda} n_1(\lambda) \cdot d + \Phi(\lambda)\right),
\end{aligned}
\tag{3.1}
$$

where we used that $r_{10} = -r_{01}$ and that $t_{10}t_{01} = 1 - r_{01}^2$. The phase shift $\Phi(\lambda)$ is here defined as

$$
\Phi(\lambda) = \tan^{-1}\left(\frac{\mathrm{Im}\left(r_{01}^* r_{12}\left(1 - r_{01}^2\right)\right)}{\mathrm{Re}\left(r_{01}^* r_{12}\left(1 - r_{01}^2\right)\right)}\right),
\tag{3.2}
$$

with Re meaning the real part and Im meaning the imaginary part.

This result holds true for all wavelengths λ. That means if we vary the wavelength λ we obtain an oscillating behavior in dependence on λ due to the oscillating cosine term. However, it depends on the ratio (film thickness)/wavelength whether the reflectance exhibits oscillations or not. In the limiting case of very thin layers, only a part of one oscillation or even less can be recognized in the reflectance spectrum. The minimum layer thickness depends on the spectral resolution of the used spectrometer and the exact knowledge on the refractive index of the layer. We will discuss this later in Chapter 5. In the limiting case of a large thickness d, the oscillations become very rapid. Then, again the spectral resolution of the used spectrometer determines the maximum detectable layer thickness. Thicker layers will result in a nonoscillating reflectance spectrum.

As we assumed for the moment a nonabsorbing layer, the phase shift Φ is zero as long as the substrate is also nonabsorbing. For an absorbing substrate, only the value

of the complex reflectivity r_{12} becomes relevant. In Section 3.1.3, we will discuss the magnitude of Φ for partially absorbing and strongly absorbing substrates.

Next, we consider the two beam interference of beam B_1 with a beam B_{m+1}, with $m \geq 2$. When we calculate the superposition of B_1 with B_{m+1} we find that the phase of the oscillating term is $m(4\pi/\lambda)n_1(\lambda) \cdot d + \Phi_m(\lambda)$ and the magnitude of the oscillating term decreases with $\left(\sqrt{R_{10}R_{12}}\right)^{m-1}$. For example, we consider an unsupported silicon wafer of several 10 μm thickness at wavelengths in the near-infrared region around 1300 nm wavelength. The silicon wafer is almost transparent in this wavelength range and the reflectivity $R_{12} = R_{01}$ amounts to $R_{01} = 0.31$. If we set the magnitude of the first-order interference of B_1 with B_2 to 1, the second-order interference $B_1 + B_3$ is on the order of 0.31, the third-order interference is on the order of $(0.31)^2 = 0.0961$, and the fourth-order interference is still on the order of $(0.31)^3 = 0.0298$. This means for such a wafer, up to the fourth-order interference should be recognizable. Indeed, when calculating the reflectance of such a wafer and applying a fast Fourier transform for thickness evaluation, these interferences appear in the power spectrum as equidistant peaks with decreasing magnitude as can be seen in Figure 3.2 (for explanation of this method and the power spectrum we refer here to Chapter 5.1). Also, in experiments on highly polished thin silicon diaphragms, they can actually be observed. In most cases, however, the remaining low absorption in the silicon material and the surface roughness strongly reduce the magnitudes of the higher order interferences. We will discuss these two effects in the following two sections.

Now, we extend our consideration on the multiple beam interference of all reflected beams for normal incidence ($\alpha = 0$). That means that we have to build the sum

$$B = A_0\left(r_{01} + t_{10}r_{12}t_{01}\exp\left(i\frac{4\pi}{\lambda}n_1(\lambda)\cdot d\right)\cdot\sum_{m=0}^{\infty}\left(r_{10}r_{12}\exp\left(i\frac{4\pi}{\lambda}n_1(\lambda)\cdot d\right)\right)^m\right). \quad (3.3)$$

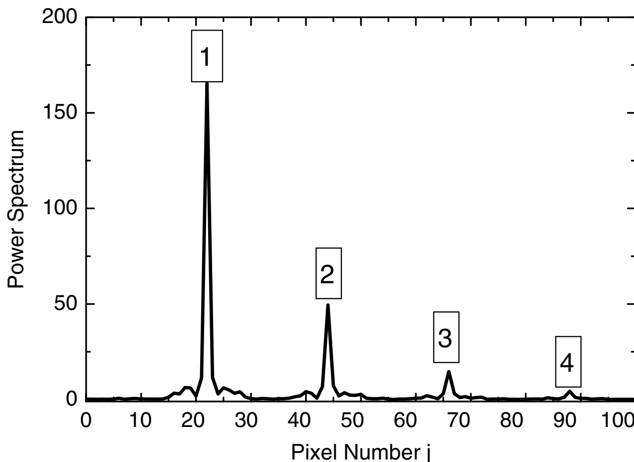


Figure 3.2 Calculated power spectrum of a silicon wafer with $d = 50$ μm, showing interferences up to the fourth order.

As each term z in the infinite sum in (3.3) fulfills $|z| < 1$, the infinite sum tends to $1/(1-z)$. If we additionally use again that $r_{10} = -r_{01}$ and that $t_{10}t_{01} = 1 - r_{01}^2$, we obtain for the magnitude B:

$$B = A_0 \frac{r_{01} + r_{12}\exp(i(4\pi/\lambda)n_1(\lambda)\cdot d)}{1 + r_{01}r_{12}\exp(i(4\pi/\lambda)n_1(\lambda)\cdot d)}. \tag{3.4}$$

The reflectance $R(\lambda, d)$ is finally obtained from $|B|^2/|A_0|^2$ to

$$R(\lambda, d) = \frac{R_{01}(\lambda) + R_{12}(\lambda) + 2\sqrt{R_{01}(\lambda)R_{12}(\lambda)}\cdot\cos((4\pi/\lambda)n_1(\lambda)\cdot d + \Phi(\lambda))}{1 + R_{01}(\lambda)R_{12}(\lambda) + 2\sqrt{R_{01}(\lambda)R_{12}(\lambda)}\cdot\cos((4\pi/\lambda)n_1(\lambda)\cdot d + \Phi(\lambda))}. \tag{3.5}$$

The phase shift $\Phi(\lambda)$ is now

$$\Phi(\lambda) = \tan^{-1}\left(\frac{\mathrm{Im}\left(r_{01}^* r_{12}\right)}{\mathrm{Re}\left(r_{01}^* r_{12}\right)}\right). \tag{3.6}$$

Obviously, the multiple beam interference does not affect the periodicity of the oscillating term. However, it renders the interpretation of the reflectance more difficult than for the two beam interferences in (3.1) since the oscillating term now also enters the denominator. The single film reflectance is determined by the reflectivities at the interface film – front medium R_{01} and at the interface film – substrate R_{12}. A simplification is obtained for a free-standing layer that is not supported by a substrate. Then, $R_{12} = R_{01}$.

To illustrate how the reflectance develops in dependence on the wavelength and the thickness, we consider the following two examples:

1) Unsupported layer, refractive index of fused silica (SiO_2) [38], thickness variable $d = 50, 100, 200, 400$, and 800 nm, surrounding medium (front and rear medium) with $n_0 = 1$, spectral range from 200 to 1500 nm,
2) Unsupported layer, refractive index of silicon nitride (Si_3N_4) [38], thickness variable $d = 50$, 100, 200, 400, and 800 nm, surrounding medium with $n_0 = 1$, spectral range from 200 to 1500 nm.

These examples are of more academic interest since such thin films cannot be prepared without a supporting substrate. However, to exclude the influence of the reflection at the interface film–substrate, we consider here the case of unsupported thin films. The phase shift Φ is $\Phi = 0$ because the materials are nonabsorbing and we consider unsupported layers.

The calculated spectra of thin films of fused silica are summarized in Figure 3.3. All spectra exhibit oscillations according to the oscillating term either in (3.5) or in (3.1). The number of maxima and minima in the spectrum depends on the thickness of the film and increase with increasing thickness d. For the smallest thickness $d = 50$ nm, there cannot be recognized a full oscillation. Absorption does not play a role because SiO_2 is transparent in the considered wavelength range.

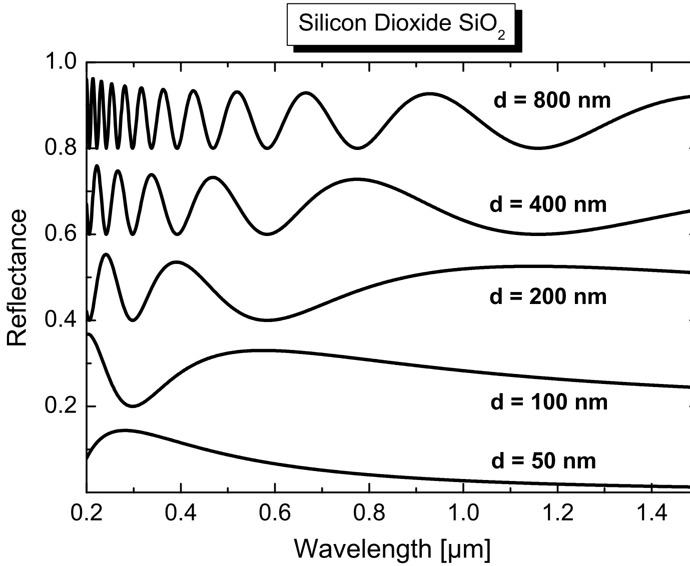


Figure 3.3 Calculated spectra of thin unsupported films of fused silica. The thickness of the film amounts to $d = 50$, 100, 200, 400, and 800 nm. The spectra are shifted along the ordinate for better presentation by adding a multiple of 0.2. Optical constants of SiO_2 from Ref. [38].

The spectra of the second example, thin films of silicon nitride, Si_3N_4, are shown in Figure 3.4. The spectra are rather similar to that of fused silica. However, due to the higher refractive index of Si_3N_4 compared to SiO_2, the oscillations start earlier at

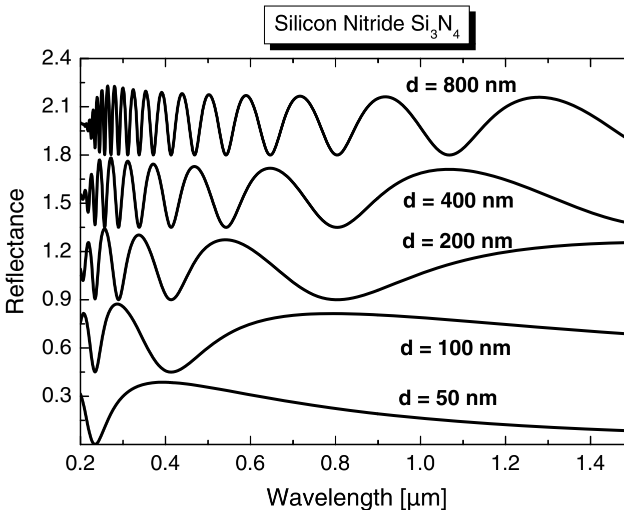


Figure 3.4 Calculated spectra of thin unsupported films of silicon nitride. The thickness of the film amounts to $d = 50$, 100, 200, 400, and 800 nm. The spectra are shifted along the ordinate for better presentation by adding a multiple of 0.45. Optical constants of Si_3N_4 from Ref. [38].

Table 3.1 Exemplaric refractive indices of SiO_2 and Si_3N_4 at different wavelengths (from Ref. [38]).

Wavelength (nm)	Refractive index SiO_2	Refractive index Si_3N_4
300	1.4878	2.1552
500	1.4624	2.0404
700	1.4553	2.0166
900	1.4518	2.0062
1100	1.4492	2.0014

smaller film thickness and more maxima and minima can be observed for the thicker films. The reason is that for the phase change in the oscillating cosine term in (3.5), the ratio (optical path)/wavelength = $(n_1 d)/\lambda$ is important. The higher refractive index of Si_3N_4 then causes the same oscillation as for SiO_2 but with smaller film thickness d. Exemplaric refractive indices for fused silica and silicon nitride are given in Table 3.1. Absorption plays a role only at wavelengths below 300 nm; for longer wavelengths, Si_3N_4 is nonabsorbing in the considered wavelength range. The influence of the absorption can be best recognized for the film thicknesses $d = 400$ and $d = 800$ nm at wavelengths below 250 nm.

3.1.2 Influence of Absorption in the Layer

If the refractive index of the layer is a complex number $n_1 + i\kappa_1$, the magnitudes of the electric and magnetic field of the electromagnetic wave that propagates through the layer decrease. For calculation of the decrease, we consider the phase term $\exp(i\,\mathbf{k}_1 \cdot \mathbf{r})$ in the layer medium with

$$\mathbf{k}_1 = \frac{2\pi}{\lambda}(n_1 + i\kappa_1) \cdot \begin{pmatrix} \sin(\beta) \\ \cos(\beta) \end{pmatrix} \tag{3.7}$$

and $r = (x, z)^T$. As we already know from Snell's law, it is

$$(n_1 + ik_1)\sin(\beta) = n_0 \sin(\alpha) \tag{3.8}$$

and hence, is real valued. On the other hand, the cosine term is a complex number. From this it follows

$$\begin{aligned} \exp(i\,\mathbf{k}_1\mathbf{r}) = \exp\left(i\frac{2\pi}{\lambda}(n_0 \sin(\alpha) \cdot x + (n_1 \operatorname{Re}(\cos(\beta)) - \kappa_1 \operatorname{Im}(\cos(\beta))) \cdot z)\right) \\ \times \exp\left(-\frac{2\pi}{\lambda}(\kappa_1 \operatorname{Re}(\cos(\beta)) + n_1 \operatorname{Im}(\cos(\beta))) \cdot z\right). \end{aligned} \tag{3.9}$$

We recognize that in general (for $\alpha \neq 0°$) the wave in the layer oscillates in x- and z-direction and decreases exponentially only in z-direction. Then, the planes of constant phase are different from the planes of constant amplitude. A simpler

relation is obtained for $\alpha = 0^\circ$, for which also $\beta = 0^\circ$, and hence the exponential decrease is $\exp\left(-\frac{2\pi}{\lambda}\kappa_1 z\right)$.

Remember that the intensity of the wave at position (x, z) is proportional to the square of the magnitude at position (x, z). The difference to the intensity at position (x, z) for $\kappa_1 = 0$ is lost in the medium by exciting vibrations of molecules or lattice vibrations (phonons). This process is called absorption.

The decrease in the magnitude by absorption hence influences the thickness determination by introducing a phase factor $\exp(-(4\pi/\lambda)\kappa_1 \cdot d)$ (for normal incidence). Then, the reflectance of the single film on a substrate in (3.5) changes to

$$R(\lambda, d) = \frac{R_{01}(\lambda) + R_{12}(\lambda) \cdot \exp(-(8\pi/\lambda)\kappa_1 \cdot d) + 2\sqrt{R_{01}(\lambda)R_{12}(\lambda)} \cdot \exp(-(4\pi/\lambda)\kappa_1 \cdot d) \cdot \cos((4\pi/\lambda)n_1(\lambda) \cdot d + \Phi(\lambda))}{1 + R_{01}(\lambda)R_{12}(\lambda) \cdot \exp(-(8\pi/\lambda)\kappa_1 \cdot d) + 2\sqrt{R_{01}(\lambda)R_{12}(\lambda)} \cdot \exp(-(4\pi/\lambda)\kappa_1 \cdot d) \cdot \cos((4\pi/\lambda)n_1(\lambda) \cdot d + \Phi(\lambda))}. \tag{3.10}$$

The important result is that the oscillating term is strongly decreased by the absorption. This may cause problems for the detection of the oscillating term if its amplitude becomes smaller than the signal-to-noise ratio (SNR) of the used detector. Moreover, the phase shift $\Phi(\lambda)$ is nonzero and different for different wavelengths.

In Table 3.2, we give the maximum thickness d_{max} for a signal-to-noise ratio of $SNR = 1000:1$ for different materials at various wavelengths. It is calculated from the condition that the magnitude of the oscillating term is on the size of $1/SNR$:

$$d_{max} = \frac{\lambda}{4\pi\kappa_1(\lambda)} \cdot \log\left(4 \cdot SNR \cdot \sqrt{R_{01}(\lambda) \cdot R_{12}(\lambda)}\right). \tag{3.11}$$

Here, the cosine term is approximated by the factor 2 (the difference between minimum and maximum value of the cosine) and we used the approximation that the

Table 3.2 Maximum thickness d_{max} of an unsupported absorbing layer of different materials for a signal-to-noise ratio $SNR = 1000$.

λ (nm)	Silicon (Si)		Germanium (Ge)		ITO		Aluminum	
	κ	d_{max} (μm)	κ	d_{max} (μm)	κ	d_{max} (μm)	κ	d_{max} (μm)
350	2.9911	0.072	2.704	0.078	0.04496	3.90	4.239	0.054
400	0.3649	0.66	2.215	0.108	0.03774	5.21	4.861	0.054
500	0.06978	4.19	2.399	0.126	0.03287	7.26	6.08	0.054
600	0.02586	13.4	1.3667	0.267	0.03479	8.02	7.26	0.054
700	0.009429	42.6	0.467	0.894	0.04181	7.56	8.31	0.055
800	0.003843	118.3	0.3209	1.47	0.05417	6.43	8.45	0.062
900	0.001847	277.7	0.1851	2.86	0.07309	5.10	8.3	0.070
1100	0.00060759	1028	0.1088	5.93	0.14202	2.64	10.875	0.066
1300	0.00041611	1770	0.07812	9.73	0.30788	1.58	13.147	0.065
1500	0.00035245	2408	0.02568	34.1	0.73963	1.11	15.4	0.064
1700	0.00030487	3153	0.002013	492.3	1.238	0.827	17.567	0.064

Optical constants of the materials from Refs [37–39].

denominator of (3.10) is almost 1 because the other terms in the denominator are drastically reduced by the absorption.

For calculation of the values in the table, we assumed an unsupported layer for what $R_{12} = R_{01}$. The optical constants of the materials were taken from Refs [37–39].

Approximately, the maximum thickness is given by

$$d_{\max} \approx (0.4\text{–}0.75)\frac{\lambda}{\kappa_1(\lambda)} \tag{3.12}$$

for almost all materials and for SNR between 1000 and 5000.

The example of aluminum demonstrates that for a metal the maximum thickness is less than 100 nm. This is true for all metals. Silicon and germanium become the more transparent the longer the wavelength becomes, but the onset of the transparency is shifted to longer wavelengths for germanium because of the higher intrinsic absorption. For optical thickness determination of silicon wafers with typical dimensions between 100 and 800 μm, it seems appropriate to use the near-infrared spectral region above 1100 nm wavelength. The tin-doped indium oxide, ITO, exhibits a window of small absorption in the visible spectral range between 400 and 900 nm, but the absorption is still high as only films of maximum thickness $d_{\max} \approx 5$–8 μm can be measured optically.

In another series of calculations on $d_{\max}$, we assumed the layers supported by a glass of N-BK7 with optical constants from Ref. [66]. This glass is transparent ($\kappa = 0$) for all considered wavelengths between 350 and 1700 nm. The effect of the glass substrate on $d_{\max}$ is a reduction of $d_{\max}$ by 5–8% of the values given in Table 3.2, except for aluminum, where the reduction even amounts to 25–30% of the given values.

Nonabsorbing materials with $\kappa = 0$ are not listed in Table 3.2, but it follows from (3.11) that for these materials the maximum thickness is unlimited. In fact, there does not exist any real material with $\kappa = 0$, but with very low values for κ. For example, the glasses used in the telecommunication for long-distance transmission (mostly SiO_2–GeO_2 mixtures) exhibit a very low signal attenuation by absorption in the frequency bands of telecommunication, so that only after approximately 400–500 km distance the signal must be amplified again. Another example is water in the visible spectral region that is clear and transparent, ranging from 10 to 100 m. On the other hand, the absorption is strong enough to make the deep sea dark (beyond the light scattering by small dissolved particles).

As an example of the influence of the absorption on the reflectance spectrum of a thin film, we consider thin unsupported ITO films of thickness $d = 0.5, 1, 2,$ and 3 μm in the wavelength range from 0.2 to 1.2 μm. From the discussion of $d_{\max}$ and its values listed in Table 3.2, we can expect the biggest influence in the visible spectral region where these films should exhibit oscillations. The calculated spectra of the films are summarized in Figure 3.5.

As expected we can recognize from the spectra that with increasing film thickness d, the oscillations decrease in their magnitude due to the absorption. For $d = 7$ μm (not shown here), the oscillations have decreased beneath the signal-to-noise ratio.

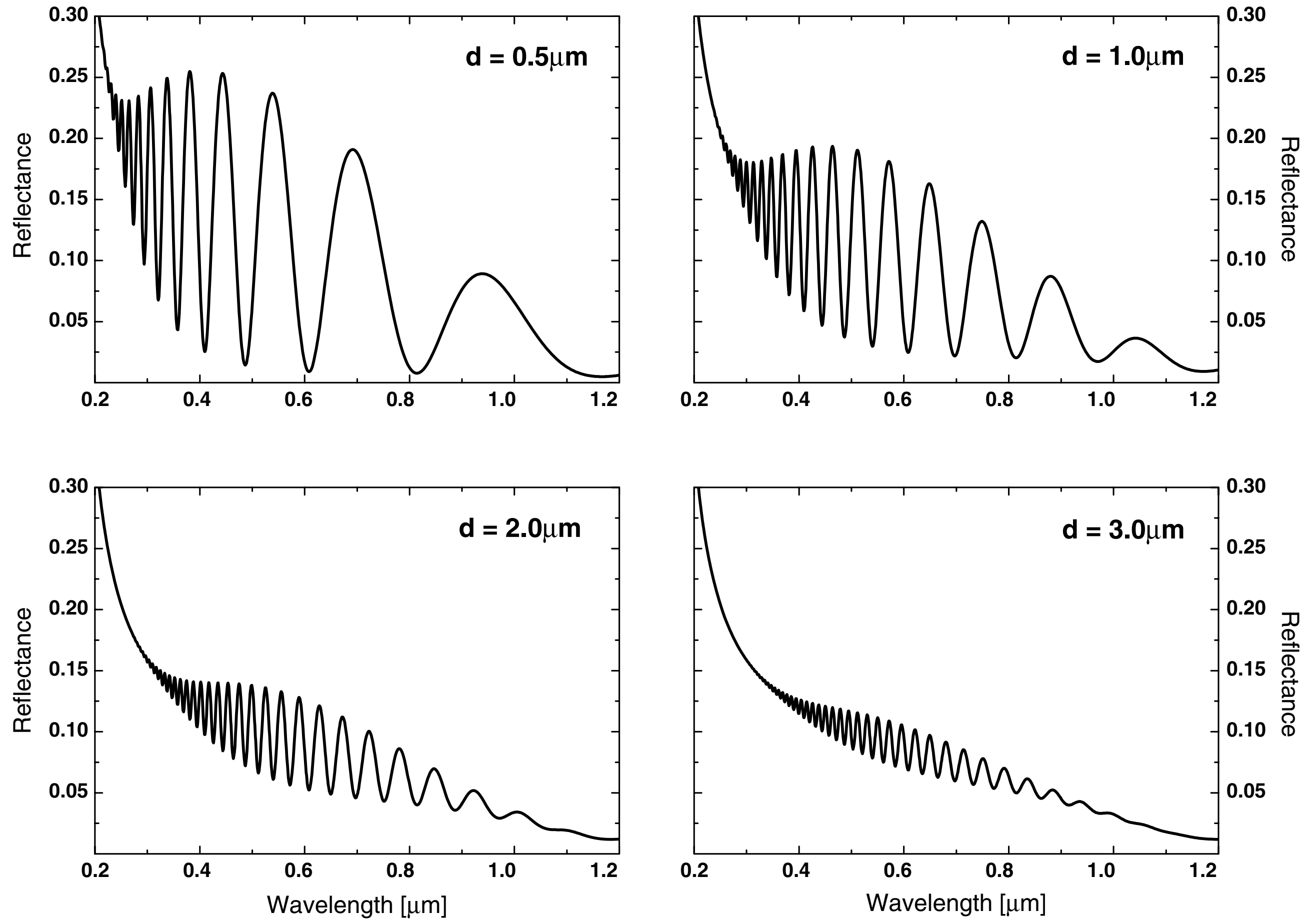


Figure 3.5 Calculated reflectance spectra of unsupported ITO with thickness $d = 0.5$, 1, 2, and 3 μm. Optical constants are taken from Ref. [39].

3.1.3
Partial Incoherence due to Thick Substrates

In general, a substrate of a layer stack can be treated like a layer. As, however, in most cases the substrate is rather thick compared to the other layers, it will result in very rapid oscillations in the reflectance or transmittance spectrum. Nevertheless, short-range disturbances of the surfaces of the substrate (roughness, waviness) strongly reduce these oscillations. Moreover, if the thickness of the substrate is larger than the maximum resolvable layer thickness in the used spectral range, no oscillations can be observed at all. That means the substrate can be treated incoherently, that is, the intensities reflected at each interface of the substrate can be summed up, instead of taking into account superposition of the electromagnetic field components that exhibit a phase shift due to the path through the substrate layer. This must be considered when calculating the reflectance and transmittance of the single layer or even a layer stack according to the models presented in Sections 3.2 and 3.3. If the substrate is absorbing, the absorption in the substrate must also be taken into account. For strong absorption, the intensity of the light reflected at the rear side of the substrate may have vanished. A detailed description of how the incoherent substrate is taken into account when calculating the reflectance and transmittance of a layer is given in Section 3.2. Here, we only illustrate how the reflectance of a single layer on a substrate develops in dependence on the wavelength with the following four examples:

1) Supported layer, refractive index of fused silica (SiO_2) [38], thickness $d = 400$ nm, front medium with $n_0 = 1$, refractive index of the substrate of aluminum oxide (Al_2O_3) [67], spectral range from 200 to 1500 nm.
2) Supported layer, refractive index of fused silica (SiO_2), thickness $d = 400$ nm, front medium with $n_0 = 1$, refractive index of the substrate of the glass N-BK7 [66], spectral range from 200 to 1500 nm.
3) Supported layer, refractive index of fused silica (SiO_2), thickness $d = 400$ nm, front medium with $n_0 = 1$, refractive index of the substrate of silicon (Si) [37], spectral range from 200 to 1500 nm.
4) Supported layer, refractive index of fused silica (SiO_2), thickness $d = 400$ nm, front medium with $n_0 = 1$, refractive index of the substrate of aluminum (Al) [38], spectral range from 200 to 1500 nm.

The calculated spectra are summarized in Figure 3.6. Exemplaric complex refractive indices of the silica film and the substrates are given in Table 3.3.

The influence of the substrate is clearly to recognize. All reflectance spectra of the same film are different in magnitude and curvature. Only the number of oscillations and the positions of the maxima and minima are identical.

The first two substrates – Al_2O_3 and N-BK7 – are transparent in the considered spectral range, but they clearly differ in their refractive index. Aluminum oxide is a high-refractive material for which also the optical contrast to the silica film is high (high reflection at the interface silica–aluminum oxide). Then, the oscillations can clearly be recognized. In contrast, for the crown glass N-BK7, the refractive index is

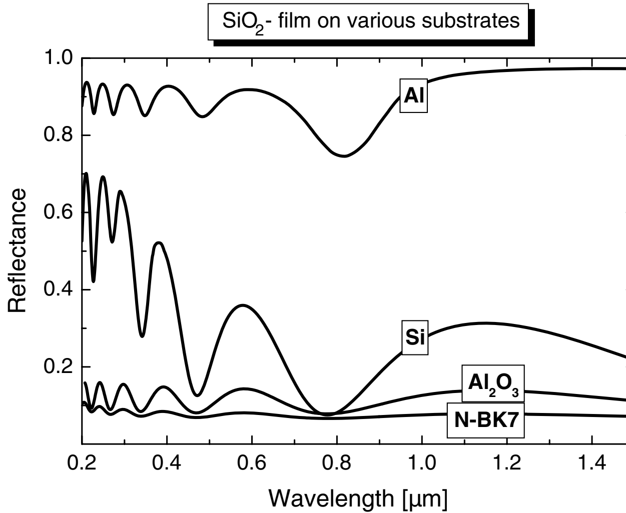


Figure 3.6 Spectra of a supported layer of fused silica (SiO_2) with $d = 400\,nm$ on different substrates: aluminum oxide (Al_2O_3), glass N-BK7, silicon (Si), and aluminum (Al). Optical constants from Refs [37–39, 66, 67].

close to the refractive index of silica. Then, the reflection at the interface silica–N-BK7 is poor and the oscillations become hard to recognize.

For silicon as substrate, we have a substrate that is absorbing at wavelengths lower than 1000 nm and that becomes transparent for longer wavelengths. Nevertheless, the strong intrinsic absorption processes in silicon in the near UV causes a high refractive index between $n = 5$ at short wavelengths and even $n = 3.4$ at long wavelengths. Therefore, the reflection at the interface silica, Si, is high and the oscillations of the silica film are clearly to recognize, even at short wavelengths where the silicon substrate is already strongly absorbing. Absorption of the substrate affects the reflection at the interface silica – Si and the phase shift $\Phi(\lambda)$. This is similar for the film on a metallic substrate where the oscillations are also clearly to recognize. The

Table 3.3 Exemplaric complex refractive indices $n + i\kappa$ of SiO_2 and the substrates Al_2O_3, N-BK7, silicon, and aluminum at different wavelengths [37–39, 66, 67].

λ (nm)	$n + i\kappa$				
	SiO_2	Al_2O_3	N-BK7	Si	Al
300	1.4878	1.8143	1.5528	$5.0004 + i{\cdot}4.1694$	$0.276 + i{\cdot}3.610$
500	1.4624	1.7746	1.5214	$4.2985 + i{\cdot}0.06978$	$0.769 + i{\cdot}6.08$
700	1.4553	1.7635	1.5131	$3.7835 + i{\cdot}0.0094285$	$1.83 + i{\cdot}8.31$
900	1.4518	1.7578	1.509	$3.6246 + i{\cdot}0.001847$	$2.06 + i{\cdot}8.3$
1100	1.4492	1.7536	1.5062	$3.5554 + i{\cdot}0.0006076$	$1.202 + i{\cdot}10.875$

deeper minimum around 825 nm wavelength is caused by the coincidence of an interference minimum of the film and the minimum in the reflectivity of aluminum caused by an interband transition in aluminum.

For further discussion, we reconsider (3.5) for the two cases that the cosine term becomes either maximum, that is, $+1$, or minimum, that is, -1. Drawing a line through all points with the cosine term being maximum defines the *upper envelope* of the reflectance spectrum of a single film on a substrate. Vice versa, the line through all minima defines the *lower envelope* of the reflectance spectrum of a single film on a substrate.

The reflectance values given by the upper envelope follows from (3.5) as

$$R_{\mathrm{up_{env}}}(\lambda) = \left| \frac{n_2(\lambda) - n_0(\lambda)}{n_2(\lambda) + n_0(\lambda)} \right|^2. \tag{3.13}$$

It corresponds to the reflectance of the interface substrate–front medium.

Analogously, the reflectance given by the lower envelope follows from (3.5) as

$$R_{\mathrm{low_{env}}}(\lambda) = \left| \frac{n_2(\lambda) - n_1^2(\lambda)}{n_2(\lambda) + n_1^2(\lambda)} \right|^2. \tag{3.14}$$

It is determined only by the complex refractive indices of the film (n_1) and the substrate (n_2).

It is worth to discuss here the phase shift $\Phi(\lambda)$ in the two cases of absorbing substrates. We have calculated it separately and have normalized it to $(4\pi/\lambda)n_1(\lambda)d$, the main argument in the cosine term in (3.5). The resulting curves are depicted in Figure 3.7.

For silicon as substrate, the phase shift $\Phi(\lambda)$ amounts to maximum 2.2% of the main argument in the cosine term. For increasing wavelength, the absorption in

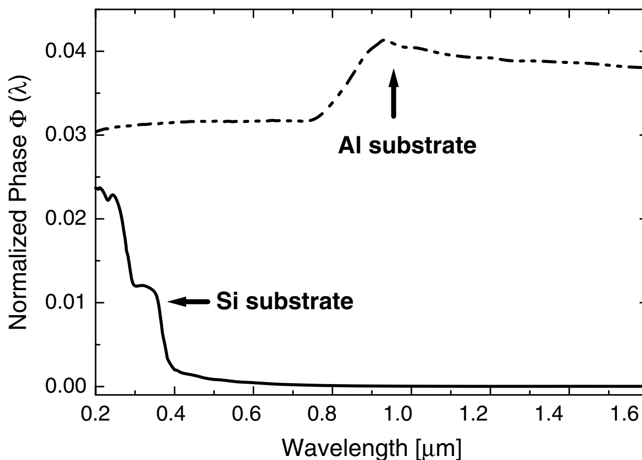


Figure 3.7 Phase shift $\Phi(\lambda)$ for a silica film with thickness $d = 400$ nm on a silicon substrate and an aluminum substrate, normalized to $(4\pi/\lambda)n_1(\lambda) \cdot d$.

silicon tends to zero and hence also the phase shift tends to zero. Unlike silicon, the absorption in aluminum does not vanish for what the phase shift is always nonzero for an aluminum substrate. It amounts between 3 and 4.5% of the main argument of the cosine term and may therefore affect the thickness determination with a fast Fourier algorithm.

3.1.4 Partial Incoherence due to Roughness

The first question that arises when dealing with roughness of a surface is how large must be the imperfection to contribute significantly to scattering of the incident light, or when does a surface appears to be rough? Looking at light scattering by particulate matter (e.g., [25]), one distinguishes between Rayleigh scattering and Mie scattering. Herewith, the scattering particles are classified according to the ratio of particle size to the wavelength of the incident light. At ratios below a certain but not very distinct limit, they satisfy the conditions for Rayleigh scattering and the scattering can be described in a simpler manner than for Mie scattering. A similar condition or size limit is desirable to have also for the scattering at rough surfaces. Below this limit, the effect of the roughness is negligible and the reflection and transmission at the surface or interface can be treated by the regular Fresnel coefficients. In fact, there exists such a condition. If h is the height of the irregularity and γ is the grazing angle, then a measure of the effective surface roughness is $(4\pi h/\lambda)\sin\gamma$ [68]. A reasonable limit to distinguish smooth surfaces from rough surfaces is the Rayleigh criterion for rough surfaces derived from this effective surface roughness

$$h < \frac{\lambda}{p \cdot \sin\gamma} \quad p = 8, 16, \text{ or } 32. \tag{3.15}$$

Macroscopic roughness of an interface causes incident electromagnetic radiation to be scattered at this interface, resulting in the loss of intensity of light collected by the detector (see Figure 3.8). It is thus important to be able to take into account this scattering in calculating the layer reflectance and transmittance.

Periodical roughness can be considered in detail [68, 69]. Random rough surfaces, however, can be considered only in average because parts of the surface are stochastically higher or lower than a constant average level. In surface topography metrology, several roughness parameters have been established that allow characterization of the roughness of a surface. In optical thickness measurements, most of these parameters are, however, less useful.

Therefore, to correct for roughness in thin film reflection modeling, scalar correction factors are used. Assuming that the total reflectivity of an interface can be broken up into a Gaussian distribution centered on the ideal interface, a Debye–Waller factor is derived with its width being parameterized by the rms height R_q of the irregularities of height h. Debye [70] and Waller [71] first understood how statistical irregularities affect the interference when they studied the influence of thermal vibrations on the intensity of X-ray scattering. It was also recognized later that static displacements like the irregularities of a rough surface have a similar effect.

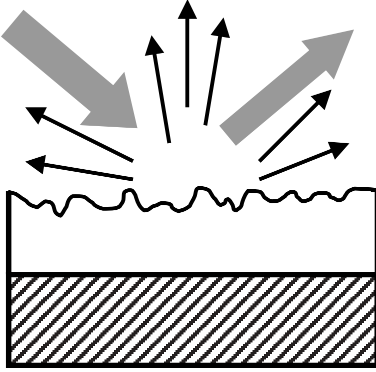


Figure 3.8 Scattering of light at a rough surface with irregularities of height h.

In fact, dealing with thin-film reflection, the Fresnel coefficients are modified by such a Debye–Waller factor. The modified coefficients at the interface between medium $m-1$ and medium m are [72–75]

$$r_{m-1,m} = r_{m-1,m}^{(0)} \cdot \exp\left(-2(Kn_{m-1})^2\right), \tag{3.16}$$

$$r_{m,m-1} = r_{m,m-1}^{(0)} \cdot \exp\left(-2(Kn_m)^2\right), \tag{3.17}$$

$$t_{m-1,m} = t_{m-1,m}^{(0)} \cdot \exp\left(-\frac{1}{2(K(n_m - n_{m-1}))^2}\right), \tag{3.18}$$

$$t_{m,m-1} = t_{m,m-1}^{(0)} \cdot \exp\left(-\frac{1}{2(K(n_{m-1} - n_m))^2}\right) \tag{3.19}$$

for both s- and p-polarization with the abbrevation $K = 2\pi R_q/\lambda$. With this in mind it becomes easy to include roughness effects in the calculation of reflectance and transmittance by using these modified complex Fresnel expressions. Névot and Croce [76] modified the Debye–Waller factor by defining an rms roughness parameter on each side of the interface by R_{q1} and R_{q2}.

We demonstrate in Figure 3.9 the influence of the roughness with calculated spectra of a silica film with $d = 400$ nm on an alumina substrate. The silica–air interface is assumed to be rough with a variable rms height R_q varying from 0 to 50 nm in 10 nm steps. The spectra show that with increasing roughness the magnitude of the oscillation decreases. A value of $R_q = 50$ nm has already strongly reduced the oscillations mainly in the short wavelength range below 500 nm.

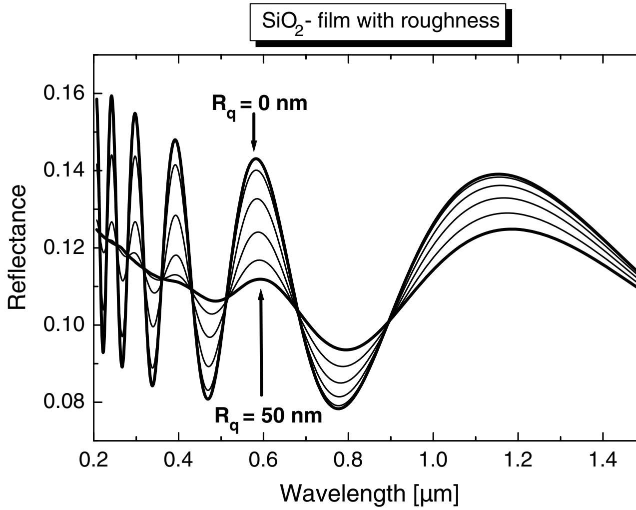


Figure 3.9 Calculated spectra of rough silica films ($d = 400$ nm) on alumina substrate. The rms height R_q of the silica–air interface amounts to $R_q = 0$, 10, 20, 30, 40, and 50 nm. Optical constants are taken from Ref. [38, 67].

3.1.5
Coherent Superposition of Transmitted Light

The favorable method for thickness determination of thin films on a substrate is the measurement of the reflectance. However, if the absorption in the film and the substrate is very small or even negligible, film thickness may also be determined from transmittance. This method is less attractive since the transmittance signal is usually high and is only weakly modulated by the thickness interference.

Without loss of generality, we can assume that the substrate is incoherent, that is, it does not introduce an additional phase factor when the light trespasses the substrate. Moreover, we neglect multiple reflections at and multiple transmissions through the different interfaces. Then, we can directly derive the magnitudes of the transmitted beams C_1, C_2, C_3, and so on in Figure 3.10:

$$C_1 = t_{20} t_{12} t_{01} A_0 \exp\left(i \frac{2\pi}{\lambda} \frac{n_1(\lambda) \cdot d}{\cos(\beta)} \right) \tag{3.20}$$

and

$$C_{m+1} = t_{20} t_{12} (r_{10} r_{12})^m t_{01} A_0 \exp\left(i \frac{2\pi}{\lambda} (2m+1) \frac{n_1(\lambda) \cdot d}{\cos(\beta)} \right) m \geq 1. \tag{3.21}$$

Then, the two beam interference of C_1 with C_2 results in

$$C_1 + C_2 = t_{20} t_{12} t_{01} A_0 \exp\left(i \frac{2\pi}{\lambda} \frac{n_1(\lambda) \cdot d}{\cos(\beta)} \right) \cdot \left(1 + r_{10} r_{12} \exp\left(i \frac{4\pi}{\lambda} \frac{n_1(\lambda) \cdot d}{\cos(\beta)} \right) \right). \tag{3.22}$$

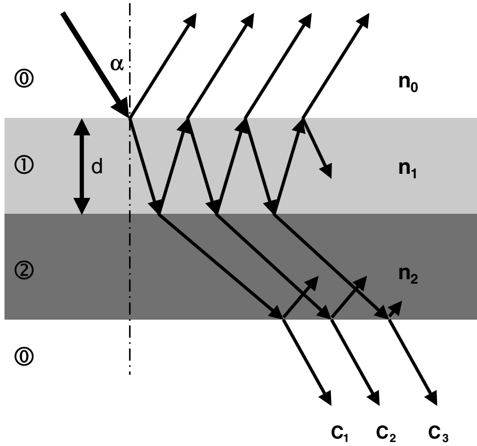


Figure 3.10 Transmission of a single layer on a substrate.

That means the interference of C_1 and C_2 results in a similar oscillatory pattern in the transmittance as the interference of B_1 and B_2 before in the reflectance. However, the magnitude of the transmittance signal is high as it is determined by the product of the transmittivities $T_{20}T_{12}T_{01}$, while the magnitude of the oscillating term is low because it is determined by the product of the reflectivities $R_{10}R_{12}$. Therefore, the oscillations are less resolved in the transmittance than in the reflectance. These results hold true also for the multiple beam interference.

3.2
Propagating Wave Model for a Layer Stack

In this model, the electromagnetic fields **E** and **H** are calculated in each layer, assuming two propagating waves in each layer, one in positive z-direction and one in negative z-direction. For example, in s-polarization the component E_y is

$$E_y(x,z,t) = E^+(x,z,t) + E^-(x,z,t) = \left(E_0^+ \exp(ik_z z) + E_0^- \exp(-ik_z z)\right)\exp(ik_x x - i\omega t). \tag{3.23}$$

Starting at the rear side medium, the fields are transformed into the front medium through all layers 1, . . ., NL. The reflectance R and the transmittance T of the layer stack then follow from the fields in the front and rear medium.

The layer stack is sketched in Figure 3.11. It consists of NL layers with thickness $d(j), j = 1, 2, \ldots,$ NL, and complex refractive index $\tilde{n}(j) = n(j) + i\kappa(j)$. The medium in front of the layer stack has the index 0 and must be nonabsorbing, meaning that $\kappa(0) = 0$ for all wavelengths in the considered wavelength range. In contrast, the rear medium with index NL + 1 can be absorbing.

The incident light may have an angle of incidence α between $\alpha = 0°$ (normal incidence) and $\alpha = 90°$ (grazing incidence). Then, both the light reflected by the stack

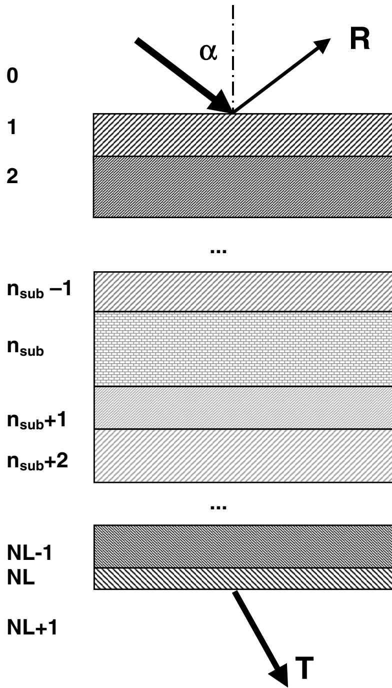


Figure 3.11 Sketch of a layerstack with NL layers including substrates.

and the light transmitted through the stack are different in p-polarization (in-plane) and in s-polarization (perpendicular to the plane of incidence), except for the two cases $\alpha = 0°$ and $\alpha = 90°$, where $R_p = R_s$ and $T_p = T_s$.

The layer stack may consist of one or more layers on a substrate, for example, the layers of an antireflective coating on a glass substrate. As we already have seen from the discussion of a single layer on a substrate, the substrate influences the reflectance and the transmittance of the complete stack. The propagating wave model allows to consider the partial incoherence due to thick substrates. For this, the substrate is considered as a layer for which, however, the waves reflected at or transmitted through the interfaces of the substrate do not interfere. In the schematic drawing of a layer stack in Figure 3.11, the substrate has the layer index n_{sub} and is situated within the layer stack.

In the following, we will first derive in Section 3.2.1 the coherent solution without incoherent substrates and then add the incoherent substrates in Section 3.2.2.

3.2.1
Coherent Reflectance and Transmittance of a Layer Stack

Without any restriction of the generality, we assume in the following the x-axis of the coordinate system going from left to right and the positive z-axis going from top to

bottom. In s-polarization, the electric field has only one nonzero component E_y, while in p-polarization the magnetic field has only one nonzero component H_y.

For the calculation of R and T, we look first at the transformation of the fields through layer j

$$\begin{pmatrix} E_y(z_j + d(j)) \\ Z_0 H_x(z_j + d(j)) \end{pmatrix} = \begin{pmatrix} \cos(k_z(j)d(j)) & -i\frac{\mu_0}{N(j)}\sin(k_z(j)d(j)) \\ -i\frac{N(j)}{\mu_0}\sin(k_z(j)d(j)) & \cos(k_z(j)d(j)) \end{pmatrix} \cdot \begin{pmatrix} E_y(z_j) \\ Z_0 H_x(z_j) \end{pmatrix} \tag{3.24}$$

in s-polarization and

$$\begin{pmatrix} E_x(z_j + d(j)) \\ Z_0 H_y(z_j + d(j)) \end{pmatrix} = \begin{pmatrix} \cos(k_z(j)d(j)) & i\frac{N(j)}{\varepsilon(j)}\sin(k_z(j)d(j)) \\ i\frac{\varepsilon(j)}{N(j)}\sin(k_z(j)d(j)) & \cos(k_z(j)d(j)) \end{pmatrix} \cdot \begin{pmatrix} E_x(z_j) \\ Z_0 H_y(z_j) \end{pmatrix} \tag{3.25}$$

in p-polarization. Here, Z_0 is the vacuum impedance $Z_0 = 376.7\ \Omega$. The z-component k_z of the wavevector in each medium is

$$k_z(j) = \frac{2\pi}{\lambda}\sqrt{\varepsilon(j) - \tilde{n}(0)^2\sin^2(\alpha)} = \frac{2\pi}{\lambda}N(j). \tag{3.26}$$

Here, $N(j)$ is the generalized complex refractive index of layer j and $\varepsilon(j) = \tilde{n}(j)^2$ is the complex dielectric function of layer j. This 2×2 matrix method with two counterpropagating waves is also well known as *Abelès method*, as it was originally developed by Abelès [77, 78]. The matrix $\underline{\underline{\mathbf{M}}}_j$ is the so-called *characteristic matrix* of the jth layer.

The complete transfer matrix $\underline{\underline{\mathbf{T}}}$ of the layer stack is obtained from the multiplication of all transfer matrices $\underline{\underline{\mathbf{M}}}_j$

$$\underline{\underline{\mathbf{T}}} = \prod_{j=0}^{\mathrm{NL}} \underline{\underline{\mathbf{M}}}_j. \tag{3.27}$$

The reflectance R and the transmittance T of the layer stack follow from the ratios of the fields in the front medium and in the rear medium to the incident field. They can be expressed in terms of the matrix elements T_{ij} of the transfer matrix as

$$R_s = \frac{\left(T_{11}^s + (N(\mathrm{NL}+1)/\mu_0)T_{12}^s\right) - \frac{\mu_0}{N(0)}\left(T_{21}^s + (N(\mathrm{NL}+1)/\mu_0)T_{22}^s\right)}{\left(T_{11}^s + (N(\mathrm{NL}+1)/\mu_0)T_{12}^s\right) + \frac{\mu_0}{N(0)}\left(T_{21}^s + (N(\mathrm{NL}+1)/\mu_0)T_{22}^s\right)}, \tag{3.28}$$

$$T_s = \frac{2}{\left(T_{11}^s + (N(\mathrm{NL}+1)/\mu_0)T_{12}^s\right) + (\mu_0/N(0)\left(T_{21}^s + (N(\mathrm{NL}+1)/\mu_0)T_{22}^s\right)}, \tag{3.29}$$

$$R_\mathrm{p} = -\frac{\left(T^\mathrm{p}_{11} - (\varepsilon(\mathrm{NL}+1)/N(\mathrm{NL}+1))T^\mathrm{s}_{12}\right) - (N(0)/\varepsilon(0))\left((\varepsilon(\mathrm{NL}+1)/N(\mathrm{NL}+1))T^\mathrm{p}_{22} - T^\mathrm{p}_{21}\right)}{\left(T^\mathrm{p}_{11} - (\varepsilon(\mathrm{NL}+1)/N(\mathrm{NL}+1))T^\mathrm{s}_{12}\right) + (N(0)/\varepsilon(0))\left((\varepsilon(\mathrm{NL}+1)/N(\mathrm{NL}+1))T^\mathrm{p}_{22} - T^\mathrm{p}_{21}\right)}, \tag{3.30}$$

$$T_\mathrm{p} = -\frac{2(N(0)/\varepsilon(0))(\varepsilon(\mathrm{NL}+1)/N(\mathrm{NL}+1)}{\left(T^\mathrm{p}_{11} - (\varepsilon(\mathrm{NL}+1)/N(\mathrm{NL}+1)T^\mathrm{s}_{12}\right) + (N(0)/\varepsilon(0))\left((\varepsilon(\mathrm{NL}+1)/N(\mathrm{NL}+1)T^\mathrm{p}_{22} - T^\mathrm{p}_{21}\right)}. \tag{3.31}$$

3.2.2
Consideration of Incoherent Substrates

For simplicity, we consider the case that the layerstack has only one substrate. Then, we distinguish a pile 1 of layers with indices 1, 2, . . ., $s-1$ followed by the substrate with index s, followed by a second pile 2 of layers with indices $s+1$, . . ., NL. The calculation of the reflectance of the layerstack is carried out in five steps [79–82]:

Step 1: Calculation of all transfer matrices $\underline{\underline{\mathbf{M}}}_j$.
Step 2: Calculation of the reflectance and transmittance of pile 2 from substrate (s) to the end of pile 2 (b) (s $\rightarrow$ b). The results are the reflection and transmission coefficients r_sb and t_sb for both s- and p-polarization according to (3.28)–(3.31).
Step 3: Calculation of the reflectance and transmittance of pile 1 from the beginning of pile 1 (a) to the substrate (s) (a $\rightarrow$ s). The results are the reflection and transmission coefficients r_as and t_as for both s- and p-polarization according to (3.28)–(3.31).
Step 4: Calculation of the reflectance and transmittance of pile 1 in reverse direction (s $\rightarrow$ a). The results are the reflection and transmission coefficients r_sa and t_sa for both s- and p-polarization according to (3.28)–(3.31).
Step 5: Calculation of the reflectance and transmittance of the complete stack with the substrate.

The results are

$$R = R_\mathrm{as} + \frac{|t_\mathrm{as} t_\mathrm{sa} r_\mathrm{sb}|^2 \cdot \exp(-(8\pi/\lambda)\kappa_\mathrm{sub}(\lambda) \cdot d_\mathrm{sub})}{1 - |r_\mathrm{sa} r_\mathrm{sb}|^2 \cdot \exp(-(8\pi/\lambda)\kappa_\mathrm{sub}(\lambda) \cdot d_\mathrm{sub})}, \tag{3.32}$$

$$T = \frac{|t_\mathrm{as} t_\mathrm{sb}|^2 \cdot \exp(-(4\pi/\lambda)\kappa_\mathrm{sub}(\lambda) \cdot d_\mathrm{sub})}{1 - |r_\mathrm{sa} r_\mathrm{sb}|^2 \cdot \exp(-(8\pi/\lambda)\kappa_\mathrm{sub}(\lambda) \cdot d_\mathrm{sub})} \tag{3.33}$$

for both s- and p-polarization.

3.2.3
Consideration of Surface Roughness

Consideration of surface and interface roughness is in the propagating wave model not as simple to take into account as in the following *r-t-ϕ* model because it always

considers traveling waves in a layer. However, a rough surface or interface can be modeled as a thin homogeneous layer with a refractive index intermediate to the indices of the two adjacent optical media. To predict the refractive index that should be used for the thin layer, several effective media approximations (EMA, see again Section 2.7.4) can be used, as already discussed. Mainly, the Bruggeman approach has been successful in interpreting spectral ellipsometric data of rough surfaces.

Carniglia and Jensen [83] showed that the interface between layer j and layer $j+1$ with rms surface roughness R_q can also be modeled as an additional absorbing thin layer with thickness $d = 2R_q$ and optical constants (n_r, κ_r) given by

$$n_r^2 = \frac{\left(n_j^2 + n_{j+1}^2\right)}{2} \tag{3.34}$$

and

$$\kappa_r = \frac{4\pi R_q}{\lambda} \cdot \frac{\left(n_j - n_{j+1}\right)^2 \cdot \left(n_j + n_{j+1}\right)}{4n_r}. \tag{3.35}$$

3.2.4
r-t-φ Model for a Layer Stack

Analogous to the propagating wave model in Section 3.2, in the *r-t-φ* model a transfer matrix for the layer stack is calculated. This method was introduced in the book of Azzam and Bashara [3]. In contrast to the Abelès method, however, only the electric field is considered in detail. The important parameters for the calculation of the transfer matrix $\underline{\underline{\mathbf{M}}}_j$ are the Fresnel coefficients $r_{m,m-1}$ and $t_{m,m-1}$ at the interface between layer m and layer $m-1$ and the phase term $\exp(i\phi_m)$) with

$$\phi_m = \frac{2\pi d_m}{\lambda}\sqrt{\tilde{n}_m^2 - \tilde{n}_0^2 \cdot \sin^2(\alpha)} = \frac{2\pi N_m(\lambda) \cdot d_m}{\lambda}. \tag{3.36}$$

Then, the characteristic matrix $\underline{\underline{\mathbf{M}}}_j$ of the jth layer is

$$\underline{\underline{\mathbf{M}}}_j = \begin{pmatrix} \frac{1}{\exp(i\phi_m)} & 0 \\ 0 & \exp(i\phi_m) \end{pmatrix} \cdot \begin{pmatrix} \frac{1}{t_{m,m-1}} & \frac{r_{m,m-1}}{t_{m,m-1}} \\ \frac{r_{m,m-1}}{t_{m,m-1}} & \frac{1}{t_{m,m-1}} \end{pmatrix}. \tag{3.37}$$

The transfer matrix $\underline{\underline{\mathbf{T}}}$ follows again from (3.27). The reflectance and transmittance of the layer stack finally follows from (3.28)–(3.31).

So far, the *r-t-φ* model is very similar to the propagating wave model and yields the same results for the reflectance and transmittance of the layer stack. The advantage of this model is that it allows consideration of other layer parameters in a simpler way than the propagation wave model. For example, roughness of a layer interface can simply be included using (3.16)–(3.19) (see also Refs [84, 85]). Beyond this, it is possible to consider anisotropic dielectric functions [86–89] and magnetooptical materials [90–94] in a compact 4×4 matrix formalism. Another approach based on a

2×2 matrix formalism has been derived by Mansuripur [95], Yeh [96, 97], and Cojocaru [98, 99]. Postava *et al.* [100] derived formulas for light reflection and transmission by general anisotropic thick layers based on coherent and incoherent summations of partial reflected and transmitted waves. Visnovsky and Krishnan [101] developed a procedure that allows computation of optical transmission, Faraday rotation, and magnetic circular dichroism in planar layered structures.

4
The Optical Measurement

For determination of the thickness of thin films, optical techniques are usually the preferred methods for measuring thin films because they are accurate, nondestructive, and require little or no sample preparation. The two most common optical measurement types are spectral reflectometry and ellipsometry. The strong advantages of both techniques are the nondestructive contactless character of the measurement and the possibility to control processes in real time. In Sections 4.1 and 4.2, we give an introduction to reflectometric and ellipsometric measurements. Besides these two mainly used measurement techniques, other techniques exist that will be presented in Section 4.3. In Section 4.4, the most important optical components for the optical measurement are discussed.

4.1
Spectral Reflectance and Transmittance Measurement

Quality control during a production process requires a measuring setup that is robust and is adjusted to the ambient conditions at the production site. Moreover, the measured results must be stable and reproducible for a long time. Operation under vacuum conditions additionally requires a simple and safe operation. All these requirements can be easily fulfilled with a setup for spectral reflectance measurement. For that purpose mainly high-valued miniaturized spectrometer modules are used that are extremely robust and have an excellent long-term stability of the wavelength calibration. In addition, one can carry out multiplexed measurements at several sites with only one spectrometer system, for example, when using fiber multiplexers.

For a *reflectometric measurement*, a light source, a fiber, and optionally a measuring head are used to illuminate the sample with unpolarized light. The reflected light gets collected by the measuring head (optionally) and a second fiber that is connected to the detector. The direction of incidence may include an angle α with respect to the normal on the sample, but usually this angle is $\alpha = 0^\circ$ (normal incidence). The principal setup for a reflectometric measurement is sketched in Figure 4.1.

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

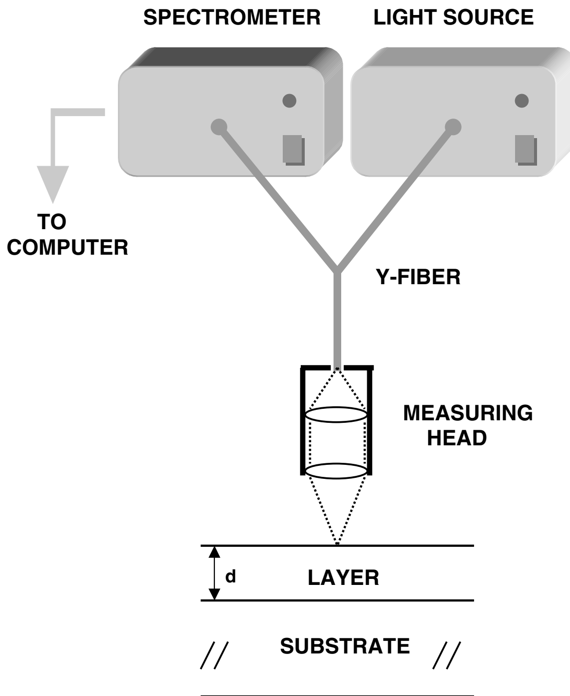


Figure 4.1 Sketch of a reflectometric thin-film measurement.

For reflection measurement typically a so-called Y-fiber is used, where two separate fibers are assembled so that the branch for the illumination and the branch for the detection of the reflected light are merged in a common branch. Hence, illumination and detection are close together in front of the sample.

Spectral reflectance measurement uses the intensity of the light and measures the amount of light reflected from a thin film or a layer stack over a range of wavelengths, with the incident light normal (perpendicular) to the sample surface. This makes spectral reflectance much simpler and less expensive than ellipsometry. On the other hand, it is restricted to less complex layer stack structures. One big advantage over the ellipsometry measurement is the much simpler setup that allows to be extended to measurement at several measuring positions and to be integrated with automated process control and monitoring.

Spectral reflectance can measure the thickness, roughness, and optical constants of a broad range of thin films. However, if there is less than one reflectance oscillation, there is less information available to determine the thickness or other parameters. Vice versa, if one attempts to solve for too many parameters, a unique solution cannot be found because there may be more than one possible combination of parameter values for which the calculated reflectance matches the measured reflectance. Depending upon the layer material and the wavelength range of the measurement, the minimum single layer thickness that can be determined using

spectral reflectance is typically in the order of 20–50 nm. The reason is that a typical reflectometer system cannot accurately measure intensity values to better than 0.1%, and therefore a reflectivity measurement is less sensitive to small changes in ultrathin-film thicknesses.

The measurement is carried out in the following ways:

1) Measurement of the light reflected at a reference sample

$$I_{ref}(\lambda) = R_{ref}(\lambda) \cdot S(\lambda) \cdot \mathrm{Sens}(\lambda), \tag{4.1}$$

where $S(\lambda)$ is the spectral distribution of the light source and $\mathrm{Sens}(\lambda)$ is the spectral sensitivity of the detector.

2) Measurement of the light reflected at the sample

$$I(\lambda) = R_{sample}(\lambda) \cdot S(\lambda) \cdot \mathrm{Sens}(\lambda). \tag{4.2}$$

Calculation of the ratio $I(\lambda)/I_{ref}(\lambda)$ and resolving for the sample reflectivity R_{sample}

$$R_{sample}(\lambda) = \frac{I(\lambda)}{I_{ref}(\lambda)} \cdot R_{ref}(\lambda). \tag{4.3}$$

This method presumes that the reflectivity R_{ref} of the reference sample is known, either as look-up table or by calculation from the optical constants and the thickness of the reference sample. The thickness of the reference sample is necessary for slightly absorbing reference materials because then the reflection on the rear side of the reference sample is correctly taken into account.

Spectral reflectance measurement can be carried out without a mounted measuring head. Then, the common branch of the Y-fiber illuminates the sample and the light gets spread on the sample according to the numerical aperture NA of the fiber (typically NA = 0.22). Vice versa, the detection fiber collects only the reflected light that enters its aperture. Therefore, the size of the detection spot is determined by the core of the detection fiber, as from the widespread illumination spot almost only the light reflected almost perpendicularly can enter the aperture of the detection fiber. Then, the measurement averages over the inhomogenities in the film thickness in the area given by the core diameter of the detection fiber. On the other hand, this allows single measurements in short times (some tens of milliseconds).

Using a measuring head, the incident light usually gets focused on illumination spots of 10–20 μm. As long as the aperture of the focused beam remains less than approximately NA = 0.087 (corresponding to an angle of 5°), there is practically no influence of the refraction at the interface air–layer and no influence of the polarization of the incident wave.

Using a microscope as measuring head, the aperture can be quite large, depending on the used microscope objective. Then, one should take into account the refraction and the polarization when analyzing the measured reflectance. An angle of refraction of $\beta = 15°$ causes an error of 3.5% in the film thickness if not considered. A microscope is usually optimized for the visible spectral range and hence additionally may reduce the useful wavelength range. This automatically also reduces the

maximum thickness that can be determined with the used spectrometer. Although antireflection coatings on the lenses of a microscope objective reduce the reflection, the intensity loss is high as longer measuring times are needed.

Spectral reflectance measurement is not free of errors. Sources of systematic errors are as follows:

- The angle of incidence. This has already been discussed. As long as the angle of incidence is restricted to $\pm5^\circ$ deviation from the normal incidence, the errors are negligible.
- An error source is that the detectors and amplifier circuits are not perfectly linear.
- Optical constants data sets. The analysis of a spectral reflectance measurement with a linear regression requires the knowledge of data sets for the optical constants of a material. The errors made in determination of these data enter the fitting procedure. They have a small but noticeable effect on the results of the fitting procedure.

The *spectral transmittance* measurement requires another setup as shown in Figure 4.2. For transmission measurement, two separate fibers are needed, one for the illumination of the sample and one for the detection of the light that has passed the sample. For transmission measurement, it is necessary that the light beam passing through the sample is a parallel beam. This is achieved using collimators at the end of both fibers.

The transmission measurement determines the amount of light that has passed the sample from the illuminated front side to the rear side. It includes all reflections at both surfaces as well as the absorption in the sample and the scattering in the sample and at the surfaces. At least, a reference measurement at the air gap between the two

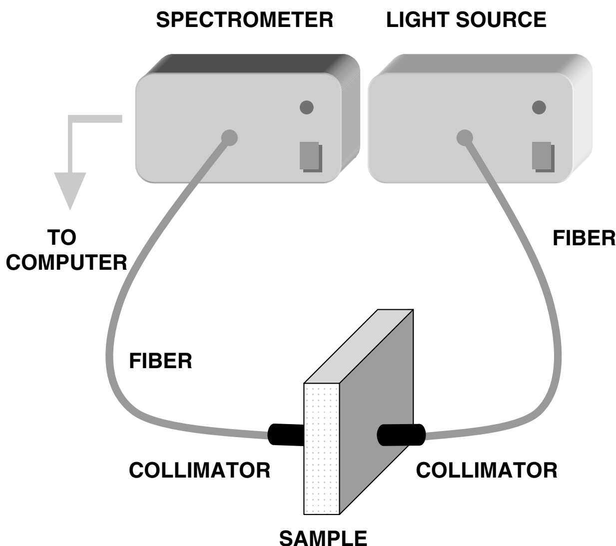


Figure 4.2 Sketch of a thin-film measurement in transmission.

collimators is necessary to take into account the spectral distribution $S(\lambda)$ of the light source and the detector sensitivity $\mathrm{Sens}(\lambda)$

$$I_{\mathrm{ref}}(\lambda) = S(\lambda) \cdot \mathrm{Sens}(\lambda). \tag{4.4}$$

The intensity I_t of the transmitted light is

$$I_t(\lambda) = (1-R_{\mathrm{front}}(\lambda)) \cdot (1-R_{\mathrm{rear}}(\lambda)) \cdot S(\lambda) \cdot \mathrm{Sens}(\lambda) \cdot A(\lambda), \tag{4.5}$$

with A being a factor that describes the intrinsic absorption of the sample and R_{front} and R_{rear} are the reflecitivities of the front and rear surfaces, respectively. The transmittance $T(\lambda)$ then is

$$T(\lambda) = \frac{I_t(\lambda)}{I_{\mathrm{ref}}(\lambda)} = (1-R_{\mathrm{front}}(\lambda)) \cdot (1-R_{\mathrm{rear}}(\lambda)) \cdot A(\lambda). \tag{4.6}$$

It is common practice to use a reference sample instead of the air gap to eliminate the reflectivities in above equation. For example, for a coated glass it is appropriate to use an uncoated glass of the same type as reference sample.

4.2 Ellipsometric Measurement

Ellipsometry is a technique originally developed to study polarized light reflected from solids coated with thin films and from liquid surfaces [2]. However, with the development of electronics and computers, this over 100-year-old technique became relevant because it made it possible to fit the experimental data to the physics-based first-principles equations in a short time. We must distinguish between single wavelength ellipsometry (SWE) and spectral ellipsometry (SE). SWE can measure only two parameters and can be used for simple nonabsorbing single-layer systems. SE, however, can analyze complex structures such as multilayers, interface roughness, inhomogeneous layers, anisotropic layers, and much more.

Azzam and Bashara [3] published in 1977 the book *Ellipsometry and Polarized Light* that has become the key source in ellipsometry. More books covering the theory of ellipsometry, fundamental principles, instrumentation, and applications were published later [1, 4–7, 102]. For a brief introduction to ellipsometry, we advise the reader to refer to the tutorial of J. A. Woollam & Co., Inc. [103].

For an ellipsometric measurement, a light source that provides unpolarized light and a polarizer are used to illuminate the sample with a light beam in an accurately known polarization state. Optionally, an optical retarder is placed between the polarizer and the sample. The direction of incidence includes an angle α with respect to the normal on the sample. Specular reflection of the beam from the sample surface leads to an emergent beam in an elliptical polarization state. It trespasses an analyzer and gets detected by an optical detector. Optionally, an optical retarder is placed between the sample and the analyzer. Usually, the analyzer is rotated (rotating

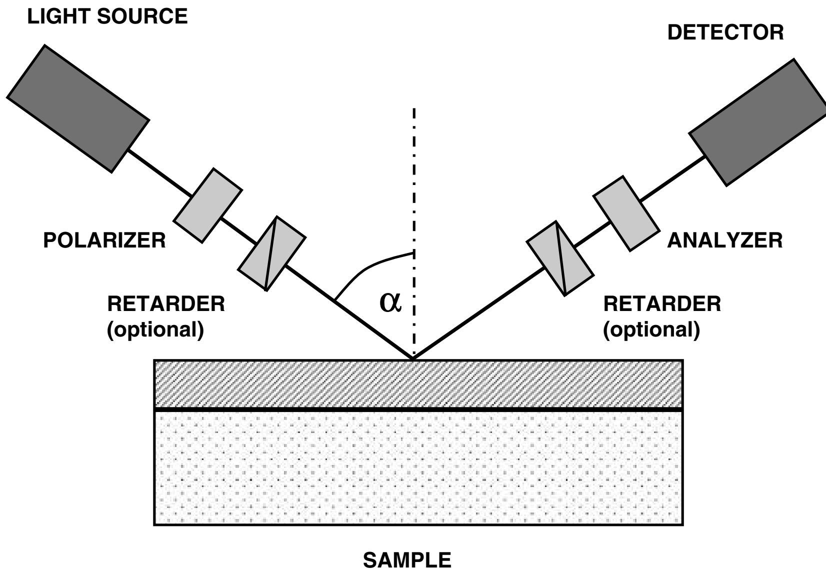


Figure 4.3 Sketch of an ellipsometric thin-film measurement.

analyzer ellipsometry, RAE) to enable at least eight different measurements, but it may also be that the polarizer is rotated (rotating polarizer ellipsometry, RPE) [104]. The principal setup for an ellipsometric measurement is sketched in Figure 4.3.

Ellipsometry measures the complex ratio ϱ of the reflection coefficient of the p-polarized component and of the reflection coefficient of the s-polarized component of the reflected light:

$$\varrho = \frac{r_\mathrm{p}(\alpha)}{r_\mathrm{s}(\alpha)} = \tan(\psi) \cdot \exp(i\Delta), \tag{4.7}$$

where $\tan(\psi)$ is the amplitude ratio and Δ is the phase shift of the p- and s-reflection coefficients according to the Fresnel equations (2.46) and (2.48). They are the *ellipsometric parameters* often given as $\tan(\psi)$ and $\cos(\Delta)$.

Different measurement techniques of the polarization after reflection exist. For them, other components like modulators or compensators can be added. Modern ellipsometer adjust all components automatically and calculate the ellipsometry parameters very fast. The optical and structural information on the sample such as complex dielectric functions and/or thicknesses is calculated from the ellipsometry parameters by fitting ψ and Δ. This is the most critical point of ellipsometry. Ellipsometry is typically used for films with thickness ranging from subnanometers to a few micrometers.

Single-wavelength ellipsometry is very sensitive for films of only a few angstroms thick (1 Å = 0.1 nm) if the optical constants are well known at this wavelength. On the other hand, for determination of optical constants with SWE, it is recommended to have films larger than 50 nm. Spectroscopic ellipsometry measures the change in

polarization of light simultaneously at different wavelengths. The commercially available spectral range covers 150 nm to 33 μm. Measurement of thinner films is achieved by working at shorter wavelengths (e.g., in the UV spectral region). It also allows the determination of the properties (thicknesses and complex refractive indices) of a layer stack. For that purpose, however, physical models for the layer stack (Chapter 3) and the optical constants (Chapter 2) are necessary. Similar to the spectral reflectance measurement, a regression analysis is used to get the best fit of calculated data to measured data. The most commonly used in ellipsometry is the Levenberg–Marquardt algorithm. We will come back to this algorithm in Chapter 5. For a review on modeling of spectroscopic ellipsometric data, we refer to Refs [105, 106]. Spectroscopic ellipsometry applied for the determination of the refractive index of only one single film hopefully results in 100 values of the refractive index and one thickness value, as one measures typically 100 values of Δ and Ψ. Using a parametrization of the refractive index strongly simplifies the task as the number of values to be determined is drastically reduced. Nevertheless, it is recommended not to determine optical constants for films with less than 10 nm thickness.

If the sample is an ideal bulk, the real and imaginary parts of the pseudo-complex dielectric function may be calculated from

$$\varepsilon_1 + i\varepsilon_2 = \sin^2(\psi)\left[1 + \tan^2(\psi)\left(\frac{1-\varrho}{1+\varrho}\right)\right], \tag{4.8}$$

with the knowledge of the incidence angle.

Ellipsometry has been extended to various measurement techniques. The major techniques are phase or polarization modulation ellipsometry (PME) [107–111], two-channel phase modulation ellipsometry [112–114], and variable angle spectroscopic ellipsometry (VASE) [115–117]. PME ellipsometers incorporate one or more oscillating retarders. For the two-channel PME, the analyzing polarizer is replaced with a Wollaston prism. Two independent photodetectors are used, one for each light beam from the Wollaston prism. VASE uses various incident angles and can be applied to characterize composition, crystallinity, roughness, doping concentration, and other material properties associated with a change in optical response in addition to the primary parameters, film thickness and optical constants. All ellipsometer types are also available in a multichannel configuration where an array of detector elements are used to span a broad wavelength range from the UV to the NIR.

Ellipsometric studies are generally carried out in the reflection mode rather than in the transmission mode, requiring either opaque substrates or substrates in which the backreflection is minimized or suppressed by different methods. Bader *et al.* [118, 119] used a transmission and reflection photoellipsometry method to study electrochromic materials and their multilayer systems deposited on thick substrates.

Essentially, any of the automatic spectroscopic ellipsometer designs can provide 100–300 point spectra in Δ and Ψ over the near-infrared to near-ultraviolet wavelength range within half an hour (typical measuring times 1–5 s per measuring point).

Ellipsometry is not free of errors. Sources of systematic errors are as follows:

- **Azimuthal alignment of optical elements**: Each optical element must be aligned to ensure a high-quality ellipsometric measurement. The azimuthal alignment errors are typically on the order of 0.02° for each element. All alignment errors of the several optical elements add up and may result in a significant systematic error in the experiment.
- **The 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 $\approx 0.02°$. In addition, the used light source is not perfectly collimated, so the sample is actually illuminated with a light beam with a distribution of angles of incidence.
- **Calibration**: Spectroscopic ellipsometers that use retarders and compensators must calibrate the amount of phase shift as a function of wavelength. Another error source is that the detectors and amplifier circuits are not perfectly linear.
- **Optical constants data sets**: The analysis of a spectral ellipsometry measurement requires the knowledge of data sets for the optical constants of a material. The errors made in determination of these data enter the fitting procedure. They have a small but noticeable effect on the results of the fitting procedure.

4.3
Other Optical Methods

4.3.1
Prism Coupling

Prism coupling is a method that utilizes the total internal reflection (TIR) of a light beam (mostly a laser beam) at the base of a prism to generate an evanescent wave that couples under certain conditions into the medium put in front of the prism. If the evanescent wave couples into this medium, the internal total reflection gets attenuated, which can be detected with a photodetector that measures the intensity of the totally reflected light. Three setups have been established that use this method.

- Surface plasmon resonance method (SPR)
- Attenuated total reflection method (ATR)
- The prism coupler

The *surface plasmon resonance* method measures the absorption by the medium at a fixed wavelength in dependence on the angle of incidence on the base of the prism. Two methods have been developed, the Otto configuration [120] and the Kretschmann configuration [121]. The Otto configuration is used to examine metal surfaces by coupling the evanescent wave into the metal and exciting a surface plasmon. The surface plasmon is a collective excitation of the free electrons in the metal that travels along the surface of the metal. For a comprehensive discussion of surface plasmons and their excitation, we refer to Raether [122, 123]. The Kretschmann configuration differs in

that way that the base of the prism is coated with a thin metal film (often gold). A surface plasmon gets excited in this film for a certain angle of incidence and travels along the interface metal film–air. If a medium is placed on this metal film, the propagation of the surface plasmon gets disturbed and the excitation of a surface plasmon in the metal film is now most efficient at another angle of incidence.

The *attenuated total reflection* method operates by measuring the changes in the total internal reflection of an infrared beam by coupling of the evanescent wave generated in an optically dense crystal in direct contact with the sample. In regions where the sample material absorbs, the total reflection gets altered by the attenuation of the evanescent wave. This method is well suited for chemical analysis of liquid and solid materials in direct contact with the ATR crystal.

An accurate method for obtaining refractive index *and* thickness for unknown dielectric materials is the use of a *prism coupler* [124–127]. The basic experimental setup is shown schematically in Figure 4.4. By means of a spring-loaded clamp, the film under test is pressed against the base of a coupling prism. It sits on a xy-translation stage that is mounted on a precision rotary table. The laser beam is linearly polarized (TE or TM) and must be of TEM_{00} cross section. The point where the beam strikes the prism base is the coupling spot. At this point, the refractive index n and the thickness d of the film are measured. The optical system is adjusted so that the coupling spot remains practically stationary on the prism base when the rotary table is rotated through the angular range where coupling is possible. TIR of the input laser beam at the prism base leads to a large intensity at the detector and an evanescent wave in front of the prism in the air gap between prism and sample. If the angle gets swept, the reading on the detector fluctuates. For particular angles – the mode angles – the evanescent wave couples into the thin film, resulting in drastically reduced intensity at the detector. By monitoring the angles for which these attenuations of the TIR occur, thickness and index of refraction of the film can be calculated.

For understanding of this method, one must consider the thin film as planar waveguide in which nonradiating eigenmodes can propagate along with the film. This presumes that the film has a higher refractive index than the bordering media. These propagating nonradiating eigenmodes can be described by waves that propagate in the film with effective refractive index n_{eff} but decrease exponentially in their amplitude in the bordering. Taking into account the planar geometry, the wave-number $\eta = (2\pi/\lambda)n_{eff}$ is given by the solution of the equation [128, 129]

$$\tan(\kappa d) = \frac{\kappa(\gamma_1 + \gamma_2)}{\kappa^2 - \gamma_1\gamma_2}, \tag{4.9}$$

with

$$\kappa = \frac{2\pi}{\lambda}\sqrt{n_{waveguide}^2 - n_{eff}^2}, \tag{4.10}$$

$$\gamma_1 = \frac{2\pi}{\lambda}\left(\frac{n_{waveguide}}{n_{substrate}}\right)^{2p} \cdot \sqrt{n_{eff}^2 - n_{substrate}^2}, \tag{4.11}$$

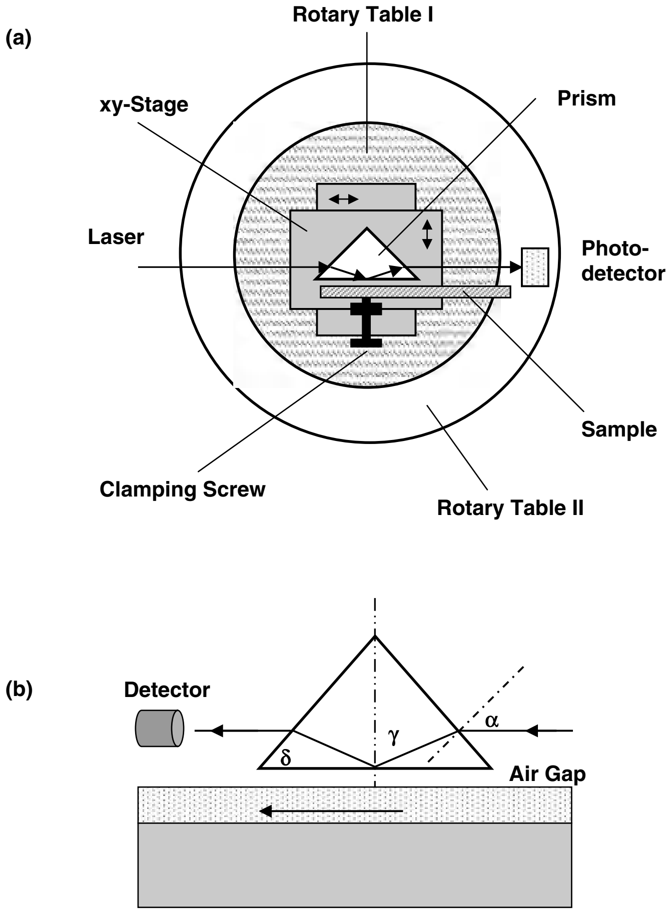


Figure 4.4 Basic prism coupler setup.

and

$$\gamma_2 = \frac{2\pi}{\lambda}\left(\frac{n_{\text{waveguide}}}{n_{\text{air}}}\right)^{2p} \cdot \sqrt{n_{\text{eff}}^2 - n_{\text{air}}^2}. \tag{4.12}$$

The solutions differ for TE polarization ($p = 0$) and TM polarization ($p = 1$). Equation 4.9 can have multiple solutions η_m with $m = 0, 1, 2, \ldots$. If the propagation constant of the evanescent wave

$$\beta = \frac{2\pi}{\lambda} n_{\text{p}} \cdot \sin\theta \tag{4.13}$$

coincides with the propagation constant η_m of any of the waveguide modes, the evanescent waves couple into the waveguide, leading to a drastic reduction of the totally reflected intensity at the prism base. At least, there must be two angles θ where the total reflection gets attenuated to solve for the film refractive index and

the thickness simultaneously. One angle allows only for the determination of the refractive index. If three or more mode lines are observable, a mean and standard deviation calculation is also possible.

According to Figure 4.4b, the angle θ is a function of the angle of incidence α, the prism angle δ, and the refractive index of the prism n_P

$$\theta = \delta + \sin^{-1}\left(\frac{\sin\alpha}{n_P}\right) \tag{4.14}$$

so that it is sufficient to record the intensity of the totally reflected light versus the incident angle α.

Figure 4.5 shows the result of such a measurement on a planar waveguide in glass prepared by ion exchange. The thickness of this waveguide was obtained to be $d = 9.574\,\mu m \pm 0.9\,\mu m$ and the refractive index was obtained to be $n = 1.525 \pm 0.0042$. The used prism of the glass SF10 had a prism angle $\delta = 60°$ and a refractive index $n_P = 1.72314$ at wavelength $\lambda = 632.8\,nm$ (He–Ne laser). The measurement was conducted in TE polarization.

Obviously, the total reflection at the prism base gets attenuated at four incident angles, from which the corresponding effective refractive indices, the film thickness and the film refractive can be calculated. In a model calculation of TE and TM modes of a planar waveguide with these parameters, up to eight TM and TE modes were obtained. Corresponding values for n_{eff} are listed in Table 4.1. It follows that only the low-numbered modes can be clearly resolved experimentally with a prism coupler.

Precise accuracy and resolution values depend on film type, thickness range, and rotary table resolution. Typical values for commercial prism couplers are $\pm$(0.5% + 5 nm) for the thickness accuracy and $\pm$0.3% for the thickness resolution. For the additional refractive index determination, the accuracy is $\pm$0.001 with a resolution of

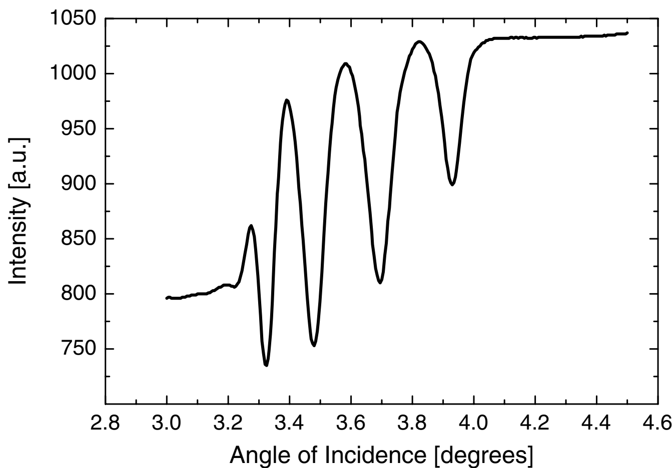


Figure 4.5 Totally reflected light versus incident angle in a prism coupler measurement of a waveguide in glass prepared by ion exchange.

Table 4.1 Effective refractive indices of the TM and TE modes of a planar waveguide with refractive index $n = 1.525$ and thickness $d = 9.574\,\mu m$ on a glass substrate with $n_{sub} = 1.5$ at wavelength $\lambda = 632.8\,nm$ (He–Ne laser).

Mode number	n_{eff} TM modes	n_{eff} TE modes
0	1.524 669 550	1.524 673 604
1	1.523 678 509	1.523 694 649
2	1.522 027 930	1.522 063 964
3	1.519 720 079	1.519 783 388
4	1.516 759 495	1.516 856 737
5	1.513 155 469	1.513 292 049
6	1.508 928 883	1.509 107 747
7	1.504 139 872	1.504 357 091

± 0.0005. With standard prisms, films and bulk materials with refractive index 3.35 and below are measurable. The thickness range is from 0.2 to 150 μm.

4.3.2 Chromatic Thickness Determination

A simple but effective measuring technique has been established in optical profilometry for surface topography and roughness measurement: the measurement with a *chromatic white light sensor*. This sensor utilizes the chromatic aberration of the front lens of a measuring head to measure the distance sensor – surface of the object without moving the measuring head in z-direction. In principle, it is a reflectometer with aberration optics, and is sketched in Figure 4.6. Measuring the light reflected at the specimen in the focal point of the front lens, one obtains a signal only from that wavelength that is focused on the sample. The other wavelengths get spatially distributed since they are not in focus. The determination of the wavelength where the measured reflectance is peaked and a corresponding calibration to the distance allows to determine distances and step heights with nanometer resolution, depending on the measuring range of the used chromatic white light sensor. The measuring range is defined by the difference in the focal lengths for the shortest and the longest wavelength in the useful spectral range. Typical values for the measuring range of chromatic white light sensors used in surface topography measurement range between 100 and 3000 μm.

This technique can also be used for thickness measurement of transparent films since one gets a distance signal from the top surface and the bottom surface of the film. Then, however, one has to consider the refraction. Due to the refraction the focal length gets increased for all longer wavelengths that would have their focus in the layer. This is illustrated in Figure 4.7. Then, the measured distance between the top and the bottom surface must be multiplied by a correction factor CF(n, α) that depends upon the refractive index of the film and the aperture angle α of the

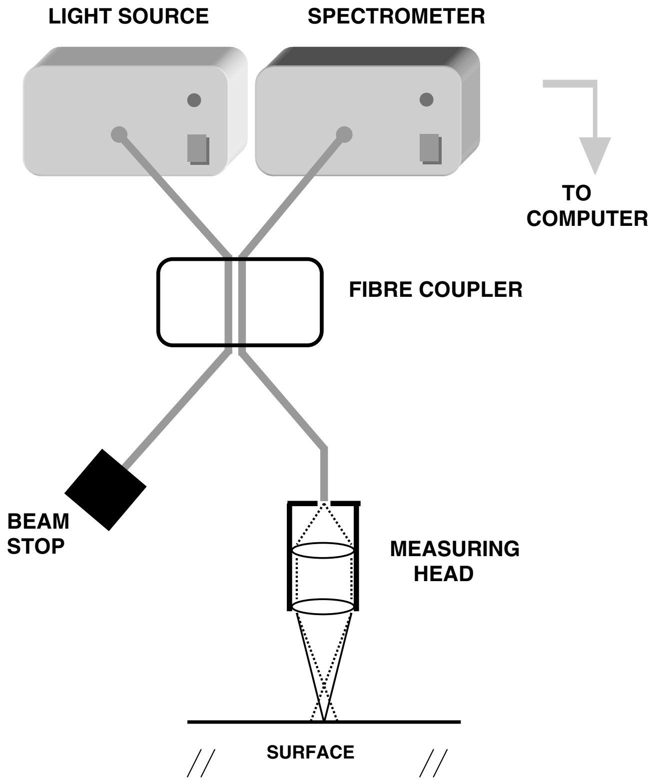


Figure 4.6 Principle of the chromatic white light sensor.

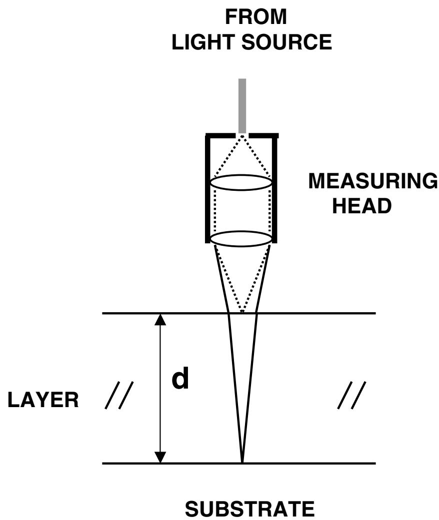


Figure 4.7 Measurement of film thickness with a chromatic white light sensor.

Table 4.2 Correction factor for the thickness determination with a chromatic white light sensor.

n	CF(n, $\alpha = 30°$)	CF(n, $\alpha = 45°$)
1.3	$1.0660 \cdot n$	$1.1867 \cdot n$
1.4	$1.0785 \cdot n$	$1.2206 \cdot n$
1.5	$1.0887 \cdot n$	$1.2472 \cdot n$
1.8	$1.1093 \cdot n$	$1.3005 \cdot n$
2.0	$1.1180 \cdot n$	$1.3229 \cdot n$
2.5	$1.1314 \cdot n$	$1.3565 \cdot n$

measuring head. The correction factor can be derived to

$$\mathrm{CF}(n, \alpha) = \sqrt{\frac{n^2 - \sin^2 \alpha}{1 - \sin^2 \alpha}}. \tag{4.15}$$

The mostly used chromatic white light sensors have an aperture angle of $\alpha = 30°$ or $\alpha = 45°$. In Table 4.2, we summarized the correction factor for these two angles and different refractive indices. As it becomes obvious from this table, the correction factor clearly increases for the larger aperture. The minimal thickness that can be measured with a chromatic white light sensor also depends on this correction factor and is $d_{\mathrm{min}} = 1/15 \times$ measuring range $\times$ CF.

4.4 Components for the Optical Measurement

4.4.1 Light Sources

4.4.1.1 Halogen Lamps

One of the most commonly used light sources in spectral reflectance or transmittance measurement and in ellipsometry are halogen lamps. They emit the broadband spectrum of a tungsten filament, which corresponds to the black body radiation at 2870 K. Figure 4.8 depicts the spectral power densities of black bodies with temperatures up to 6000 K in a semilogarithmic plot versus the wavelength. The spectral density is given by

$$I(\lambda) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{\exp(hc/(\lambda k_{\mathrm{B}} T)) - 1}. \tag{4.16}$$

Here, h is the Planck constant, k_{B} is the Boltzmann constant, c is the vacuum speed of light, and T denotes absolute temperature. With increasing temperature T, the wavelength where the spectral density is maximum λ_{max} shifts according to Wien's law as

$$\lambda_{\mathrm{max}} \cdot T = 2898\ \mu\mathrm{m} \cdot \mathrm{K}. \tag{4.17}$$

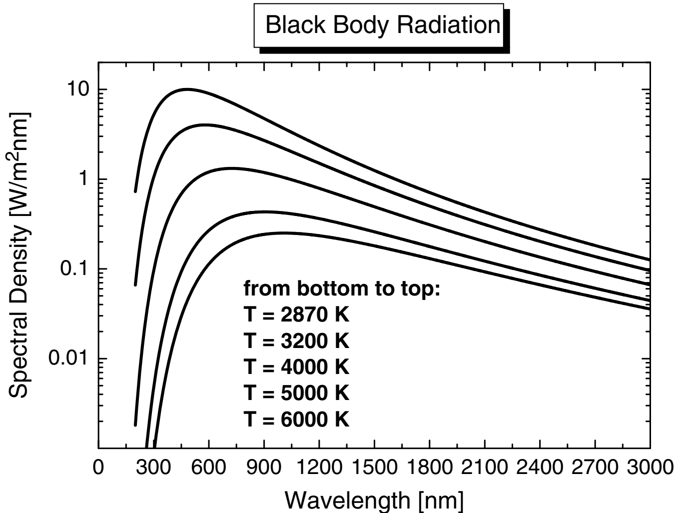


Figure 4.8 Spectra of black bodies of different temperatures.

The filling of the bulb with the halogenes iodine I_2 or bromine Br_2 allows to increase the temperature up to 3200 K because they react with the gaseous tungsten to form WI_6 or WBr_6 near the quartz bulb wall. Then, W is redeposited on the filament that either increases the lifetime of the filament or allows to increase the temperature. Despite this high temperature, the amount of light in the visible spectral region is rather low and only a little amount of blue/violet light or even UV radiation is obtained (see Figure 4.8). The most radiation is emitted in the near-infrared region. The useful spectral range is 360–2500 nm.

Halogen lamps have the advantage of a relatively long lifetime of typically 2000 h. That means approximately 3 months of permanent use. Moreover, they show very low noise and a low drift in the intensity with time. The main disadvantage is that they generate a lot of heat.

The natural black body radiation spectrum of a halogen lamp can be modified using filters or coated reflectors. The special coating on the reflector reduces the amount of long-wavelength radiation so that the fraction of violet and blue light gets stronger. Then, the apparent color temperature is increased compared to the black body radiation but at the cost of the spectral density. An example is given in Figure 4.9 for two halogen lamps with reflector and one lamp without reflector. They all have the same wattage. The decrease in the intensity at long wavelengths is caused by the detector sensitivity that decreases for longer wavelengths. The dip in the spectra at around 940 nm wavelength is caused by absorption in the UV-grade optical fiber.

4.4.1.2 **White Light LED**

White light LEDs are based on short-wavelength LEDs that exhibit strong emission in the blue visible spectral region, covered with a layer of photoluminescent material (phosphor) that extends the emission to longer wavelengths. Figure 4.10 shows the typical spectrum of such an LED covered with the phosphor YAG:Ce. Outside these

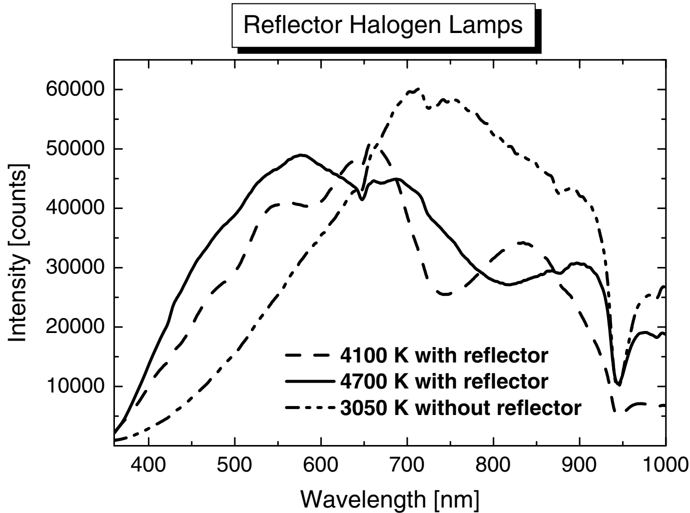


Figure 4.9 Comparison of two halogen lamps with reflector and color temperatures $T = 4100$ K and $T = 4700$ K with a halogen lamp without reflector and color temperature $T = 3050$ K. The wattage of the lamps is identical.

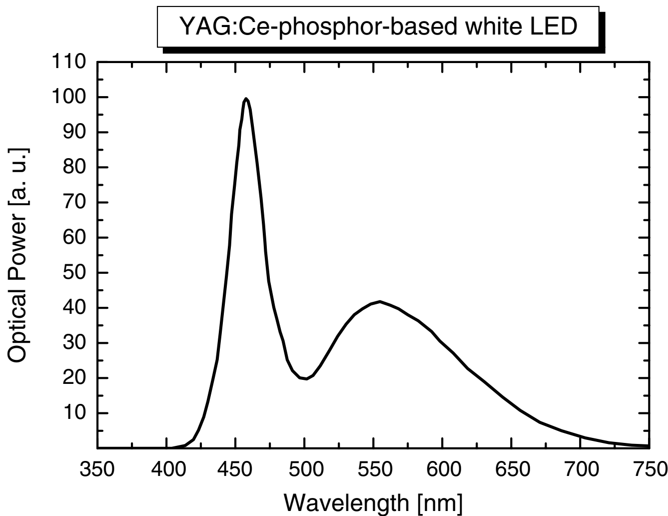


Figure 4.10 Typical spectrum of a white LED based on the phosphor YAG:Ce.

luminescence regions, no further radiation is emitted. Therefore, the use of white light LEDs in the spectrometric determination of layer thickness is restricted. The main advantages are high luminance, easy to couple into a fiber, and an extremely long lifetime of typically 20 000 h or more, or approximately 2.5 years of permanent use.

4.4.1.3 **Superluminescence Diodes**

Superluminescent diodes (SLD) are LEDs that exhibit a high luminance in a narrow wavelength region. As they emit only in a narrow band, they are often used

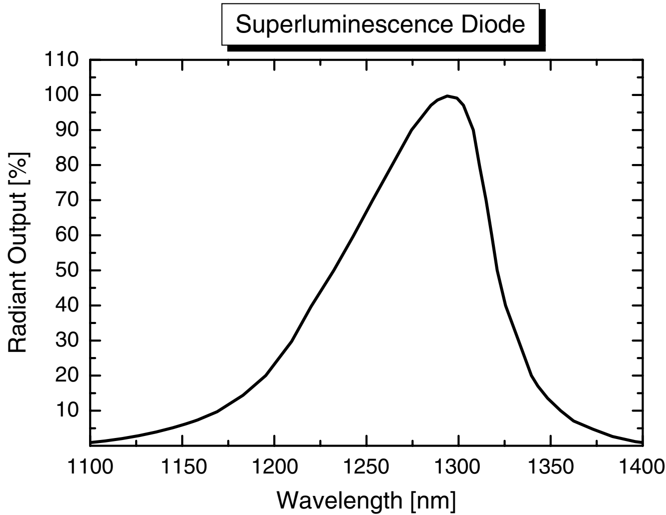


Figure 4.11 Spectrum of an SLD with the center of emission around 1300 nm wavelength.

in high-resolution spectrometers. The main advantages are the same as for white light LEDs: high luminance, easy to couple into a fiber, and an extremely long lifetime of typically 20 000 h or higher, or approximately 2.5 years of permanent use. Figure 4.11 exemplarily shows the emission spectrum of an SLD with its center of emission around 1300 nm wavelength.

4.4.1.4 **Xenon High-Pressure Arc Lamps**

Xenon high-pressure arc lamps are based on arc discharge in a quartz bulb. The filling with noble gas Xenon yields discrete emission lines of Xenon that, however, are extremely broadened as a result of the high pressure of the Xenon gas filling: typically 8 bar (116 psi) when cold and 40 bar (580 psi) when hot. This high pressure requires specific safety precautions when exchanging the bulb and during operation. Moreover, the high voltage needed to ignite the arc discharge also requires specific safety precautions. The biggest advantage is the high luminance over a broad wavelength region from 200 to 2000 nm with a color temperature of 6000 K (see Figure 4.12). The spectral output of Xenon lamp is not altered as the device ages (even toward the end of the lifetime) and occurs instantaneously upon ignition. The lifetime is comparable to that of halogen bulbs, approximately 2000 h or 3 months of permanent use.

Remaining discrete emission lines especially in the range between 900 and 1000 nm restrict the applicability of these lamps. Moreover, the intensity of the emitted light is not constant in time. A stabilization of the arc discharge and the required specific safety precautions make such lamps expensive. When using Xenon high-pressure arc lamps, it is strongly recommended to protect eyes and skin from direct UV irradiation. This kind of lamp is also available as a flash lamp.

4.4.1.5 **Deuterium Lamps**

Deuterium lamps became the workhorse in UV spectrometry due to their stable almost continuous emission of radiation in the wavelength range from 115 to

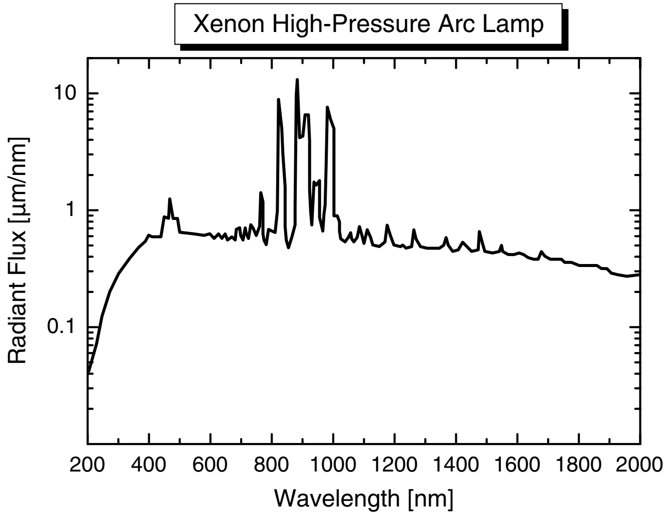


Figure 4.12 Typical spectrum of a Xenon high-pressure arc lamp.

900 nm. Also deuterium lamps are based on an arc discharge in a bulb filled with deuterium at low pressure of about 10 mbar. Deuterium D_2 is a heavy hydrogen, with an additional neutron in the atomic core of hydrogen. This additional neutron doubles the mass of the hydrogen atom. Deuterium lamps emit in the range from 115 to 900 nm, but have a useful spectral range between 185 and 400 nm with a continuous emission of UV radiation. In the visible spectral region, its emission is low. A typical emission spectrum from 200 to 400 nm is shown in Figure 4.13.

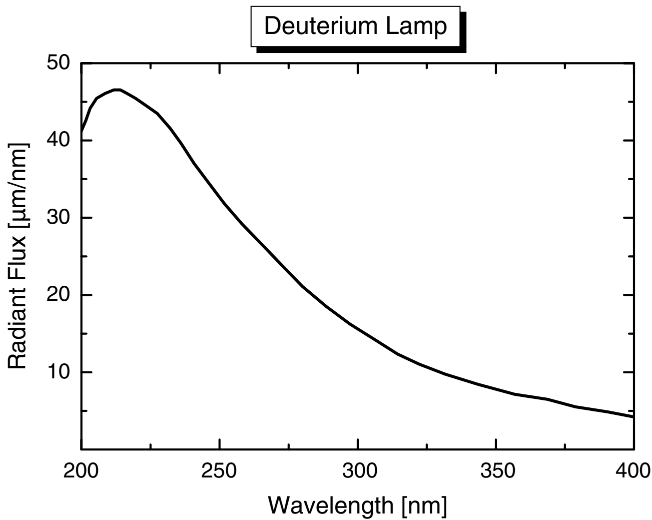


Figure 4.13 Typical UV spectrum of a deuterium lamp.

The lifetime is comparable with the lifetime of halogen bulbs (approximately 2000 h or 3 months of permanent use). The longest lifetime of any deuterium lamps has so far been reported for the L2-4000 series lamps of Hamamatsu Photonics K.K. with 4000 h. When using deuterium lamps, it is strongly recommended to protect eyes and skin from direct irradiation.

A new technique for commercially used Deuterium lamps is the RF-induced discharge. It allows a compact setup for mobile applications and for applications with power consumption of less than 5 W.

4.4.2 Optical Components

4.4.2.1 Lenses and Mirrors

Lenses are important components for measuring heads in reflection and transmission spectrometers. The materials of lenses are usually various glasses, fused silica, and calcium fluoride CaF_2, depending on the spectral range where they are used. While fused silica and CaF_2 are used for the ultraviolet region, in the visible and near infrared mainly various glasses are used. The glasses can roughly be divided into flint glasses and crown glasses. The difference is the refractive index. The higher refractive index of the flint glasses is caused by heavy metals that are introduced as oxides of heavy metals (Ba_2O_3, As_2O_3, PbO_2, etc.). On the other hand, they also cause absorption and, therefore, reduce the transmittance at wavelengths in the blue violet spectral region.

The German Ernst Abbe introduced a number ν, called Abbe number, for further characterization of the glasses. The Abbe number contains information on the refractive index n_d at the yellow helium line (587.1562 nm) and on the dispersion as difference between the refractive indices n_F at the red hydrogen line (486.1327 nm) and n_C at the cyan hydrogen line (656.2725 nm):

$$\nu_d = \frac{n_d - 1}{n_F - n_C}. \tag{4.18}$$

With this Abbe number, one can distinguish low refractive crown glasses ($n_d < 1.6$) with $\nu_d \geq 50$ and high refractive flint glasses ($n_d > 1.6$) with $\nu_d < 50$. Table 4.3 gives an example of a typical crown glass and a flint glass.

Table 4.3 Properties of a typical crown glass and flint glass.

Crown glass	Flint glass
Low refractive index $n_d < 1.6$	High refractive index $n_d > 1.6$
High Abbe number $\nu_d \geq 50$	Low Abbe number $\nu_d < 50$
Example: N-BK7 (SCHOTT AG)	Example: SF6 (SCHOTT AG)
$n_d = 1.5168$	$n_d = 1.80518$
$\nu_d = 64.17$	$\nu_d = 25.43$

Table 4.4 Refractive index of the crown glass N-BK7 (SCHOTT AG) at several wavelengths [66].

Wavelength (nm)	Refractive index	Wavelength (nm)	Refractive index
312.7	1.54 862	632.8	1.51 509
334.1	1.54 272	643.9	1.51 472
365.0	1.53 627	656.3	1.51 432
404.7	1.53 024	706.5	1.51 289
435.8	1.52 668	852.1	1.50 980
480.0	1.52 283	1014	1.50 731
486.1	1.52 238	1060	1.50 669
546.1	1.51 872	1529.6	1.50 091
587.6	1.51 680	1970.1	1.49 495
589.3	1.51 673	2325.4	1.48 921

Although transparent in the visible and near-infrared region, and for fused silica and also CaF_2 in the ultraviolet region down to 175 nm, all these materials exhibit dispersion, that is, a wavelength dependence of the refractive index. For example, the refractive index of N-BK7 at different wavelengths is given in Table 4.4. The refractive index decreases with increasing wavelength.

Illuminating a focusing lens with a parallel white light beam, the focal length is different for different colors. The reason is that due to the higher refractive index for blue light it gets refracted more than red light. Therefore, the focal point lies closer to the lens for blue light than for the red light. This *chromatic aberration* must be avoided when using lenses in spectrometers or in measuring heads. It can be strongly reduced by combining a focusing lens of crown glass with a dispersing lens of flint glass. The focal length of this achromatic lens is almost constant over a wide spectral range as can be seen from Figure 4.14 for a standard lens and an achromatic lens with focal lens $f = 50$ mm.

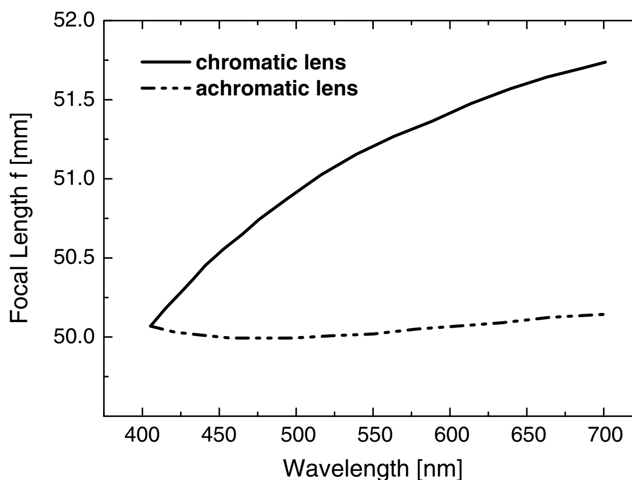


Figure 4.14 Chromatic aberration of a standard lens and an achromatic lens with focal length $f = 50$ mm.

When using mirrors as imaging elements, the chromatic aberration is absent. However, spherical aberrations and other errors must be corrected (correction plates), and they must be corrected for lenses (aspheric shapes and gradients).

4.4.2.2 **Polarizers and Analyzers**

A polarizer is a device that converts an electromagnetic wave of undefined or mixed polarization into a beam with well-defined polarization. The common types of polarizers are linear polarizers and circular polarizers.

Linear polarizers exploit the properties of birefringent crystals such as quartz or calcite crystals. In these crystals, a beam of unpolarized light incident on their surface is split into two rays, the *ordinary ray* that obeys Snell's law of refraction and the *extraordinary ray*. Both rays experience different refractive indices in the crystal and are in general in different polarization states, though not in linear polarization states except for certain propagation directions relative to the crystal axis. Various types of birefringent polarizers have been developed mostly based on prisms (Nicol prism, Glan–Thomson prism, Glan–Foucault prism, Glan–Taylor prism, Wollaston prism, Rochon prism, and Sénarmont prism). We briefly describe here only the Glan–Taylor prism and the Wollaston prism.

For the *Glan–Taylor prism* in Figure 4.15 both prisms of calcite are polished so that the cutting edges of the crystal are parallel to the optical axis. Then, the ordinary and the extraordinary beam have the same angle of refraction. For perpendicular incidence, both beams propagate along the same way but with different speed. The cutting angle α is chosen so that the ordinary beam gets totally reflected while the extraordinary beam gets transmitted. This is achieved because the refractive index of the ordinary beam is higher than that of the extraordinary beam. The second prism is used to compensate the refraction of the extraordinary beam. At the end of the prism, only the p-polarized extraordinary beam exits the prism in the same direction as the incident beam.

The Glan–Foucault prism is built similar to the Glan–Taylor prism, except that the crystal is cut perpendicular to the optical axis. The result is that the transmitted beam is now s-polarized.

Very similar to the Glan–Foucault prism is the Glan–Thompson prism, except that the air gap between the prisms is replaced by a transparent glue.

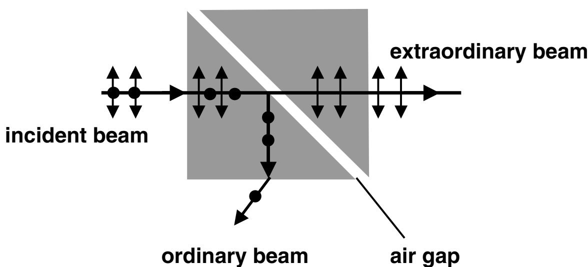


Figure 4.15 Light propagation through a Glan–Taylor prism.

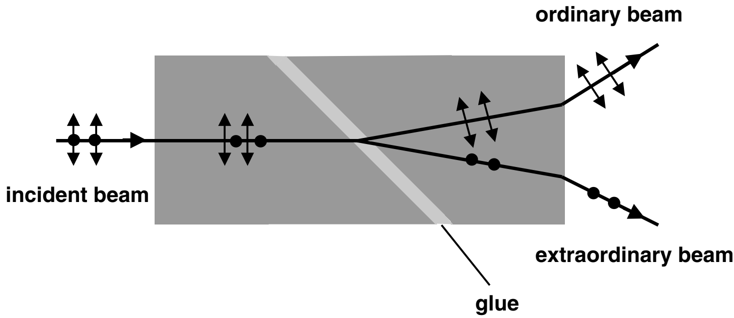


Figure 4.16 Light propagation through a Wollaston prism.

The *Wollaston prism* consists of two calcite prisms bonded with a transparent glue (see Figure 4.16). Incident polarized or unpolarized light gets divided into two perpendicular linearly polarized beams.

If one places a perfect polarizer in a polarized beam, the intensity of the light that passes through the polarizer is given by Malus' law.

$$I = I_0 \cos^2(\theta), \tag{4.19}$$

where I_0 is the initial intensity and θ is the angle between the initial polarization direction of the light and the axis of the polarizer.

In practice, the actual transmission is somewhat lower than this because some light is lost in the polarizer. Real polarizers do also not perfectly block the polarization orthogonal to their polarization axis. The ratio of the transmission of the unwanted component to the wanted component is called *extinction ratio* and varies from around $1:500$ to about $1:10^6$ depending on the polarizer.

4.4.2.3 **Optical Retarders**

Retarders or wave plates are used to split a light wave into two orthogonal linear polarization components and to produce a phase shift between them. That means they change the polarization form of the wave without polarizing or inducing an intensity change in the light beam. A special retarder is the compensator that generates a phase shift of $\pi/2$.

Retarders that cause a change of linearly polarized light into circularly polarized light are known as *quarter wave retarders*. Their unique property is to turn elliptically polarized light into linearly polarized light or to transform linearly polarized light into circularly polarized light. For that purpose, birefringent uniaxial materials are used. The resulting retardance is given by

$$\delta = (n_e - n_o)\frac{d}{\lambda}, \tag{4.20}$$

where d is the thickness of the retarder, λ is the wavelength of light, and n_e and n_o are the refractive indices of the extraordinary beam and of the ordinary beam.

A retarder that produces a $\lambda/2$ phase shift is known as a *half wave retarder*. Typical retarder materials are crystals such as calcite, magnesium fluoride, and most commonly quartz, and oriented polymers.

In automated ellipsometers in which oscillating retarders are used, the oscillating phase retardation is obtained by applying a mechanical stress, an electric field, or a magnetic field to the transparent optical material.

4.4.3 Optical Fibers

For the reflectometric measurement, optical fibers are used to bring the light from the light source to the object where it gets reflected and a second fiber collects the reflected light and brings it to the entrance slit of the spectrometer. Usually, the two separate fibers are combined in a bifurcated fiber, the so-called Y-fiber, with two branches for connection to the light source and the spectrometer and one common branch leading the two fibers to the measuring object.

Optical fibers are circular dielectric waveguides. They consist of a central core surrounded by a concentric cladding and a protective mantle of one or two layers (see Figure 4.17).

Light propagation through the fiber is established by the fact that the core has a higher refractive index (n_1) than the cladding (n_2). Then, light entering the fiber at one end will be totally reflected at the interface core–cladding if the critical angle of total reflection is exceeded (see Figure 4.18). Contrary, only light that enters the fiber with

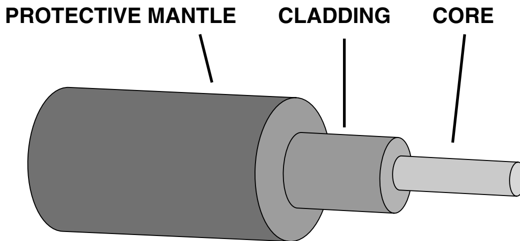


Figure 4.17 Constitution of an optical fiber: core, cladding, and protective mantle.

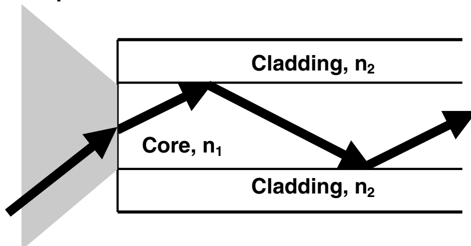


Figure 4.18 Geometrical optics model of total internal reflection for the transmission of light through an optical fiber.

an angle $\alpha < \alpha_{max}$ will be transmitted through the fiber. The sine of α_{max} is known as numerical aperture NA of the fiber, that is, $NA = \sin(\alpha_{max})$. Typical values for NA of fibers for reflectometry are $NA = 0.22$ and sometimes $NA = 0.11$.

In the more elaborated description of propagating nonradiating eigenmodes, the transmitted light is composed of modes with each of them described by a wave that propagates in a medium with effective refractive index n_{eff}. Taking into account the cylindrical symmetry of the fiber, the wavenumber $\eta = (2\pi/\lambda)n_{eff}$ is given by the solution of the equation

$$p\frac{J_{n-1}(p)}{J_n(p)} + q\frac{K_{n-1}(q)}{K_n(q)} = 0, \tag{4.21}$$

with $p = (D/2)\sqrt{((2\pi/\lambda)n_1)^2 - \eta^2}$ and $q = (D/2)\sqrt{\eta^2 - ((2\pi/\lambda)n_2)^2}$. J_n and K_n are the Bessel function of first kind and the modified Bessel function of second kind with order n. D is the core diameter of the fiber.

Equation 4.21 can have multiple solutions η_m with $m = 1, 2, \ldots$, for each order n. It is therefore common practice to number the corresponding eigenmodes with indices n and m. The eigenmodes of a fiber are abbreviated with LP_{nm}, LP meaning linearly polarized. Figure 4.19 shows the intensity distribution of the LP modes LP01, LP02, LP11, and LP21.

As the above solutions are strongly dependent upon the fiber and cladding materials, not all LP modes that are solution of (4.21) can actually travel through the fiber. Introducing the parameter V defined as

$$V = \frac{2\pi}{\lambda}\frac{D}{2}\sqrt{n_1^2 - n_2^2} \tag{4.22}$$

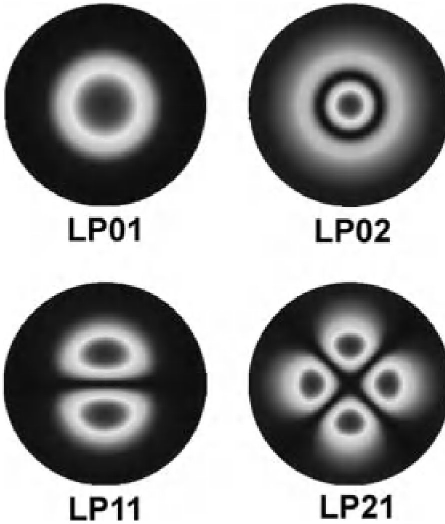


Figure 4.19 Transmission of light through an optical fiber with distinct fiber modes: the intensity distribution of the four LP modes LP01, LP02, LP11, and LP21.

Table 4.5 Cutoff values $V_{c,nm}$ of some lower modes LP_{nm}.

$V_{c,nm}$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	0	3.832	7.016	10.173
$n=1$	2.405	5.520	8.654	11.792
$n=2$	3.832	7.016	10.173	13.324

there exists a cutoff value $V_{c,nm}$ for each mode LP_{nm}, except for LP_{01}, for which $V_{c,01}=0$. If V becomes less than a certain $V_{c,nm}$, this mode LP_{nm} and higher modes cannot travel through the fiber. The mode LP_{01} can always propagate through the fiber and is therefore called *ground mode*. Vice versa, the mode LP_{nm} exists only for light with wavelengths $\lambda < \lambda_{c,nm}$ with

$$\lambda_{c,nm} = \frac{\pi D}{V_{c,nm}} \sqrt{n_1^2 - n_2^2}. \tag{4.23}$$

A list of cutoff values $V_{c,nm}$ for some lower modes is given in Table 4.5.

The most commonly used material for optical fibers is fused silica (SiO_2) with index-modifying dopants such as GeO_2. Then, the refractive index gets increased as can be seen from Figure 4.20, where the refractive index of pure SiO_2 is compared with the refractive index of GeO_2-doped SiO_2 with a concentration of 13.5% GeO_2. The curves are calculated with a three-term Sellmeier formula with the coefficients from Ref. [130] (wavelength in microns) given in Table 4.6.

It is important to prepare the silica to a certain extent free of SiOH because the OH^- can be excited by light and absorbs at the same wavelengths as water (H—OH) exhibits absorption. Optical fibers can be distinguished in UV fibers and IR fibers.

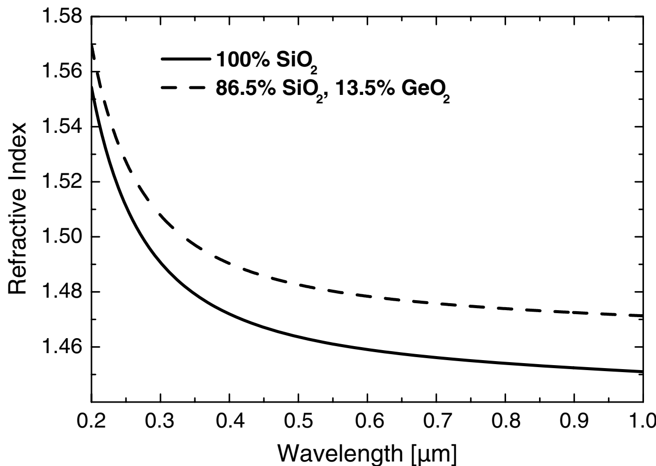


Figure 4.20 Refractive index of pure SiO_2 and GeO_2-doped SiO_2 with 13.5% GeO_2.

Table 4.6 Sellmeier coefficients of fused quartz (SiO_2) and GeO_2-doped SiO_2 with 13.5% content of GeO_2.

	100% SiO_2		86.5% SiO_2, 13.5% GeO_2	
j	A_j	B_j	A_j	B_j
1	0.696 750	0.0 069 066	0.711 040	0.0 042 700
2	0.408 218	0.0 115 662	0.451 885	0.0 129 408
3	0.890 815	99.00 559	0.704 048	99.93 707

This distinction considers the amount of SiOH in the fiber. The IR fiber has only a small amount of remaining SiOH and therefore does not exhibit strong absorption bands in the spectral range between 900 and 1650 nm. Figure 4.21 shows the comparison of the light transmitted through a UV fiber (solid line) and an IR fiber (dash–dot-dot line) of the same core diameter and same length in the wavelength range 900–1650 nm. The absorption by SiOH strongly reduces the intensity of the transmitted light between 1200 and 1500 nm. The absorption around 1415 nm is as strong such that it can also be recognized in the spectrum of the IR fiber as a small bump.

Illumination with deep UV radiation below 220 nm wavelength can lead to *solarization* of fibers. This means that the UV radiation damages the structure of the fused silica fiber. When the silica recrystallizes, it forms crystalline inclusions of crystobalite, a certain mode of crystalline quartz. They act as scattering centers for the UV radiation and therefore reduce the transmission of the fibers in the range between 220 and 250 nm. Solarization can partly be reversed by heating of the fiber, but mainly can be taken into account by referencing.

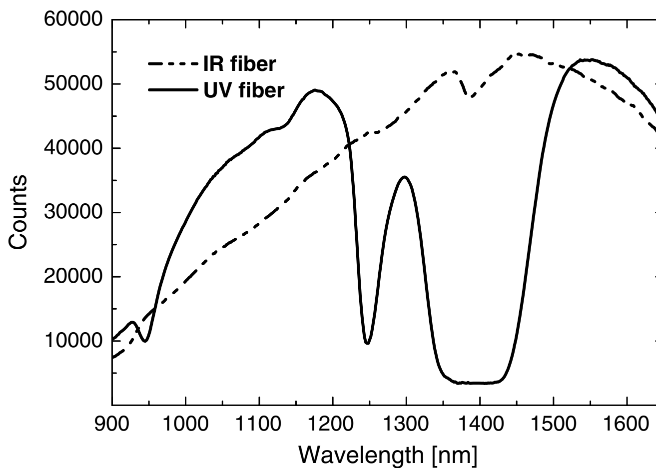


Figure 4.21 Comparison of the transmitted light of a UV fiber (solid line) and an IR fiber (dash–dot-dot line) of the same core diameter and length in the wavelength range 900–1700 nm.

4.4.4 Miniaturized Spectrometers

Classical spectrometers consist of an input slit, a rotating dispersive element, an output slit, and a single photodetector. The main advantages of this setup are a high sensitivity and very low stray light. The rotating dispersive element, either a prism or a grating, allows separation of multichromatic radiation into its spectral components. Using multiple dispersive elements and several slits, a broad wavelength range from the UV (190 nm) up to the near infrared (3000 nm) is available with only two different photodetectors. This setup is realized in several commercially available spectrometers mainly used for chemical analysis.

The main disadvantages of this setup – serial measurement and moveable parts – are overcome with the development of detector arrays. They allow parallel measurement at many wavelengths and only need fixed components. Therefore, additionally, they allow a compact and robust setup as miniaturized spectrometer.

The principles of such a miniaturized spectrometer are as follows:

- Imaging of the entrance slit on the image sensor.
- Diffraction of the incoming broadband light ("white light") into spectral components ("colors") with a diffraction grating.
- Consideration of only the first-order diffraction.
- The light releases electrons (photo effect) that are stored in each pixel of the diode line (condensator).
- The electrons get drained (electrical current, analogue) and are transformed into discrete voltage steps (digital) (A/D conversion).

There are three types of spectrometers available: (a) with reflection grating and mirrors (collimating and focusing), (b) with transmission grating and mirrors (collimating and focusing), and (c) with transmission grating and lenses (collimating and focusing). They are illustrated in Figure 4.22.

4.4.4.1 **Gratings**

Gratings are optical elements that divide broadband light or other electromagnetic radiation in its components by diffraction (see also Section 2.5). They have the advantages of higher dispersion and lower production costs compared to prisms as dispersive elements in spectrometers. A grating consists of a series of equally spaced parallel grooves formed in a coating deposited on a suitable substrate. The way in which the grooves are formed separates gratings into two basic types, holographic and ruled. For an overview on diffraction gratings, we refer to the *Diffraction Grating Handbook* [131] from Newport Corp. or the *Diffraction Gratings: Ruled and Holographic Handbook* from HORIBA Jobin-Yvon S.A. [132].

The first step in the production of a *holographic reflection grating* is the coating of an appropriate polished substrate with a photosensitive material (photoresist). The photoresist is exposed by intersecting beams of a laser. They generate a series of parallel equally spaced interference fringes whose intensities vary in a sinusoidal pattern. This pattern impinges on the resist and is transferred into the

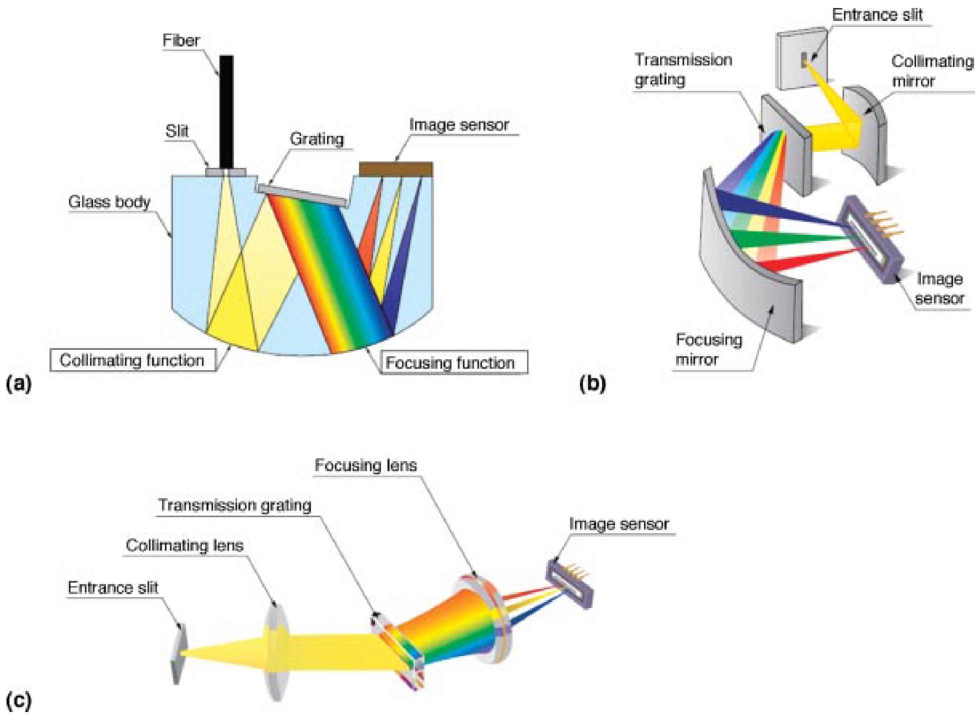


Figure 4.22 Schematics of miniaturized spectrometers. (a) With reflection grating and mirrors, (b) with transmission grating and mirrors, (c) with transmission grating and lenses. Copyright Hamamatsu Photonics K.K.

resist by developing the resist. The substrate is then coated with a reflective material. The whole grating can be replicated by the same process used for ruled gratings.

In the production of a *holographic transmission grating*, a transparent glass or silica substrate is coated with a thick film of a photosensitive material. The photoresist gets exposed similar to the holographic reflection gratings, resulting in a sinusoidal interference pattern in the photoresist. Depending on the properties of the photoresist, the transmission grating is finally covered by a protection glass or a silica plate.

The production of *ruled gratings* is more complex. Grooves are scribed in a reflective layer with a diamond mounted on an engine. For an original or master grating, usually glass or copper is used as substrate that is polished to a 10th of the wavelength and is coated with a thin layer of aluminum by vacuum deposition. The ruling of parallel, equally spaced grooves is a slow process and is accompanied by many tests of the ruling before the actual ruling. A series of test rulings must be made to check the grating for efficiency, groove profile, and stray light. After extensive testing, an original grating is ruled on a large substrate. High-quality ruled gratings are then formed by replication from this master grating.

The theory of diffraction by a grating has been already discussed in Section 2.5 for several gratings. Here, we only briefly repeat the basics and discuss the properties of holographic gratings and of ruled gratings.

If ε is the angle between the incident light and the normal to the grating and β is the angle between the diffracted light and the normal to the grating, then these angles satisfy the following *grating equation*:

$$g \cdot (\sin(\varepsilon) \pm \sin(\beta)) = m\lambda, \tag{4.24}$$

where m is the order of diffraction ($m = 0, \pm 1, \pm 2, \ldots$), λ is the diffracted wavelength, and g is the *grating period* (the distance between successive grooves). The reciprocal of the grating period is called *groove density*, $L = 1/g$. All components of light corresponding to $m = 0$ follow a straight line given by $\beta = \varepsilon$. It is not possible to separate the wavelengths for this order.

For a given set of angles of incident and diffraction, the grating equation is valid at more than one wavelength, giving rise to several orders of diffracted radiation. For example, if the wavelength $\lambda = 600$ nm gets diffracted in the first order $m = \pm 1$, then also the wavelength $\lambda = 300$ nm gets diffracted in the second order $m = \pm 2$ at the same angle of diffraction. The number of orders produced is limited by the groove spacing and the angle of incidence, which obviously cannot exceed 90°. The maximum spectral bandwidth that can be obtained in a specified order without spectral overlap from adjacent orders is called *free spectral range*. If the grating spacing decreases, the free spectral range increases. Order overlap can be compensated either by the intelligent use of sources, detectors, and filters or by using a permanently installed long-pass optical filter in the fiber entrance connector or an order sorting coating on a window in front of the detector.

Depending on the grating period g (or the groove density L) and the angle of incidence ε, it may not be possible to obtain diffracted light. For example, for $L = 2000$ grooves/mm, $\varepsilon = 0°$, and $\lambda = 500$ nm we obtain $\sin(\beta) = 1$ for the first-order $m = \pm 1$. All wavelengths $\lambda > 500$ nm cannot be diffracted by this grating. The spectral range covered by a grating is the same for ruled and holographic gratings with the same grating period.

Angular dispersion of a grating describes how grating separates light into its components. It corresponds approximately to the projection of the angle of refraction on the detector plane. The equation for dispersion is given by

$$\frac{d\beta_m}{d\lambda} = \frac{m}{g \cdot \cos(\beta_m)}. \tag{4.25}$$

Multiplying both sides with the focal length of the system, inverting both sides, and using the abbrevation $dx = f \times d\beta$ gives the *reciprocal linear dispersion*

$$D = \frac{d\lambda}{dx} = \frac{g \cdot \cos(\beta_m)}{m \cdot f}. \tag{4.26}$$

Angular dispersion can be increased by increasing the angle of incidence or decreasing the distance between successive grooves.

Looking at the *resolving power* of a grating, there is no difference for holographic and ruled gratings with identical groove spacing. Holographic gratings are, however, available with higher groove densities. The resolving power of a grating is the product of the diffracted order in which it is used and the number of grooves illuminated by the incident radiation

$$\frac{\lambda}{\Delta\lambda} = \frac{m \cdot w}{g} \tag{4.27}$$

with w being the width of the entrance slit in millimeters. The resolving power is a property of only the grating and does not depend on the optical and mechanical properties of the spectrometer system.

However, due to the widening caused by the angular dispersion the actual *spectral resolution* of the grating becomes

$$\frac{\lambda}{\Delta\lambda} = \frac{m \cdot w}{g \cdot \cos(\beta_m)} = w\frac{\mathrm{d}\beta}{\mathrm{d}\lambda}. \tag{4.28}$$

The entrance slit of the spectrometer also limits the spatial spread of the light that enters the spectrometer. Therefore, the slit width is an essential factor in determining the spectral resolution. The narrower the slit width, the more the resolution is gets improved.

One of the most important properties of a grating is the *grating efficiency*. Diffraction is usually symmetric with respect to the energy diffracted into the order $+m$ and $-m$ (see Section 2.5). However, this homogeneous distribution over all diffraction orders can be modified by the groove profile, the groove dimensions, the groove materials, and other groove and grating properties such as the reflectivity, as in the case of ruled gratings. The result is that the diffracted energy is redirected into a certain desired order. The grooves of a ruled grating have a sawtooth profile with one side longer than the other. This geometry concentrates diffracted radiation in a specific region of the spectrum, increasing the efficiency of the grating in that region. The wavelength at which maximum efficiency occurs is the *blaze wavelength*. Holographic gratings are generally less efficient than ruled gratings because of their sinusoidal profile. Their sinusoidal shape can, however, be altered to approach a triangle or sawtooth profile. Then, the efficiency increases.

The efficiency of a ruled grating not only depends on the shape but is also different for p-and s-polarized light as the reflection is polarization dependent. Figure 4.23 shows an example of the absolute efficiency of a ruled grating with 1800 grooves/mm with a blaze wavelength $\lambda_{\mathrm{blaze}} = 500$ nm. The useful spectral range is from 250 to 1050 nm. The different curves belong to s-polarized light (dashed–dot-dot line), p-polarized light (dashed line), and unpolarized light (solid line).

4.4.4.2 **Detectors**

The detectors of miniaturized spectrometers are diode line arrays based on the following two technologies:

- CMOS (complementary metal oxide semiconductor)
- CCD (charged coupled device)

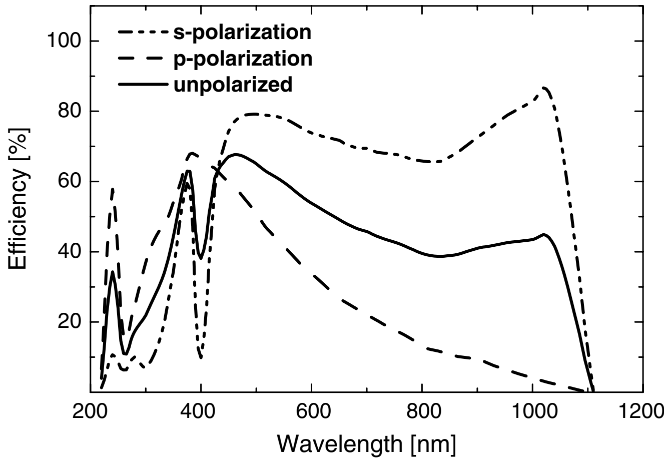


Figure 4.23 Absolute efficiency of a ruled grating with 1800 grooves per mm, blazed at 500 nm. The different curves belong to s-polarized light (dashed–dot-dot line), p-polarized light (dashed line), and unpolarized light (solid line). This grating can be used for the spectral range from 250 to 1050 nm. (Available from Optometrics LLC, USA.)

Both use the principle of the inner photoeffect where electrons get released by light but remain in the material. They change from the valence band to the conduction band and get stored in the potential well, a region in the semiconductor component, from which they are read out. The properties of the released electrons depend on various parameters: the absorption coefficient, the recombination time of the generated electron–hole pairs, the diffusion path, and the chemical and physical structure of the material above the photosensitive layer.

CMOS Detectors CMOS sensors are based on a combination of photodiode with condensator for the charge collection and an amplifier with data connections for the readout and reset. This is realized by a pair-wise combined p-channel and n-channel field effect transistors (MOSFETs).

For a MOSFET, the conductivity in a certain volume – the channel – can be influenced by applying a voltage on an isolated control electrode, the gate. The current between two electrodes (source and drain) through the channel depends on the voltage U_{DS} between drain and source and on the voltage U_{GS} at the control electrode. Source and drain are separated from the bulk by p–n junctions that are in reverse biasing. Figure 4.24 shows a picture of a p-channel MOSFET. For an n-channel MOSFET, only the n-type and p-type silicon regions must be exchanged to p-type and n-type silicon regions. A conductive p-channel (n-channel) is formed in the n-region (p-region) only if the voltage U_{GS} is sufficiently high. The advantage of a MOSFET is that the current between source and drain through the channel can be controlled without power consumption.

To use a MOSFET as photodiode, one has to replace the metal electrode by a transparent electrode. In modern semiconductor technology, the gate is formed by a

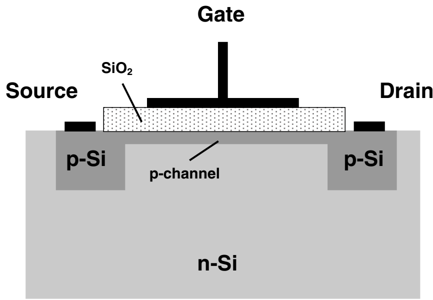


Figure 4.24 Picture of a p-channel MOSFET.

thin, heavily doped polycrystalline silicon. Unlike crystalline silicon, polycrystalline Si has a high conductivity that is further enhanced by doping. Using a p-type MOS as photodiode, the gate gets biased negatively. Then, a potential well is established below the isolator in the n-Si region. If ultraviolet, visible, or infrared light hits a silicon atom in the n-Si region in Figure 4.24, it generates an electron–hole pair. The free electron gets collected in the charge pool (potential well), while the hole migrates away from the pool. Figure 4.25 depicts a CMOS unit with photodiode and amplifying n-type MOSFET. Such a unit is also called active pixel sensor (APS). The photodiode generates electrons proportional to the amount of the incident light. They are stored in the condensator (potential well) and are converted into voltage by the downstream MOSFET that additionally amplifies this voltage. From this MOSFET, it is read out by further downstream electronics.

Compared to CCD sensors, one big advantage is that the charge-to-voltage conversion already happens in the pixel and the charges need not to be shifted. In consequence, the blooming is much lower. A disadvantage is that there is much photoinsensitive electronics between the photosensitive diodes.

As the conversion charge-to-voltage already happens in the pixel, the sensitivity of CMOS sensors is very high. Linear arrays of CMOS photodiodes feature excellent

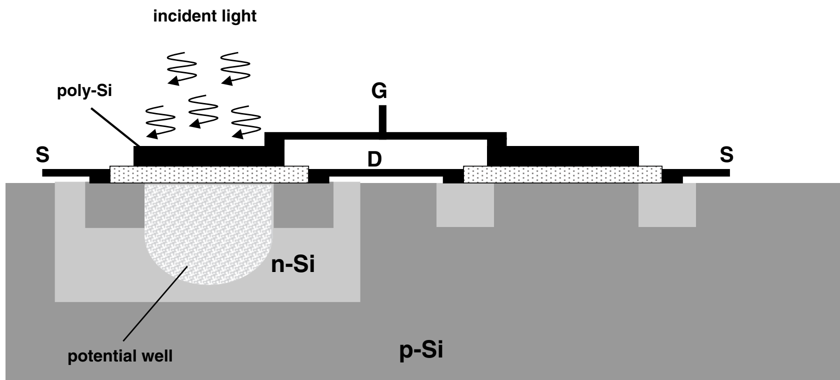


Figure 4.25 Picture of an illuminated CMOS unit.

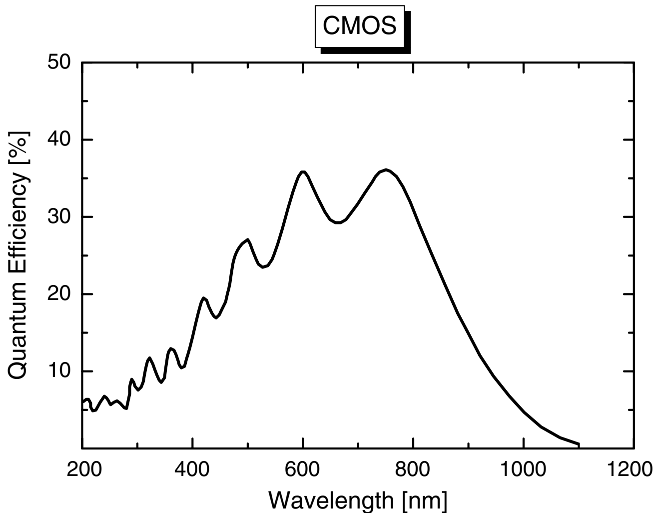


Figure 4.26 Quantum efficiency of a silicon CMOS image detector (e.g., S10114-1024Q from Hamamatsu Photonics K.K.) in dependence on the wavelength of incident radiation. The oscillations are caused by a thin transparent protective quartz window in front of the detector.

linearity with respect to incident light, low internal noise, and a wide spectral response. They have a long lifetime, and they are mechanically rugged, compact, and lightweight. In Figure 4.26, the quantum efficiency of a CMOS image sensor is shown (e.g., S10114-1024Q from Hamamatsu Photonics K.K.). The array has a broad spectral response from 200 nm to approximately 1000 nm wavelength, with a high level of ultraviolet sensitivity. The quantum efficiency tends to be zero for wavelengths longer than 1000 nm because for the intrinsic photoeffect in Si at least an energy of 1.1 eV (band gap of Si) is necessary to transport an electron from the valence band to the conduction band. This energy corresponds to a wavelength of 1127 nm.

InGaAs Detectors For measurements at longer wavelengths, another semiconductor has been proven to be well suited: indium gallium arsenide, InGaAs. The photosensitivity of a linear InGaAs detector array (G9204-512D and G9206-02, TE-cooled, from Hamamatsu Photonics K.K.) is plotted in Figure 4.27 versus the wavelength. Depending on the cooling of this sensor, the spectral range can be extended from 1700 to 2200 nm, but then with less sensitivity at wavelengths between 800 and 1200 nm.

CCD Detectors CCD means "charge coupled device" and is an electronic device that can transport electronic charges. CCDs consist of an array of isolated photodiodes. In contrast to the CMOS photodiodes, the charge is not directly converted into a voltage, but must be transported from the photodiode to an amplifier. The technique was invented by William Boyle and George E. Smith in 1969 at the Bell Laboratories as a device for data storage. However, it was never realized as data storage element.

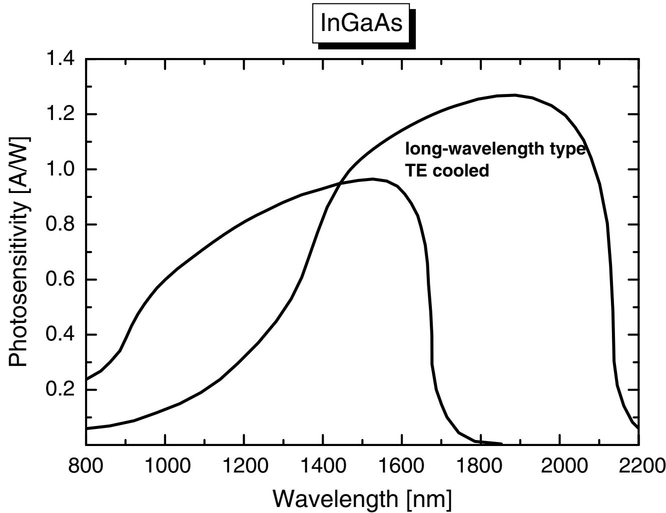


Figure 4.27 Photosensitivity of a linear InGaAs detector array (G9204-512D and G9206-02, TE-cooled, from Hamamatsu Photonics K.K.) in dependence of the wavelength of the incident radiation.

Instead, this technique is used to shift photoinduced charges in many small steps (vertical and horizontal shift registers) to a central analog-to-digital converter. The principle of this shifting is comparable to a chain of buckets: one filled bucket is drained into the next empty bucket in the chain, so on and so forth. The shift is forced electronically by applying external voltages. Boyle and Smith obtained for their invention the Nobel Prize in physics in 2009.

As the photodiodes are isolated from each other, it may happen that the capacity of the charge pool is exceeded and excess charges flow into the pools of neighboring photodiodes. This effect is called *blooming*. Blooming occurs for long-duration intensive illumination of the pixel. It can be reduced by additional vertical and lateral overflow drains that direct the excess charges away from the pixel. The disadvantage is that these additional electric lines restrict the size of the pixels and, hence, the photosensitivity of the pixels.

Compared to CMOS sensors, the dynamics of a CCD pixel is approximately double the dynamics of a CMOS pixel. Dynamics is here defined as the ratio of the saturation limit and the photosensitivity. Moreover, the noise is less than that in CMOS.

The quantum efficiency is plotted in Figure 4.28 for a front-illuminated CCD image sensor, a back-thinned CCD image sensor, and a back-thinned linear CCD detector (e.g., S10127, S10200-02, and S10420-1106 from Hamamatsu Photonics K.K.). As can be recognized, the quantum efficiency increases by almost a factor 2 if the photodiode gets thinned at its rear side so that it becomes transparent and then coated by an antireflective coating. This design is, however, more expensive.

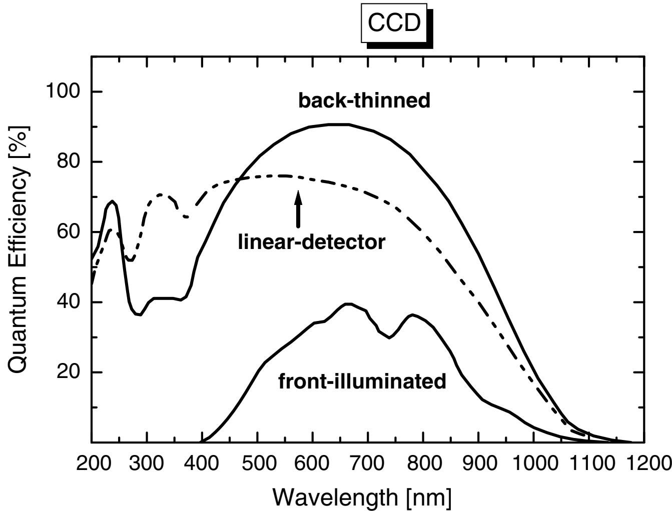


Figure 4.28 Quantum efficiency of a front-illuminated CCD image sensor, a back-thinned CCD image sensor, and a back-thinned CCD linear detector (e.g., S10127, S10200-02, and S10420-1106 from Hamamatsu Photonics K.K.) in dependence on the wavelength of incident light.

4.4.4.3 **System Properties**

Spectral Resolution The resolution of the complete optical system depends not only on the grating resolving power but on focal length f, slit size, the optical quality of all components, and the system alignment. Therefore, the resolution of the optical system is usually much less than the resolving power of the grating.

One important quantity for the determination of the spectral resolution is the spectral range. For a spectrometer with a diode line array with N pixels of width Δp (mm), a grating with grating period g (μm), and a focal length f (mm) of the system, the wavelength $\lambda(j)$ at pixel j is

$$\lambda(j) = 1000 \cdot g \cdot \left\{ \sin\left[\arctan\left(\frac{\Delta p}{f} \left(j - \frac{N}{2} + s \right) \right) + \beta_{N/2} \right] - \sin \varepsilon \right\}. \tag{4.29}$$

Here, $\beta_{\mathrm{N/2}}$ is the diffraction angle of the wavelength in the middle of the detector for which the grating is calculated, ε is the incident angle, and s is the pixel shift. The pixel shift is usually $s = 0$, but can also amount to a nonzero value. The pixel width may be different for CCD detectors and CMOS detectors. The minimum wavelength $\lambda_{\min} = \lambda(1)$ and the maximum wavelength $\lambda_{\max} = \lambda(N)$ are the wavelengths that are imaged to the first and the last pixel of the detector without pixel shift s (see Figure 4.29).

Spectral resolution is not unique but there exist the following three terms for spectral resolution:

- The Rayleigh criterion (accordingly to DIN), $\Delta\lambda_{\mathrm{Rayleigh}}$.
- The line width as full width at half maximum (FWHM), $\Delta\lambda_{\mathrm{FWHM}}$.
- The pixel dispersion as (spectral range)/(number of pixels), $\Delta\lambda_{\mathrm{Pixel}}$.

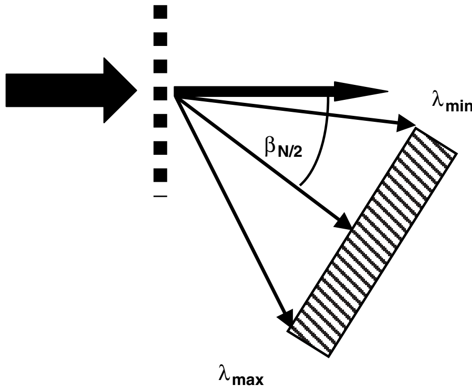


Figure 4.29 Spectral range of a spectrometer according to (4.29).

The spectral resolution is usually determined by examination of closely spaced absorption or emission lines. To separate two closely spaced lines, its distance must be $\Delta\lambda_{Rayleigh}$. It is given by the condition that the maximum in the interference pattern of λ_2 falls in the first minimum in the interference pattern of λ_1. This is illustrated in Figure 4.30. The solid black line indicates the resulting sum in the intensity. The two lines with $I_{max,1} = I_{max,2}$ are therefore separated if the intensity of the saddle point is equal to or less than 81% (corresponding to $8/\pi^2$), that is, $I_{saddle} \leq 0.81 \cdot I_{max}$. This condition is called *Rayleigh criterion*.

A more well-known and practical alternative is the spectral half-width or FWHM $\Delta\lambda_{FWHM} = \lambda_2(I_{max}/2) - \lambda_1(I_{max}/2)$. It describes the broadening of an emission or absorption line by the optical system and corresponds to the spectral width of a peak at 50% of the peak power value. It must be smaller than the spectral width of the line

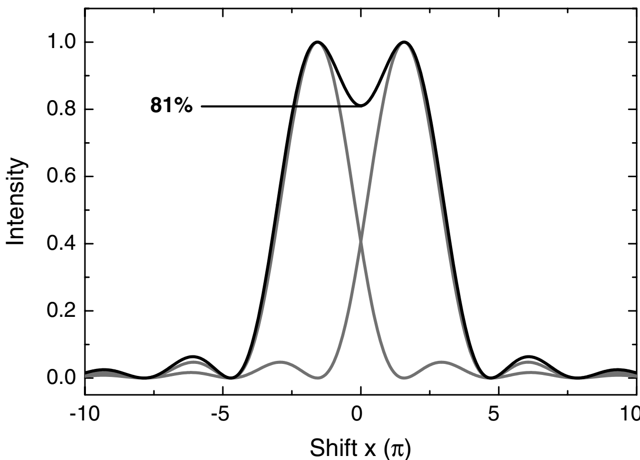


Figure 4.30 Illustration of the Rayleigh criterion.

itself. $\Delta\lambda_{FWHM}$ is about 80% of the resolution defined by Rayleigh criterion: $\Delta\lambda_{FWHM} \approx 0.8 \cdot \Delta\lambda_{Rayleigh}$.

$\Delta\lambda_{Pixel}$ is the linear dispersion of the spectrometer. To find a connection between $\Delta\lambda_{Pixel}$ and $\Delta\lambda_{Rayleigh}$, one has to consider the width of the entrance slit and the imaging properties of the spectrometer. Approximately, it is $\Delta\lambda_{Rayleigh} \approx 3 \cdot \Delta\lambda_{Pixel}$.

In Table 4.7, we give wavelengths of often-used spectral lines from various elements used for determination of the spectral resolution.

Table 4.7 Wavelengths of often used spectral lines.

Wavelength (nm)	Description	Spectral line	Element
2325.4		Infrared mercury line	Hg
1970.09		Infrared mercury line	Hg
1529.882		Infrared mercury line	Hg
1060.0		Neodymium-glass laser	Nd
1013.98	t	Infrared mercury line	Hg
852.11	s	Infrared cesium line	Cs
769.8979	A′	Red potassium line	K
766.4907	A′	Red potassium line	K
760.82	A	Red oxygen line	O
706.5188	r	red helium line	He
686.72	B	Red oxygen line	O
656.2725	C	Red hydrogen line	H
643.8469	C′	Red cadmium line	Cd
632.8		He–Ne gas laser	He-Ne
589.5923	D1	Yellow sodium line	Na
589.2938	D	Yellow sodium line (mid of D1 and D2)	Na
588.9953	D2	Yellow sodium line	Na
587.5618	d	Yellow helium line	He
579.0654		Yellow mercury line	Hg
576.9596		Yellow mercury line	Hg
546.074	e	Green mercury line	Hg
527.03 602	E	Green iron line	Fe
486.1327	F	Cyan hydrogen line	H
479.9914	F′	Blue cadmium line	Cd
435.8343	g	Blue mercury line	Hg
430.79 048	G	Blue iron line	Fe
404.6561	h	Violet mercury line	Hg
396.8468	H	Violet calcium line	Ca
393.3666	K	Violet calcium line	Ca
365.0146	i	Ultraviolet mercury line	Hg
334.1478		Ultraviolet mercury line	Hg
312.5663		Ultraviolet mercury line	Hg
296.7278		Ultraviolet mercury line	Hg
280.4		Ultraviolet mercury line	Hg
253.6519		Ultraviolet mercury line	Hg
248.3		Ultraviolet mercury line	Hg

Wavelength Calibration Initially, the AD converter delivers the spectral signal in counts for each pixel. To assign it to a wavelength, the pixel number k must be transformed into a corresponding wavelength $\lambda(k)$. This is usually done with a multiorder polynomial:

$$\lambda(k) = \sum_{n=0}^{N} p_n \cdot k^n. \tag{4.30}$$

The coefficient p_0 is the wavelength of the first pixel. The choice for the maximum order N of the polynom depends upon the nonlinear behavior of the spectrometer.

The calibration can be done using one or more suited spectral lamps (e.g., a mercury arc lamp) exhibiting single narrow emission lines. Look again at Table 4.7 where we give wavelengths of often-used spectral lines from various elements for determination of the spectral resolution and for spectral calibration. The parameters p_n are calculated by regression.

Dynamic Resolution Measurement of tiny intensity changes and stability of the signal depends upon each other and is mainly determined by the electronic noise because the stability of the light path is maintained. To obtain a measure for the accuracy of an intensity value, it is common practice to measure repeatedly N times the intensity with a fixed integration time. The standard deviation σ of this N measurements defines the accuracy of the intensity measurement or the *signal noise*.

The dynamic range of the spectrometer is given by the ratio of the intensity in saturation I_{sat} and the signal noise σ and corresponds to the *signal-to-noise ratio* $\mathrm{SNR} = I_{sat}/\sigma$. The *useful dynamic range* $\mathrm{dynR} = (I_{sat} - I_{dark})/\sigma$ gets diminished by the dark current I_{dark}. I_{dark} is a small electrical output of the detector array without incident light. It is caused by thermal generation of carriers in each pixels, mainly due to SiO_2–Si interface states (for Si as base detector material). It is, therefore, strongly correlated with the operation temperature. Another contribution to I_{dark} comes from the downward electronics in the AD converter.

The SNR keeps stable with increasing temperature; however, as the dark current increases with temperature, the useful dynamic range decreases with temperature. A limiting factor for the useful dynamic range is the fluctuation of the light source. Typical values for miniaturized spectrometers are $I_{dark} = 500$–2500 counts for the dark current, a dark noise of 1–10 counts rms, a signal noise of $\sigma = 10$–100 counts, and an intensity in the saturation of $I_{sat} = 2^{12} = 4096$ counts (12 bit) or $I_{sat} = 2^{16} = 65\,536$ counts (16 bit). Hence, the SNR is typically in the range of $\mathrm{SNR} = 250:1$ to $\mathrm{SNR} = 10\,000:1$.

Ghosts and Stray Light Stray light influences the dynamic resolution because it limits the dynamic range. However, the measurement of tiny intensity changes are affected little, as the noise is stronger than the stray light in most cases. Moreover, if the source of the stray light is known it can be taken into account in the calculation of the signal. Typical stray light performance is $<0.05\%$ at long wavelengths and $<0.10\%$ at short wavelengths.

Spurious spectral lines, the so-called *ghosts*, arise from periodic errors in groove spacing or irregularities of the reflecting surfaces. Ruled diffraction gratings exhibit stray light and ghosting since they cannot be manufactured without defects. Interferometrically controlled ruling engines minimize ghosts but cannot eliminate them. Holographic gratings are produced optically, for which groove form and spacing are almost perfect. Hence, holographic gratings are free from ghosts.

A method to determine the stray light is the long-pass filter method. It uses white light obtained by passing through a filter with particular cutoff wavelength. Stray light is defined by the ratio of the transmittance in the wavelength transmitting region I_t to the tansmittance in the wavelength blocking region I_b. The stray light level, SLL, then is

$$\mathrm{SLL} = 10 \cdot \log_{10}\left(\frac{I_b}{I_t}\right). \tag{4.31}$$

Usually, I_t and I_b are measured close to the cutoff wavelength. Typical values of SLL range from -35 dB to -27 dB, or $I_b = 0.0003 - 0.002 \times I_t$.

5
Thin-Film Thickness Determination

Evaluation of the thickness of one or more layers in a layer stack can be made by two methods: a fast Fourier transformation (FFT) analysis and a regression analysis with χ^2 test. The FFT analysis is restricted to thicknesses of more than approximately 1 μm. The reason is that it needs at least one full oscillation in the spectrum. If the spectral range of the spectrometer is given by λ_{min} and λ_{max}, the first complete oscillation in the spectrum of a film with refractive index $n(\lambda)$ can be recognized if one maximum appears at λ_{min} and the second maximum appears at λ_{max}. Then, the minimum thickness available from FFT is

$$d_{min} = \frac{1}{2((n(\lambda_{min})/\lambda_{min}) - (n(\lambda_{max})/\lambda_{max}))}. \tag{5.1}$$

This is in the order of 0.2 μm for a spectral range from 360 to 1000 nm and a typical refractive index of $n = 1.5$. It increases with decreasing wavelength range.

As the spectrometer has N pixel, the maximum number of minima and maxima is $N/2$, from which d_{max} theoretically follows as

$$d_{max} = \frac{N/2}{2((n(\lambda_{min})/\lambda_{min}) - (n(\lambda_{max})/\lambda_{max}))} = \frac{N}{2} \cdot d_{min}. \tag{5.2}$$

This, however, presumes that the focusing of each diffracted wavelength on a single pixel is perfect. In fact, the focus of long wavelengths is bigger than that of short wavelengths, and is often smeared over more than one pixel. Therefore, in practice one can reach about two-thirds of the above d_{max} as maximum resolvable thickness.

The regression analysis with χ^2-test is more complicated than the FFT as it uses the calculation of a layer stack as described in Chapter 3. On the other hand, it is well suited for small layer thicknesses down to a few 10 nanometers depending upon the layer material. In principle, it can be applied also to larger thicknesses, but the effort drastically increases with increasing thickness and the accuracy decreases. A reasonable (but not fixed) upper application limit is $d \approx 10$ μm.

Both methods are discussed in more detail in the following two sections.

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

5.1 Fast Fourier Transform

5.1.1 Single Layer

Without any restriction of the generality, we assume in the following a single layer on a substrate. Then, the reflectance spectrum is given by

$$R(\lambda, d) = \frac{R_{01}(\lambda) + R_{12}(\lambda) + 2\sqrt{R_{01}(\lambda)R_{12}(\lambda)} \cdot \cos((4\pi/\lambda)n_1(\lambda) \cdot d + \Phi(\lambda))}{1 + R_{01}(\lambda)R_{12}(\lambda) + 2\sqrt{R_{01}(\lambda)R_{12}(\lambda)} \cdot \cos((4\pi/\lambda)n_1(\lambda) \cdot d + \Phi(\lambda))}, \tag{5.3}$$

where R_{01} is the reflectivity of the interface air–layer and R_{12} is the reflectivity of the interface layer–substrate. It is obtained from the multiple beam interference of all reflected beams (see again Section 3.1 and (3.5)). The thickness of the film is d and its refractive index is n_1. We omit in the following the phase shift $\Phi(\lambda)$ as we will consider only nonabsorbing materials for the films and substrates or its value is negligible.

It is composed by a wavelength-dependent underground given by the lower envelope $R_{\mathrm{low_env}}$ in (3.14) and an oscillating part. This can be recognized from Figure 5.1 where the calculated reflectance spectrum of a SiO_2 layer with $d = 800$ nm on an alumina (Al_2O_3) substrate is plotted versus the wavelength. The task is to determine the thickness d from evaluation of the oscillating term. A quick and easy method for this is to apply the FFT to the measured reflectance spectrum. For a

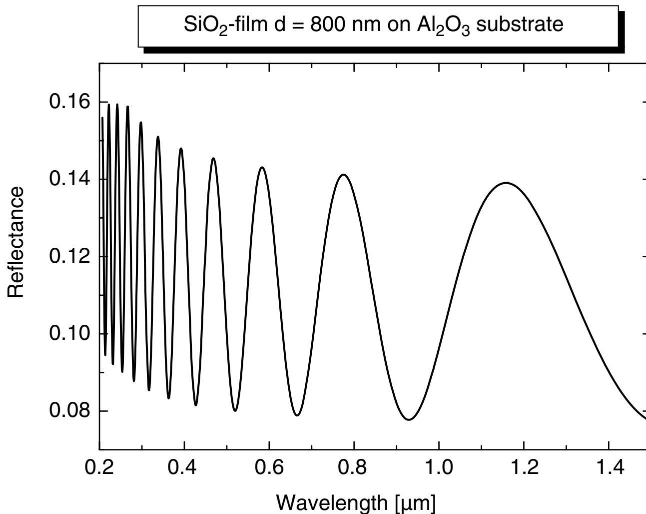


Figure 5.1 Calculated reflectance spectrum of a single SiO_2 layer with $d = 800$ nm on Al_2O_3 substrate.

successful application of the FFT, the layer thickness should be larger than approximately 1 µm. This thickness is a reasonable but not very distinct limit for the applicability of the FFT. As it depends upon the refractive index of the layer, for large refractive indices this limit may be even lower.

The spectrum in Figure 5.1 is aperiodic in the sense that the distance between two neighboring maxima is short at short wavelengths and becomes larger for longer wavelengths. The reason is that the cosine term in (5.3) depends upon $1/\lambda$ but it is plotted versus λ. Therefore, the first step in the FFT analysis is to rescale the spectrum by introducing a wavenumber $\nu = 1/\lambda$ and to plot the reflectance spectrum versus ν. The rescaled spectrum in Figure 5.2 exhibits now a periodic oscillation.

The next step is to interpolate the rescaled spectrum on N equidistant sampling points $\Delta\nu = (\nu_{max} - \nu_{min})/N = (1/\lambda_{min} - 1/\lambda_{max})/N$ because all FFT methods require the input data to be sampled at evenly spaced intervals. Moreover, N must be a power of 2 (radix-2 method), 4 (radix-4 method), 8 (radix-8 method), or higher multiples of 2. The reason is that the FFT algorithm divides the number of data points into two parts of length $N/2$. The higher the M of the radix-M method is, the faster the applied method is. On the other hand, the higher the M is the less possible powers of 2 are available for N. For example, for the radix-2 method N may be 256, 512, 1024, 2048, and 4096. For the radix-4 method, N can be 256, 1024, and 4096, but cannot be 512 or 2048 because these numbers are not a power of 4. It is recommended to use a number N larger than or equal to the number of sampling points in the original spectrum. Usually, the number of pixels of the spectrometer is a power of 2, so that N can be chosen identical to the number of pixels of the spectrometer or larger.

Step 3 is the FFT in effect. Introducing the parameter $t = n_1 d$ as optical thickness, the periodic oscillating spectrum looks like a $\cos(\nu \cdot t)$, which is a linear combination of $\exp(i\nu t)$ and $\exp(-i\nu t)$. The variable parameter is here ν, and we are interested in t.

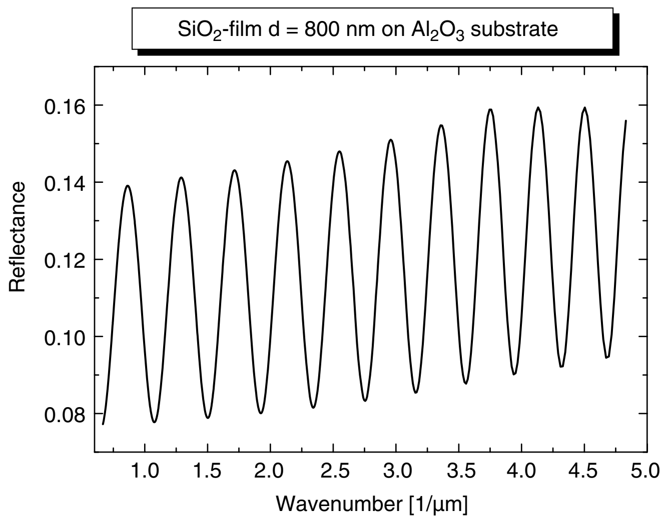


Figure 5.2 Rescaled calculated reflectance spectrum of a single SiO_2 layer on Al_2O_3 substrate.

Therefore, the Fourier transformation is

$$R(t) = \int_{-\infty}^{\infty} R(\nu) \cdot \exp(i\nu t) \mathrm{d}\nu \tag{5.4}$$

but applied to only N discrete values $R(\nu_j)$. The corresponding discrete Fourier transform (DFT) compares the reflectance spectrum with a sum over exponential terms $\exp(i\nu_j t) = \exp(i\nu_j n_1 d)$ with $\nu_j = j \cdot \Delta\nu$:

$$R(t) = C_0 + \sum_{j=1}^{N-1} C_j \cdot \exp(i(j \cdot \Delta\nu \cdot t)). \tag{5.5}$$

The coefficients C_j are complex numbers in general.

The advantage of the fast Fourier transform over this DFT is its speed that is achieved by an especially developed algorithm by Cooley and Tukey [133]. It reexpresses the discrete Fourier transform of size N in smaller discrete Fourier transforms. The special case with one DFT with the even-numbered points and the second with the odd-numbered points is often called *Danielson–Lanczos lemma* since this method was noted by those two authors in 1942 [134]. For more information on Fourier transform methods and the FFT, we refer to Appendix B.

The result of the FFT algorithm are wavenumbers $\nu_m = m \cdot \Delta\nu$ and $-\nu_m$ for which the absolute value $|C_m|^2$ of the complex Fourier coefficient C_m is maximum. From these wavenumbers, the layer thickness d is obtained as

$$d_{\mathrm{FFT}} = \frac{t}{n_1} = \frac{m}{2 \cdot n_1 \cdot (\nu_{\max} - \nu_{\min})} = \frac{m}{2 \cdot n_1 \cdot ((1/\lambda_{\min}) - (1/\lambda_{\max}))}. \tag{5.6}$$

Note that the FFT delivers the discrete integer number m. The boundaries $\nu_{\max} = 1/\lambda_{\min}$ and $\nu_{\min} = 1/\lambda_{\max}$ are given by the used spectrometer. The refractive index n_1 must be given by the user. In the above equation, it corresponds to a mean value in the spectral range, for example, the refractive index at the central wavelength of an SLD.

Usually, one plots the absolute value $|C_j|^2$ of the complex Fourier coefficient C_j versus the index j. The resulting spectrum is called *power spectrum*. The corresponding *power spectral distribution* (*PSD*) is obtained by dividing the power spectrum by the frequency step, which is here $1/(2 \cdot (\nu_{\max} - \nu_{\min}))$. The power spectrum of the reflectance spectrum of the silica film with $d = 800$ nm in Figure 5.1 is shown in Figure 5.3. For better presentation, the plot is restricted from $j = -50$ to $j = 50$ instead of from -1024 to 1024 (the number N of sampling points in the FFT was $N = 2048$). In this PSD, there is a peak at the indices $m = j = 10$ and $m = j = -10$ and at $j = 0$. The peak at $j = 0$ (value $|C_0|^2 = 241$) results from the background of the reflectance signal. From $m = 10$ and a mean refractive index $n_1 = 1.48$ for SiO_2 follows a thickness of 811.3 nm. This result deviates from the correct value of $d = 800$ nm used in the calculation of the reflectance spectrum by 1.4%. Although the FFT analysis is a quick and easy method for film thickness evaluation, it is incorrect to a certain extent. We will discuss this item in detail in the following.

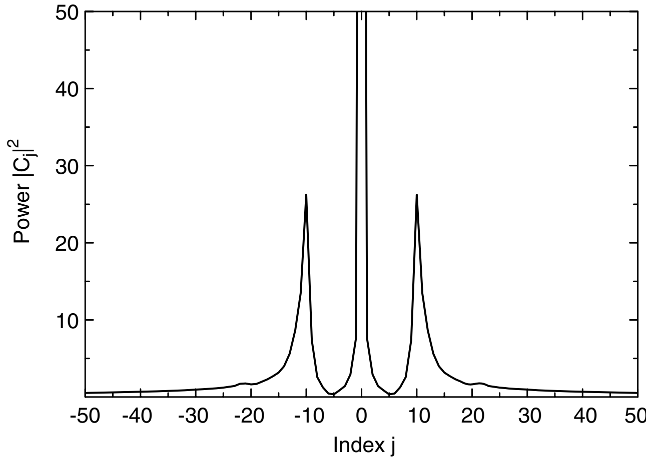


Figure 5.3 Power spectrum of the reflectance spectrum of the silica film in Figure 5.1.

By comparing (5.4) with usual formulations of the Fourier transform and discrete Fourier transform, we cannot recognize any special feature. However, as we have the refractive index $n_1(\nu)$ dependent upon ν, the optical thickness t is also dependent on ν, $t = t(\nu)$. This is different from the usual Fourier transforms, where t and ν are independent variables. To replace $n_1(\nu)$ by a mean value n_1 may be critical mainly in the case of high dispersion of the refractive index in the visible and near-ultraviolet spectral region. But if the dispersion also is almost negligible, as in the near-infrared region, the assumption of a constant refractive index is fallacious since it leads to inaccurate thickness values. This is demonstrated in Table 5.1. In this table, we summarized the results of the FFT analysis of calculated reflectance spectra of free-standing layers with thickness $d = 20, 50, 100, 200, 500$, and $800\,\mu m$ in the spectral range 1260–1360 nm using 1024 sampling points for the materials silicon (Si), sapphire (Al_2O_3), and fused silica (SiO_2). The optical constants were taken from [37, 38, 67] and were interpolated to the 1024 sampling points. The thickness range

Table 5.1 Thickness values, deviations, and effective refractive index from the application of the FFT analysis on calculated reflectance spectra of films of silicon, sapphire, and fused silica in the near infrared.

	Si n_1(1310 nm) = 3.503			Al_2O_3 n_1(1310 nm) = 1.75			SiO_2 n_1(1310 nm) = 1.4468		
d (μm)	d_{FFT} (μm)	$\Delta d/d$ (%)	n_{eff}	d_{FFT} (μm)	$\Delta d/d$ (%)	n_{eff}	d_{FFT} (μm)	$\Delta d/d$ (%)	n_{eff}
20	21.52	7.6	3.769	19.14	4.3	1.675	23.16	15.8	1.675
50	52.6	5.2	3.685	52.65	5.3	1.843	52.1	4.2	1.508
100	105.2	5.2	3.685	100.51	0.51	1.759	98.41	1.59	1.424
200	210.4	5.2	3.685	201.01	0.51	1.759	202.61	1.31	1.466
500	516.44	3.29	3.618	507.31	1.46	1.776	503.63	0.73	1.457
800	832.04	4.01	3.644	808.83	1.1	1.769	810.45	1.3	1.466

following from the spectral range is 8.57–4387 μm for $n_1 = 1$. The refractive index at 1310 nm (in the middle of the spectral range) amounts to $n(\mathrm{Si}) = 3.503$, $n(\mathrm{Al_2O_3}) = 1.75$, and $n(\mathrm{SiO_2}) = 1.4468$. As the obtained thickness values always differ from the correct value used in the calculations, we also give the deviation from the correct value in percentage and calculated an effective refractive index n_{eff} that should have been used in (5.6) to obtain the correct thickness value.

It becomes obvious from the values in Table 5.1 that the deviations are not unique neither with respect to the thickness nor with respect to the material. In the following, we discuss the reason for the deviations and how we can improve the FFT analysis for layer thickness determination.

First of all, the FFT computation assumes that a signal is periodic in the measurement data block. This means that the length of the data block is an integer multiple of the signal period. Only in this case does the FFT calculate exactly the frequency. Remember that in layer thickness determination the optical thickness t is the "frequency" and $\nu = 1/\lambda$ corresponds to the "time" variable. However, the measured reflectance signal is often not periodic with discontinuities at the borders of the interval (ν_{min}, ν_{max}). Then, the FFT calculates the portion of given frequencies t_m on the complete signal. Hence, it is almost impossible to determine the exact frequency. Moreover, the discontinuities in the signal cause pseudo-frequencies that distort the frequency spectrum (the power spectrum) and result in lower magnitudes of $|C(t_m)|^2$. This well-known effect is called *leakage*. A simple method to correct leakage is to use window functions. A window is shaped so that it is exactly zero at the beginning and end of the data block and has some special shape in between. FFT windows reduce the effects of leakage but cannot eliminate leakage entirely. In effect, they do not affect the frequencies t_m, but weight the magnitudes $|C(t_m)|^2$ of each frequency t_m in the power spectrum anew, so that the frequency with the maximum magnitude can be better found in the power spectrum. Commonly used window functions are Hamming, von Hann (or Hanning or raised cosine), Blackman, Blackman–Harris, Welch, or Kaiser window. For an overview on window functions, we refer, for example, to Ref. [135]. For more detailed reading on FFT and concepts of the FFT, we refer to Ref. [136].

As the reflectance data are available only in the interval (ν_{min}, ν_{max}), they are automatically convoluted by a window function, the rectangular window rect(ν_{min}, ν_{max}). The Fourier transform of $R(\nu)\cdot\mathrm{rect}(\nu_{\mathrm{min}}, \nu_{\mathrm{max}})$ is $R(t)\cdot\mathrm{sinc}(t\cdot(\nu_{\mathrm{max}} - \nu_{\mathrm{min}}))$, meaning that the Fourier transformed data (the frequencies t_m) are convoluted by the sinc function $\sin(x)/x$, resulting in a broadening of the frequency domain. This broadening complicates the search for the number m where $|C_m|^2$ is maximum because neighboring points $m + j$ and $m - j$ may also exhibit a maximum in the power spectrum.

The search for the maximum magnitude in the power spectrum may be improved by determination of the center of gravity (COG):

$$\bar{m} = \frac{\sum_{k=m-M}^{m+M} k \cdot |C_k|^2}{\sum_{k=m-M}^{m+M} |C_k|^2}. \tag{5.7}$$

This is of interest particularly for asymmetric peaks or peaks where two neighboring coefficients have almost the same magnitude $|C|^2$. This method depends on the number M of neighboring points at left and right of the point m.

The biggest improvement of the FFT algorithm for layer thickness determination is given when $\cos((4\pi/\lambda)n(\lambda)d) = \cos(4\pi\nu \cdot t(\nu))$ matches almost exactly the measured spectrum periodically. But how to achieve this situation with a refractive index exhibiting dispersion? Remember that the minimum thickness $d_{\min}$ (Equation 5.1) and the maximum thickness $d_{\max}$ (Equation 5.2) of a film with refractive index $n(\lambda)$ are also determined by the refractive indices $n(\lambda_{\min})$ and $n(\lambda_{\max})$. These values are not included in (5.6). The idea of an improvement is to modify the FFT analysis in that way, that after determining the index m in the above classical way, the values $n_1(\lambda_{\min})$ and $n_1(\lambda_{\max})$ are used in (5.6) instead of a mean constant value:

$$d_{\mathrm{FFT}} = \frac{m}{2 \cdot (n_1(\lambda_{\min})/\lambda_{\min} - n_1(\lambda_{\max})/\lambda_{\max})}. \qquad (5.8)$$

This modification of the FFT can be combined with the COG.

A quite similar idea is to consider the dispersion of the refractive index $n_1(\lambda)$ already when rescaling the measured spectrum using the variable $\nu = n_1(\lambda)/\lambda$ instead of $\nu = 1/\lambda$. After rescaling, the spectrum must be interpolated on evenly spaced points in the interval $\nu_{\min} = n_1(\lambda_{\max})/\lambda_{\max}$ to $\nu_{\max} = n_1(\lambda_{\min})/\lambda_{\min}$ before the FFT is used to find the index m. The thickness is then also obtained by (5.8). The difference to the proposal above is that due to the different rescaling, the input data of the FFT may slightly be different that may affect the index m. This method can also be combined with the COG.

To prove which of these methods yields an improvement in the thickness determination with FFT, we calculated again the thickness of silicon, sapphire, and fused silica films, using the following models:

- Model M0 = unmodified FFT analysis.
- Model M1 = FFT analysis using (5.8) for thickness determination.
- Model M2 = FFT analysis considering the full dispersion and using (5.8) for thickness determination.

The results are summarized in Table 5.2. The refractive indices of silicon, sapphire, and fused silica at the wavelengths $\lambda_{\min} = 1260$ nm and $\lambda_{\max} = 1360$ nm are tabulated as follows:

Si:	n_1(1260 nm) = 3.510	n_1(1360 nm) = 3.496
Al_2O_3:	n_1(1260 nm) = 1.7507	n_1(1360 nm) = 1.7493
SiO_2:	n_1(1260 nm) = 1.4474	n_1(1360 nm) = 1.4462

We find a significant improvement in the thickness results when using (5.8), especially for the highly refractive silicon. For sapphire and fused silica, the smallest thickness of $d = 20$ μm still exhibits larger errors, but this thickness is very close to the lower boundary of the range of resolvable thicknesses (8.57 μm/n_1). Therefore, for

Table 5.2 Thickness values from the application of the FFT and its improvements applied to calculated reflectance spectra of wafers of silicon, sapphire, and fused silica in the near infrared.

d (μm)	M0 d_{FFT} (μm)	M1 d_{FFT} (μm)	M2 d_{FFT} (μm)
	Silicon		
20	21.52	20.45	20.45
50	52.6	49.999	49.999
100	105.2	99.999	99.999
200	210.4	199.98	199.98
500	516.44	490.90	499.98
800	832.04	790.90	724.993
	Sapphire		
20	19.14	18.94	18.94
50	52.65	52.08	52.08
100	100.51	99.42	99.42
200	201.01	198.84	198.84
500	507.31	501.84	501.84
800	808.83	800.11	800.11
	Fused silica		
20	23.16	22.92	22.92
50	52.1	51.57	51.57
100	98.41	97.40	97.40
200	202.61	200.54	200.54
500	503.63	498.47	498.47
800	810.45	802.14	802.14

these materials and this small thickness, another spectral range is recommended. A problem arose with M2. For sapphire and fused silica the explicit consideration of the dispersion of the refractive index yields the same results as M1. Also for silicon, it yields the same or even better results, except for the 800 μm silicon wafer. The reason is unknown yet, but the power spectrum is completely widespread and has a maximum at the wrong position. In general, M2 resulted in very noisy power spectra, also for sapphire and fused silica. Therefore, in our opinion, M1 is the favorable method to determine the thickness d of the film. For illustration, we show in Figure 5.4 the power spectra for the wafers with $d = 100$, 500, 750, and 800 μm, resulting from models M0 and M2.

The center of gravity was checked with the result that only for some few examples the thickness was obtained closer to the exact value, but without clear indication how large the number M of neighboring points must be chosen.

The use of a window function was also checked. They only partially yielded better results. Particularly for silicon with thickness $d = 500$ μm, the result of M1 of approximately 491 μm could be improved to approximately 502 μm, and for silicon with $d = 800$ μm the result of M1 of approximately 791 μm could be improved to approximately 793 μm. All other results of M0, M1, and M2 kept unchanged.

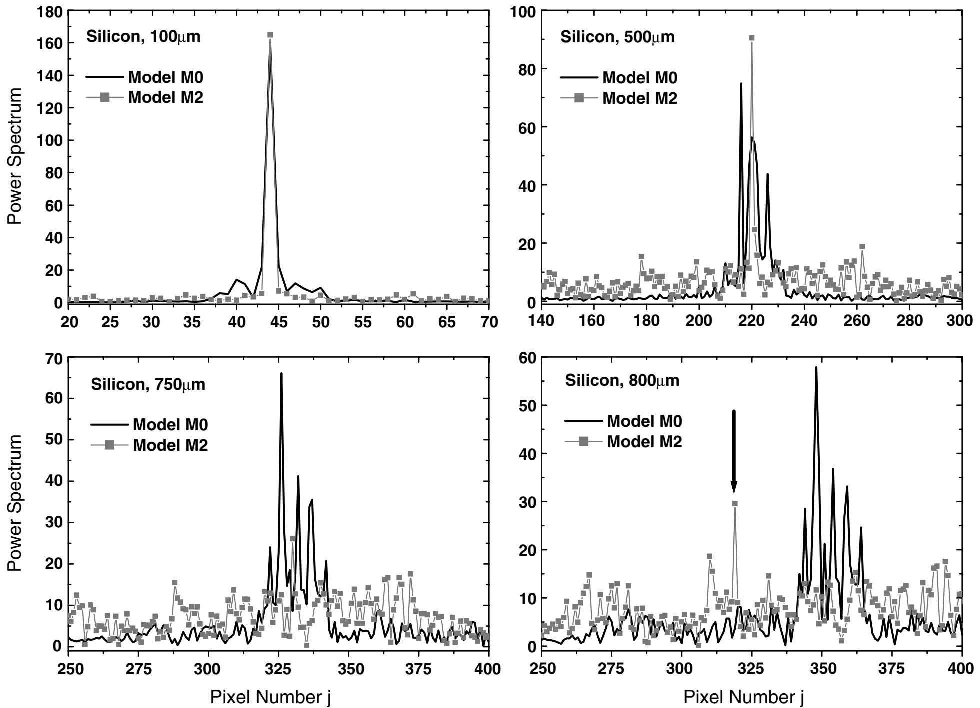


Figure 5.4 Power spectra of a silicon wafer with $d = 100$, 500, 750, and 800 μm, resulting from the models M0 and M2.

5.1.2 Layer Stack

If we have more than one thick layer, we can also expect more peaks in the power spectrum. For two layers, one gets three peaks, one for each optical layer thickness and one for the sum of optical thicknesses of the two layers. For three layers, one gets already seven peaks, as also all sums of two optical layer thicknesses result in a peak. The task is then to assign the peaks to a certain layer. This can be done manually or automatically if the actual thickness of each layer is approximately known. For completely unknown thicknesses, this assignment is difficult and is doubtless only for maximum two layers in the stack.

For two layers, the FFT finds two values m_1 and m_2. They can be inserted into (5.6), together with two mean refractive indices n_1 and n_2 for the two layers. Then, however, as shown above, errors of up to about 5% can be expected, depending on the layer material. A significant improvement is again obtained if for each layer the modified FFT algorithm M1 is applied correspondingly.

In Figure 5.5, we give an example for the reflectance spectrum and the corresponding power spectrum from FFT for a thick and a thin layer on a substrate. The reflectance spectrum is characteristically modified by the film thickness interference

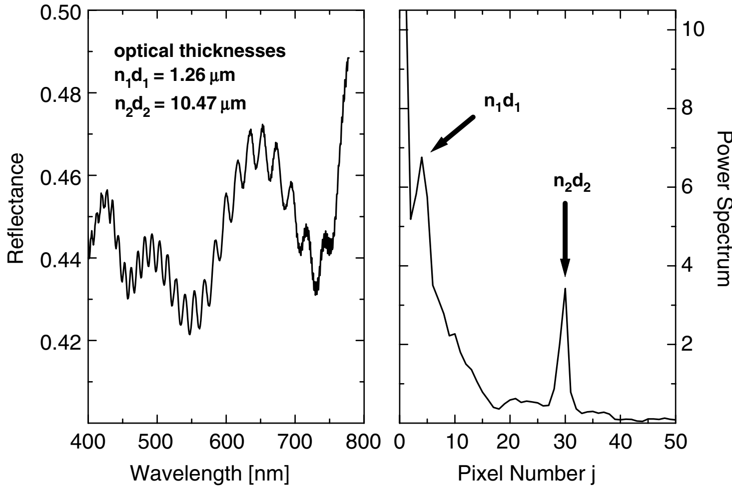


Figure 5.5 Example for a thin and a thick layer. The optical thicknesses of the two layers are $n_1d_1 = 1.26\,\mu m$ and $n_2d_2 = 10.47\,\mu m$. (a) shows the reflectance and (b) shows the corresponding power spectrum from the FFT.

patterns. The optical thicknesses of the two layers are $n_1d_1 = 1.26\,\mu m$ and $n_2d_2 = 10.47\,\mu m$. The thicker layer results in fast oscillations that are superimposed by the slower oscillations due to the thin film. The power spectrum exhibits two maxima at different pixel numbers. They can be assigned to the single optical thicknesses. The sum of the two optical thicknesses is not resolved.

5.1.3
Accuracy, Resolution, Repeatability, and Reproducibility

All optical measurements of the thickness of a thin film or layer are inaccurate to a certain amount. To determine the accuracy, it is necessary to have official standards for film thickness determination. While for layers with thickness d up to $d = 1\,\mu m$, official standards exist (e.g., from the Physikalische Technische Bundesanstalt PTB (National Bureau of Standards of Germany)) or standards are available that are traceable back to an official standard, the situation is much more complicated for thicknesses larger than $1\,\mu m$. Recently, standards have been manufactured by Precitec Optronics GmbH, Germany, that utilize an air gap between an optical flat and a gauge block. The thickness of the air gap is certified by the PTB. These standards are available with different air gaps from $d = 20\,\mu m$ to $d = 500\,\mu m$. These thicknesses require an evaluation of the measured reflectance spectra by fast Fourier transform analysis. Unfortunately, to our knowledge, a corresponding standard is still missing for the range from $d = 1\,\mu m$ to $d = 20\,\mu m$.

As already shown in Section 5.1, the accuracy of the FFT analysis does not only depend on errors in the measurement but also depend on how good the measured signal fits in the measuring window to minimize the leakage problem. Nevertheless,

as leakage is almost always present, the accuracy is limited. Typically, if the leakage correction proposed in Section 5.1.1 as model M1 is used, the accuracy is better than 2% of the thickness value for high refractive materials (e.g., semiconductor materials Si, GaAs, GaN, etc.). For materials with lower refractive index, the accuracy is in the range of 1% or even better.

The resolution in thickness determination of thin films with an FFT algorithm is given by the smallest step Δd:

$$\Delta d_{\text{Resolution}} = \frac{1}{4}\Delta d = \frac{1}{8\cdot(n(\lambda_{\min})/\lambda_{\min} - n(\lambda_{\max})/\lambda_{\max})}. \tag{5.9}$$

From various examinations on the repeatability and reproducibility of the thin-film determination with FFT, it follows that the surface roughness has influence on the result to a certain extent. In dependence on the surface roughness, the repeatability and reproducibility are in the range of $\sigma_{\text{Repeatability}} = 60$–$120\,\text{nm}$ and $\sigma_{\text{reproducibility}} = 150$–$400\,\text{nm}$.

5.2 Regression Analysis with χ^2-Test

5.2.1 Method of Thickness Determination

The regression analysis used for layer thickness determination fits a given set of measured data (spectrally resolved ellipsometric parameters, spectral reflectance) to a model that depends upon the thickness of the layers, the optical constants of the layers, and roughness parameters as adjustable parameters. The basic approach is to choose a figure-of-merit function that measures the agreement between the data and the model. Small values of the merit function represent close agreement. The parameters of the model are automatically adjusted so that the merit function achieves a minimum, yielding best-fit parameters. However, the merit function often has several local minima. Then, the task is not only to find a minimum but also to find the global minimum. Otherwise, the determined parameters of the layers may be senseless.

Any fitting procedure should provide

- parameters,
- error estimates on the parameters, and
- a statistical measure of goodness of fit.

If each data point (x_i, y_i) of a set of N measured data has its own, known standard deviation σ_i, then the maximum likelihood estimate of the model parameters $a_1, \ldots, a_M$ is obtained by minimizing the quantity chi-square χ^2

$$\chi^2 = \sum_{i=1}^{N}\left(\frac{y_i - f(x_i, a_1, \ldots, a_M)}{\sigma_i}\right)^2 \tag{5.10}$$

or the mean squared error (MSE)

$$\mathrm{MSE} = \sqrt{\frac{\chi^2}{2N-M}} \tag{5.11}$$

by iterative nonlinear regression, with $f(x, a_1, \ldots, a_M)$ being the function that calculates either the ellipsometric parameters or the reflectance and/or transmittance of the layer system with the M parameters $a_1, \ldots, a_M$. A smaller χ^2 or MSE implies a better model fit to the data.

Various methods are available to find a minimum of χ^2. Unfortunately, there is no perfect algorithm. Two methods are often used in the regression analysis for layer stacks, the *downhill simplex algorithm* due to Nelder and Mead [137] and the *Levenberg–Marquardt algorithm* [138, 139].

The downhill simplex method requires only function evaluations. A simplex is a geometrical figure consisting of $M + 1$ points and all interconnecting line segments. For example, if $M = 2$, the simplex is a triangle, for $M = 3$ it is a tetrahedron (not necessarily the regular tetrahedron). The minimization starts with an initial guess for the parameters and calculates $M + 1$ values of χ^2 as the points of the initial simplex. Then, the routine tries to find automatically the minimum volume inscribed by the simplex by replacing the worst of the $M + 1$ points by a new one that is better. The best point of any iteration is kept. If χ^2 falls of or equals a certain value, the calculation stops and the set of parameters with the best fit to the measured data is found. The downhill simplex algorithm can be used only for fitting of the thickness of the layers in the multilayer stack. A parameterization of optical constants of a layer according to any of the models presented in Chapter 2.7 is not planned, but they must be provided in tables. For a more detailed reading on the downhill simplex algorithm, we refer to Press *et al.* [140] and to Appendix C.

The Levenberg–Marquardt algorithm combines minimization of χ^2 with a Hessenberg matrix algorithm with the search for the steepest descent. For that purpose, it uses the first derivative of χ^2 versus each parameter a_i, $\partial\chi^2/\partial a_i$. This may be a long lasting method if the derivatives cannot be given as formula but must be calculated by variation of the parameter and calculating the differential quotient numerically, as in multilayer stack analysis. On the other hand, it allows fitting of many different parameters, not only layer thicknesses but also the parameters that describe optical constants in a corresponding model (Section 2.7). Therefore, this algorithm is preferably used in ellipsometry for determination of layer thicknesses, optical constants, and other parameters. It works well as long as the number of data points N in the measurement is sufficiently larger than the number of parameters M, that is, $N >> \mathrm{M}$. Again, if χ^2 falls off or equals a certain value, the calculation stops and the set of parameters with the best fit to the measured data is found. For more information, we refer to Press *et al.* [140] and to Appendix D.

Such a regression analysis requires the correct model to achieve good fits to the experimental data. Otherwise, the model needs to be reformulated. If, for example, the initial guess for a thickness value d is too far from the actual value, the regression algorithm may yield an erroneous value for the parameter. The reason is that χ^2 exhibits more than one minimum when varying the thickness d. In Figure 5.6a–c, we

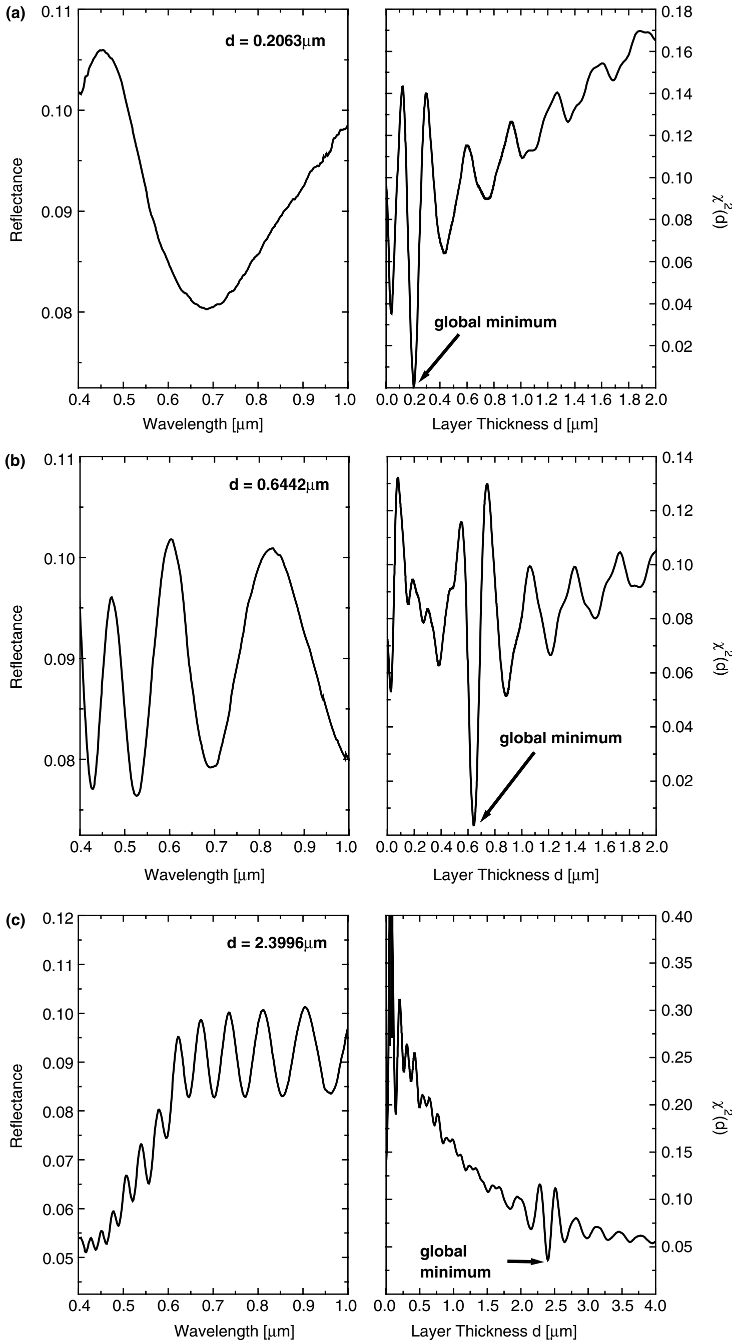


Figure 5.6 Measured reflectance spectra and $\chi^2(d)$ curves of a photoresist spin-coated on a glass with three different thicknesses: (a) $d = 0.2063\,\mu m$, (b) $d = 0.6442\,\mu m$, and (c) $d = 2.3996\,\mu m$. The thickness was obtained by a regression analysis with upstream FFT.

show on the left side of each graph the measured reflectance spectrum of a photoresist that was spin-coated on glass. Due to the different rotational speed, the thickness of the layer was different. They were determined to $d = 0.2063\,\mu m$, $d = 0.6442\,\mu m$, and $d = 2399.6\,\mu m$. On the right side of Figure 5.6a–c, we plotted χ^2 in dependence on the thickness d. Starting with $d = 0.001\,\mu m$, χ^2 exhibits several local minima, but only one minimum is the deepest. The regression algorithm must find this global minimum to determine correctly the thickness of the layer. To ensure that the true global minimum χ^2 is found, the regression algorithm can be started with a wide range of initial parameter values. This procedure is, however, time consuming and cannot be used for automated measurements. Therefore, both for the downhill simplex method and for the Levenberg–Marquardt algorithm, it is necessary to have a good estimation of the thickness of the layers that are to be determined. For example, the used downhill simplex algorithm implemented in MQLayer [8] finds the value $d = 0.6442\,\mu m$ for the photoresist in Figure 5.6b when the starting value is in the interval (0.52, 0.715 μm). Lower starting values between 0.4 and 0.52 μm lead to $d \approx 0.384\,\mu m$. Higher values between 0.715 and 0.85 μm lead to $d \approx 0.884\,\mu m$. These values belong to the next neigboring local minima in χ^2 as can be recognized from Figure 5.6b.

We used a third, self-developed algorithm for thickness evaluation (also implemented in MQLayer). It first applies an FFT to find a thickness already close to the thickness for which χ^2 has its global minimum. Then, a regression analysis iteratively finds the correct thickness value. If the upstream FFT cannot be applied because less than one oscillation can be recognized in the reflectance spectrum, this algorithm searches iteratively for the global minimum in the interval from zero to four times the minimal thickness available from FFT.

One important question concerns the lowest detectable layer thickness. This question cannot be answered by giving one single number. Instead, one has to consider both the layer properties and the substrate properties in detail. This will be representatively demonstrated for the reflectance measurement for the examples of a silica layer (SiO_2), an alumina layer (Al_2O_3), and a ZnO:Al layer (aluminum-doped zinc oxide). From SiO_2 to ZnO:Al, the refractive index increases. Moreover, ZnO:Al is absorbing in the spectral range below 530 nm wavelength. We have summarized some values of the (complex) refractive index in Table 5.3. The layers are supported by a silicon wafer (optical constants from Ref. [37]), a glass plate of N-BK7 (optical constants from Ref. [66]), or an aluminum mirror (optical constants from Ref. [38]). So, we have a variety of materials with different optical contrast. Optical constants of ZnO:Al were submitted from FHR Anlagenbau GmbH, Germany.

We start our discussion in Figure 5.7 with silicon as substrate. The smallest detectable thickness is here determined by the reduction in the reflectivity of pure silicon. The reflectivity gets reduced since the refractive index of the used oxides is less than that of silicon. This reduction can be best recognized in the UV spectral region where the two interband transitions in silicon result in the two maxima in the reflectivity at 272 and 364 nm. Only in this spectral range is the change significantly high in the reflectivity of silicon, caused by an oxide layer with 10 nm thickness, to

Table 5.3 Refractive indices of SiO_2 [38], Al_2O_3 [67], and ZnO:Al at various wavelengths.

λ (nm)	$n(SiO_2)$	$n(Al_2O_3)$	n(ZnO:Al)
200	1.5519	1.8350	$1.98 + i0.28$
300	1.4878	1.8143	$2.0045 + i0.552$
500	1.4624	1.7746	$1.972 + i0.003$
700	1.4553	1.7635	1.861
900	1.4518	1.7578	1.775

detect such a thin film. The higher the refractive index is, the more distinct is the difference in the reflectivity. We can conclude that for the materials with even higher refractive index than the refractive index of ZnO:Al also even thinner layers may be detectable, but 10 nm is a reasonable lower limit. The reason for these changes is the beginning evolution of film thickness oscillations due to interference.

When turning to glass substrates as in Figure 5.8, the situation is different. SiO_2 has a lower refractive index than the glass substrate. Therefore, a silica film again reduces the reflectivity of the substrate, similar to silicon as substrate. However, as the reflectivity of the glass does not exhibit any characteristic features, the thickness of the silica layer must now be increased to $d = 20$ nm to get a detectable change of the reflectivity. In contrast, the films of Al_2O_3 and ZnO:Al with refractive index higher than that of the glass N-BK7 increase the reflectivity. For ZnO:Al even a maximum at 323 nm appears, which is caused by the absorption of light in ZnO:Al. For these materials with higher refractive index than the refractive index of the substrate, a thickness of $d = 10$ nm is sufficient to detect the change in the reflectivity in the UV.

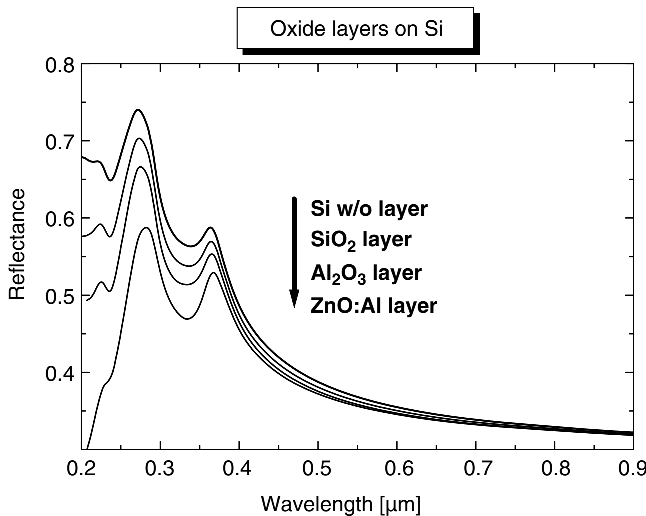


Figure 5.7 Thin oxide layers on silicon substrate. The thick line on top belongs to pure silicon.

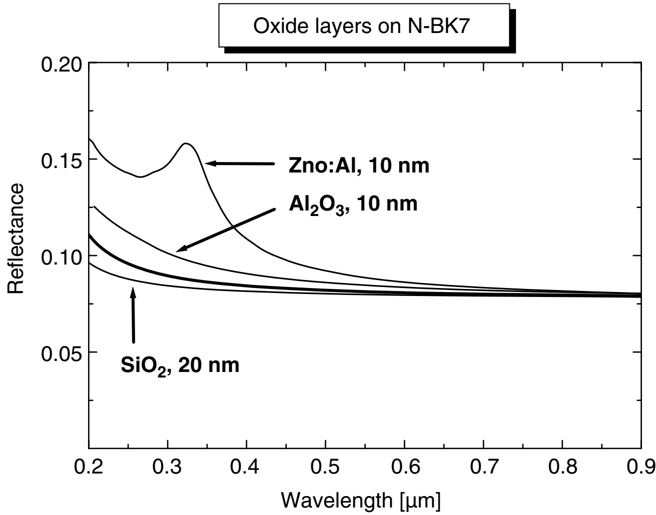


Figure 5.8 Thin oxide layers on N-BK7 glass substrate. The thick line belongs to the pure glass.

Using a highly reflective aluminum mirror as substrate, the reflectivity decreases for all layers. This can be seen from Figure 5.9. The most striking change in the reflectivity occurs for the absorbing ZnO:Al layer. A film of 10 nm ZnO:Al reduces the reflectivity down to 40 % at 200 nm wavelength. For other oxide layers, the effect is less striking, but a thickness of 10 nm is sufficient for a detectable change in the reflectivity.

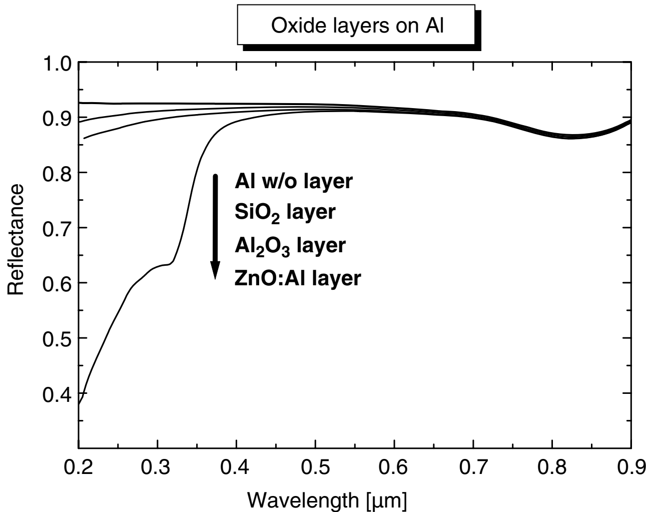


Figure 5.9 Thin oxide layers on aluminum substrate. The thick line on top belongs to the pure aluminum mirror.

From the above examples, we conclude that the minimum detectable thickness of transparent layers is between 10 and 20 nm. However, this result is valid only in the UV part of the spectral range. At longer wavelengths in the visible or even in the near infrared, such thin films do obviously not cause significant changes in the reflectivity of the pure substrate. As the ratio thickness/wavelength $= d/\lambda$ is relevant for the evolution of thickness oscillations, we can deduce that changes in the reflectivity will occur in the visible spectral range for layers with thicknesses between 20 and 50 nm, and must be further increased to 70–100 nm when going to the near-infrared spectral region. If absorption in the layer occurs, as in the case of aluminum-doped zinc oxide, the change in the reflectivity is often more striking, allowing for determination of thinner films.

The example SiO_2 on N-BK7 additionally demonstrates that the optical contrast is also of importance for the recognition of a thin layer on a substrate. While the contrast of the layer to the ambient air or vacuum as front medium ($n_1 - n_0$) is almost always high enough, the contrast between layer and substrate ($n_2 - n_1$) may be very small. As this contrast enters the reflection coefficient of the interface layer–substrate, the reflection at this interface may be very poor so that, in fact, the interference of this reflected beam with the beam reflected on top of the layer may decrease in its magnitude below the signal-to-noise ratio. Then, this interference may be undetectable. A well-established thumb rule is that the optical contrast ($n_{k+1} - n_k$) at the interface between two media k and $k + 1$ must be $(n_{k+1} - n_k) \geq 0.1$.

In contrast to the reflectance measurement, the spectral ellipsometric measurement is sensitive down to 0.1 nm film thickness. The reason is that ellipsometry uses a nonzero incident angle for which the reflectivity becomes polarization dependent. Changes in the polarization can be measured with a high accuracy and resolution. A change of 0.1 nm in the thickness causes a phase shift of 0.25°.

5.2.2 Accuracy, Resolution, Repeatability, and Reproducibility

The accuracy, resolution, repeatability, and reproducibility of the thickness determination with a regression analysis are different from that of the FFT.

To determine the accuracy, it is necessary to have official standards for film thickness determination. For layers with thickness d up to $d = 1$ μm, official standards exist (e.g., SiO_2 layers on Si from the PTB) or standards are available that are traceable back to an official standard. They are often certified by an ellipsometric measurement within a certain wavelength range. Hence, when measuring a reflectance spectrum of these coatings, one has to consider the following points:

- The best fit to the measured spectrum must fit simultaneously at many sampling points. If the number of sampling points differs from that in the certification process, this can result in another thickness than the certified one.
- The errors in the used optical constants will affect the thickness determination, as other sets of optical constants than those used for specification can lead to different results.

- When using a spectral range that not includes the wavelengths used for the certification of the standard, problems may arise because the (n, k) data are perhaps only partly available. Moreover, for longer wavelengths the minimal detectable thickness will shift to larger sizes.

These discrepancies mainly affect the accuracy of the thickness determination in both spectral ellipsometry and spectral reflectance measurement. It may, therefore, happen that the mean value of a series of repeated measurements and evaluations has an excellent standard deviation for repeatability and reproducibility but lies outside the tolerance interval of the official standard. This is demonstrated with the values in Table 5.4 for silica layers on silicon.

The spectral reflectance measurements in the range 360–1000 nm were carried out with a spectrometer with 2048 pixel. In this case, the number of sampling points was strongly extended compared to the ellipsometric measurement from the PTB and the spectral ranges do not match completely. The spectral reflectance measurements in the range from 900 to 1700 nm were carried out with an NIR spectrometer with 512 pixel. Here, the spectral range does not at all match the spectral range of the PTB measurements, and optical constants were available only from continuation of data from literature to the NIR.

The obtained thickness values clearly demonstrate that the deviation in the thickness may be larger than the tolerance of the thickness value of the standard. Nevertheless, the obtained results for the accuracy are typical and lie in the range of values less than 1% over the full thickness range, but at least 1 nm.

For the long wavelengths in the range 900–1700 nm (near-infrared region), the thinnest SiO_2 layer becomes critical. Preferably, the thinnest layer should be about 80 nm in thickness to have an accuracy of less than 1% in this spectral range.

The repeatability and the reproducibility of a reflectometric thickness measurement of nanometer-thick films are pretty high. One can achieve values of $\sigma_{\text{repeatability}} \leq 0.1$ nm and $\sigma_{\text{reproducibility}} \leq 0.3$ nm over a wide spectral and thickness range. The actual values in the measurements for Table 5.4 were $\sigma_{\text{repeatability}} = 0.06$ nm and $\sigma_{\text{reproducibility}} = 0.23$ nm for 25 repeated measurements.

For derivation of the thickness resolution, one has to resolve (5.3) for the thickness d. The result is

Table 5.4 Results of the thickness determination from regression analysis of measured reflectance spectra in comparison to the values of standards from PTB, Germany.

PTB standard 300–850 nm	**Spectral reflectance 360–1000 nm**	**Spectral reflectance 900–1700 nm**
66.5 ± 0.6 nm	66.8 nm	70.3 nm
160.8 ± 0.7 nm	163.6 nm	160.9 nm
381.5 ± 0.9 nm	382.6 nm	383.3 nm
1000.4 ± 2.1 nm	997.5 nm	998.3 nm

$$d = \frac{\lambda}{4\pi n(\lambda)} \cos^{-1}\left(-\frac{R(\lambda) \cdot C(\lambda) - A(\lambda)}{R(\lambda) \cdot D(\lambda) - B(\lambda)} \right) \tag{5.12}$$

with $A(\lambda)$, $B(\lambda)$, $C(\lambda)$, and $D(\lambda)$ being abbrevations for the terms with the reflectivities of the interfaces. The arc cosine function cannot exceed 2π so that we can replace it by 2π. Then, the resolution becomes

$$\Delta d \leq \frac{\Delta\lambda}{2 \cdot n(\lambda)} = \frac{3 \cdot (\lambda_{\max} - \lambda_{\min})}{2 \cdot n(\lambda_{\max}) \cdot N}. \tag{5.13}$$

Here, we used the relation between the wavelength resolution according to the Rayleigh criterion and the pixel dispersion. It can take values of $\Delta d \leq 1$ nm for spectrometers working in the UV/VIS region and $\Delta d \leq 2$ nm for spectrometers in the NIR. These values are valid for $n = 1$ and must be divided by the corresponding refractive index n of the film.

The determination of the complex refractive index $n + i\kappa$ from reflectance or transmittance measurements is not independent of the thickness determination. All used methods simultaneously solve for thickness d and $n + i\kappa$, in spectral reflectance and ellipsometry measurements.

Often, photometric measurements of transmittance T and reflectance R at normal incidence are used for determination of the complex refractive index of a thin transparent film on a transparent substrate (e.g., [142–148]). The optical constants are determined as solutions of the implicit system

$$T_{\text{calc}}(n, \kappa, d, \lambda) - T_{\text{meas}}(\lambda) = 0, \tag{5.14}$$

$$R_{\text{calc}}(n, \kappa, d, \lambda) - R_{\text{meas}}(\lambda) = 0. \tag{5.15}$$

To solve this system, it is necessary to know the refractive index of the substrate and the film thickness d. Unfortunately, this system is ambiguous, and sometimes a physical criterion must be applied to identify the physical solution. The general problems with R–T methods have been minimized or solved by using improved algorithms [149–152], measurements of R and T at oblique incidence [153, 154], or other additional measurements with improved evaluation algorithms [155–160].

The R–T method cannot be applied to samples with an opaque substrate. Then, only reflectance can be measured. In that case, the determination of the refractive index needs a parameterization of the refractive index with a suitable model, like the models described in Chapter 2. With this parametrization the number of unknown parameters gets drastically reduced, which allows simultaneous determination of the wavelength dependence of the refractive index and the thickness of a thin layer on a substrate using a regression analysis. In the following, we present the results of simultaneous determination of film thickness and optical constants of SiO_2 layers on Si on the PTB standards and discuss the accuracy of this method. The optical constants of the substrate silicon are taken from Ref. [37] and have been kept fixed, while the optical constants of SiO_2 have been modeled with a three-term Sellmeier and a Cauchy formula. The thickness results for the two layers with $d = 160.8$ nm and

Table 5.5 Results of the simultaneous thickness and optical constants determination from regression analysis of measured reflectance spectra in comparison to the values of standards from PTB, Germany.

PTB standard	Sellmeier model	Cauchy model	Fit with (*n*,*k*) from Ref. [38]
160.8 ± 0.7 nm	162.5 nm	162.6 nm	163.6 nm
381.5 ± 0.9 nm	380.0 nm	380.0 nm	382.6 nm
1000. 4 ± 2.1 nm	999.6 nm	998.8 nm	997.5 nm

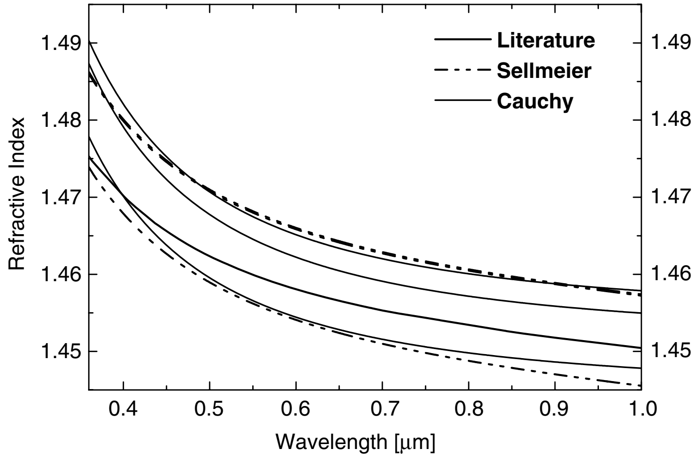


Figure 5.10 Comparison of the optical constants of SiO_2 from a regression analysis with Levenberg–Marquardt algorithm of measured reflectance data using a Sellmeier and a Cauchy parameterization.

$d = 381.5$ nm were compared with each other and with the PTB standard values in Table 5.5. The optical constants were compared in Figure 5.10 with that of SiO_2 from Ref. [38]. The obtained curves for the optical constants $n + i\kappa$ deviate from the SiO_2 data from Ref. [38] by maximum $\Delta n = 0.015$. Both fits result in a stronger increase in the refractive index than the values from literature.

6
The Color of Thin Films

Thin films of transparent or semitransparent materials do not only affect the properties of technical surfaces but also can excite human perception by the color resulting from the thickness interference. Examples are the iridescent colors of a peacock feather, the impressive colors of lustrous butterfly wings, or the colors of thin oil films on water. This chapter is, therefore, devoted to the apparent color of thin films.

In general, unlike mass, volume, or temperature, color is not merely a physical property of an object. It is rather a sensation triggered by radiation of sufficient intensity. The sensation of color depends not only on physical laws but also on the physiological processing of the radiation in the human eye. Colorimetry is the study of the dimensional relations between colors. It assumes that colors can be described by dimensional figures and that these figures can be measured.

Assuming $\phi(\lambda)$ is the spectral color stimulus function, that is, the measured spectral response of an examined sample, it can be used to calculate the apparent chromaticity coordinates x and y and the tristimulus values X, Y, and Z that results from that spectral response. $\phi(\lambda)$ may be the reflectance, the remittance, or the transmittance of the sample.

In accordance to the proposals of the Commission Internationale de l'Éclairage (CIE, International Commission on Illumination) the *tristimulus values* X, Y, and Z are defined by [161, 162]

$$X = F\int_{\lambda_1}^{\lambda_2} \phi(\lambda)\cdot S(\lambda)\cdot \bar{x}(\lambda)\cdot \mathrm{d}\lambda, \tag{6.1}$$

$$Y = F\int_{\lambda_1}^{\lambda_2} \phi(\lambda)\cdot S(\lambda)\cdot \bar{y}(\lambda)\cdot \mathrm{d}\lambda, \tag{6.2}$$

$$Z = F\int_{\lambda_1}^{\lambda_2} \phi(\lambda)\cdot S(\lambda)\cdot \bar{z}(\lambda)\cdot \mathrm{d}\lambda, \tag{6.3}$$

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

with the normalization factor F defined as

$$F = \frac{100}{\int_{\lambda_1}^{\lambda_2} S(\lambda) \cdot \bar{y}(\lambda) \cdot d\lambda}, \tag{6.4}$$

so that for $\phi(\lambda) = 1$, the value of Y amounts to $Y = 100$. The boundaries λ_1 and λ_2 in the integrals correspond to the onset wavelength $\lambda_1 = 380$ nm and the offset wavelength $\lambda_2 = 780$ nm of the visible spectral region. The quantities $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$ are the *spectral tristimulus values* (color matching functions), which depend upon the standard colorimetric observer (2° (CIE 1931) and 10° (CIE 1964)). Finally, $S(\lambda)$ is the spectral power distribution of a light source. Common light sources for colorimetry are the standard illuminants A, C, D_{50}, D_{55}, D_{65}, and D_{75}. The standard illuminant A corresponds to the black body radiation at absolute temperature $T = 2856$ K, the standard illuminant C corresponds to a daylight phase that is comparable to the black body radiation at absolute temperature $T = 6774$ K, and the standard illuminants D_{nn} correspond to a daylight phase that are comparable with the black body radiation at absolute temperature $T = nn^* 100$ K.

From the tristimulus values X, Y, and Z, the *chromaticity coordinates* x and y are obtained as

$$x = \frac{X}{X + Y + Z}, \tag{6.5}$$

$$y = \frac{Y}{X + Y + Z}. \tag{6.6}$$

Since $x + y + z = 1$, it is sufficient to determine x and y. The quantity Y gives the lightness of the color. $Y = 100$ means white, and $Y = 0$ means black.

In the following, we study the apparent color of thin films by numerical calculations of the reflectance of single layers and multiple layers on a substrate. The obtained spectra are used to determine the tristimulus values X, Y, and Z and the color coordinates x, y always for the standard illuminant D_{65} and the 2° norm observer.

For the first three figures, we assumed a single layer of SiO_2 (Figure 6.1), Si_3N_4 (Figure 6.2), or TiO_2 (Figure 6.3) on silicon substrate. These are, for example, the most commonly used antireflection coatings for solar cells. We varied the thickness of the layer from $d = 0$ nm to $d = 1000$ nm in steps of 1 nm and calculated the reflectance spectrum from 0.35 to 0.95 μm wavelength and the apparent color.

In Figure 6.1a, we summarized exemplaric spectra of SiO_2 films on silicon substrate for $d = 100$–1000 nm in steps of 100 nm. The dash–dot–dot line corresponds to the pure silicon substrate. Due to the changes in the reflectance when increasing the film thickness, different parts of the illumination get differently reflected. For example, for a film with $d = 200$ nm, the reflectance exhibits a maximum around 600 nm wavelength. For the film with $d = 300$ nm, there is a minimum in the reflectance at 600 nm, but a maximum at approximately 450 nm. From this different behavior in dependence on the film thickness, also different colors result for the films. The color chart in Figure 6.1b shows the evolution of the

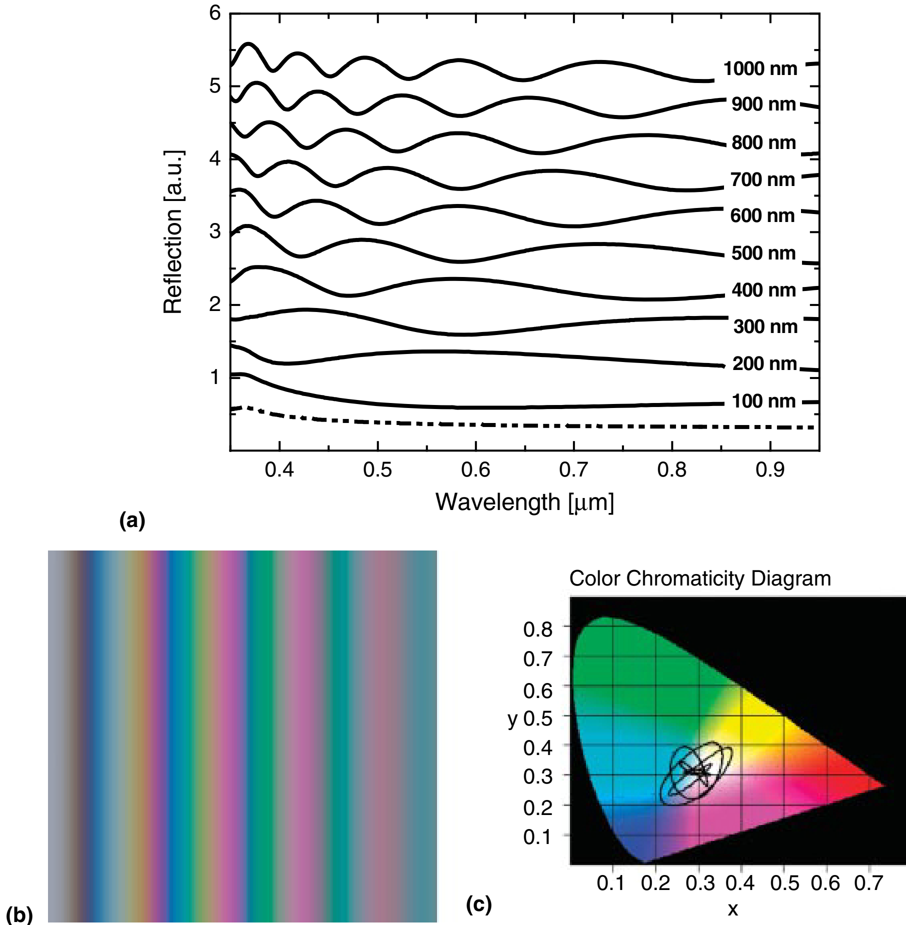


Figure 6.1 (a) Reflectance spectra, (b) color chart, and (c) color chromaticity diagram of thin SiO_2 films on silicon. For the color chart, the film thickness was varied from $d = 0$ nm to $d = 1000$ nm in steps of 1 nm. The black line in the color chromaticity diagram indicates the corresponding color coordinates. Note that the reflectance spectra are shifted by a multiple of 0.5 along the ordinate for better presentation.

apparent color with film thickness from $d = 0$ nm on the left to $d = 1000$ nm on the right. Obviously, the main color changes appear for layers with d less than approximately $d = 600$ nm. For thicker layers, the changes are less because of the increasing number of oscillations in the interference pattern. The black line in the color chromaticity diagram in Figure 6.1c indicates the color coordinates of all calculated films. It confirms that the mostly apparent colors are cyan, magenta, and yellow. Green and blue are seldom, red and orange are never obtained. Moreover, this line approaches more and more the white point with only small extensions to other colors. The main impression of these films is then a gray color. This is obtained for the films with $d > 600$ nm.

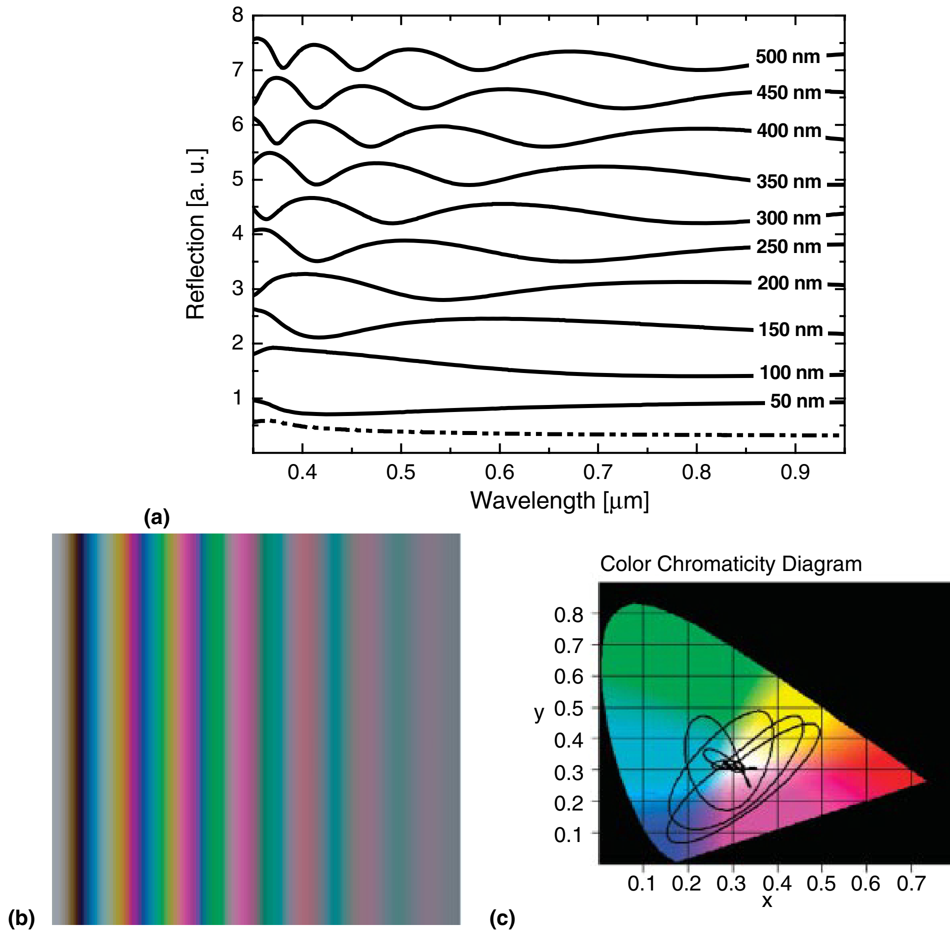


Figure 6.2 (a) Reflectance spectra, (b) color chart, and (c) color chromaticity diagram of thin Si_3N_4 films on silicon. For the color chart, the film thickness was varied from $d = 0$ nm to $d = 1000$ nm in steps of 1 nm. The black line in the color chromaticity diagram indicates the corresponding color coordinates. Note that the reflectance spectra are shifted by a multiple of 0.7 along the ordinate for better presentation.

Comparing the spectra and the color charts for thin silicon nitride films in Figure 6.2 with those of Figure 6.1, we can conclude that silicon nitride films behave similar to silicon dioxide films. However, as the refractive index of Si_3N_4 is higher than that of SiO_2, we obtain comparable effects already for thinner layers. Therefore, we plotted only the spectra for $d = 50$–500 nm. The corresponding curve for the color coordinates indicates stronger changes in the apparent color than in case of SiO_2. This is also obvious from the color chart.

Titanium dioxide (TiO_2) finally has the highest refractive index of these three materials. It approaches almost the refractive index of the silicon substrate, for which reason the dependence on the film thickness is not as strong as one could expect but is

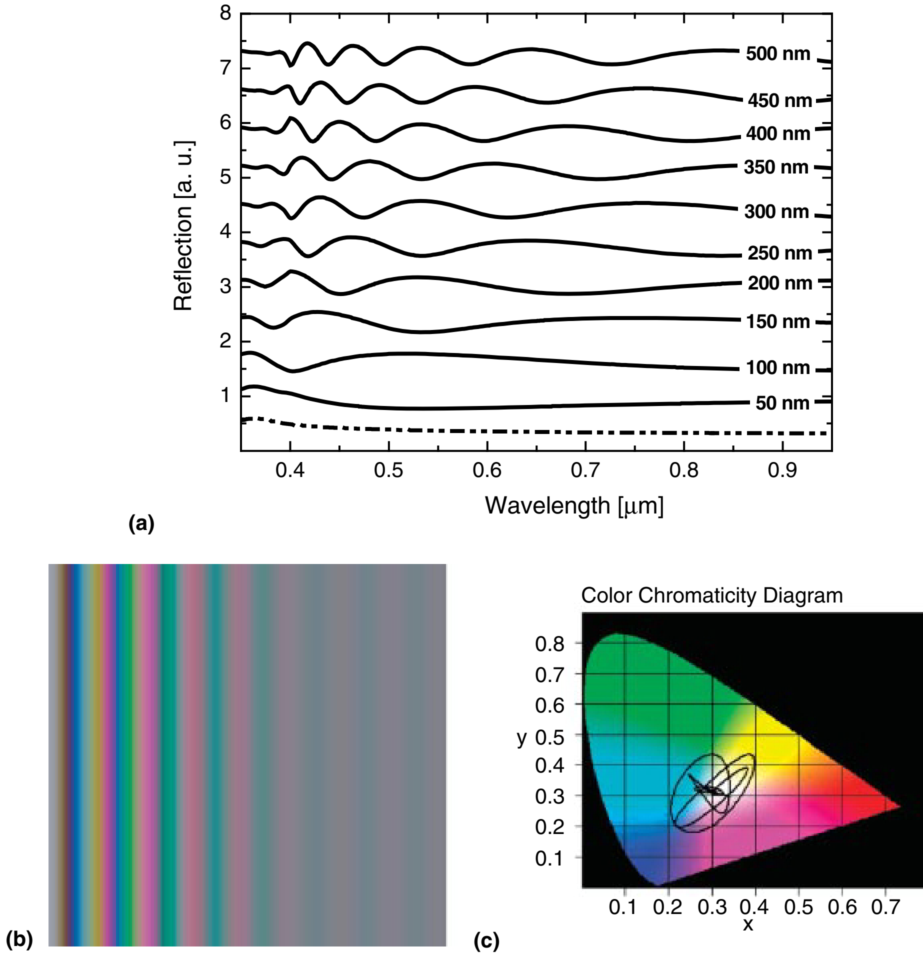
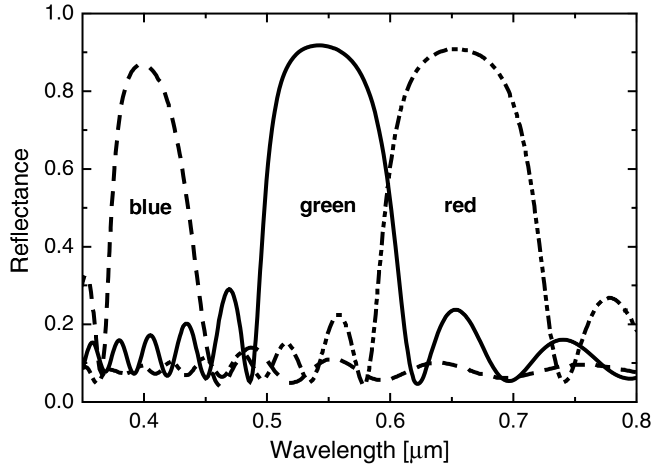


Figure 6.3 (a) Reflectance spectra, (b) color chart, and (c) color chromaticity diagram of thin TiO_2 films on silicon. For the color chart, the film thickness was varied from $d = 0$ to $d = 1000$ nm in steps of 1 nm. The black line in the color chromaticity diagram indicates the corresponding color coordinates. Note that the reflectance spectra are shifted by a multiple of 0.7 along the ordinate for better presentation.

comparable to that of Si_3N_4. This is demonstrated with selected reflectance spectra from $d = 50$ nm to $d = 500$ nm, a color chart and a color chromaticity diagram in Figure 6.3.

Looking at the color charts and color chromaticity diagrams in the above figures, one can recognize that the colors red and orange are missing. Apparently, these colors cannot be generated by a single transparent layer on a substrate. Therefore, we also investigated multilayer systems. One example for which red and orange colors are also obtained is the system Ta_2O_5–SiO_2 with alternating layers of tantalum pentoxide



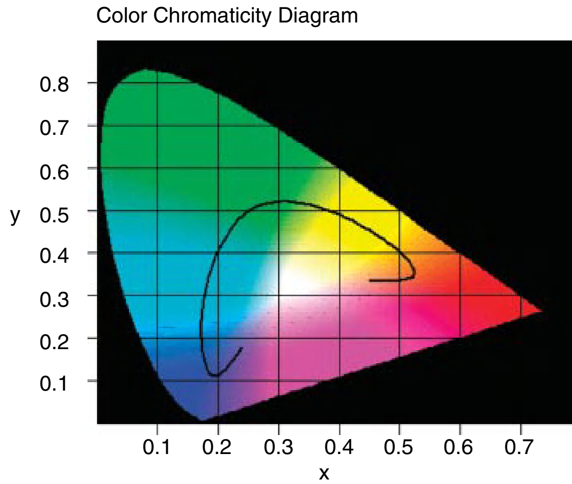


Figure 6.4 (a) Calculated reflectance spectra for a 20-layer stack of alternating Ta_2O_5–SiO_2 layers with the thickness of SiO_2 fixed to $d = 100$ nm. The thickness of Ta_2O_5 is $d = 30$ nm (blue), $d = 70$ nm, and $d = 100$ nm. (b) Color chromaticity diagram for this stack with the layer thickness of Ta_2O_5 varying continuously from $d = 30$ nm to $d = 100$ nm in steps of 1 nm.

and silica. The exemplaric spectra for a 20-layer stack shown in Figure 6.4 are for 30 nm Ta_2O_5 (blue), 70 nm Ta_2O_5 (green), and 100 nm Ta_2O_5 (red) alternating with 100 nm SiO_2. The spectra exhibit a distinct reflectance maximum whose wavelength position shifts to longer wavelengths with increasing thickness of the Ta_2O_5 layers. The chromaticity diagram at the right side of Figure 6.4 shows the colors obtained

when varying continuously the thickness of the Ta_2O_5 layers from 30 to 100 nm. Although red and orange colors are obtained, the main colors are cyan, green, and yellow. For more red colors, other variations of multilayer stacks with other materials but with alternating low refractive index material (LRI) and high refractive index material (HRI) must be used.

7
Applications

Thin films of transparent materials play an important role in many fields of technical applications. The applications can be classified into the following:

- Applications with a single unsupported layer.
  Single unsupported layers appear mainly as
  - thin glass and sapphire sheets (glass or sapphire wafers),
  - transparent polymer films (foils), and
  - semiconductor wafers.

  They are almost everywhere present as food packaging, wrapping, foils, membranes, lamination, and in display technology and solar cells, to give some examples. These layers are usually thicker than 10 μm for which a fast Fourier transform can be used for evaluation of the thickness. The main task is checking homogeneity of the thickness either along a line (e.g., in quality control of transparent foils during their production) or over the full area (e.g., for semiconductor wafers).
- Applications with one layer on a substrate.
  Films with thickness of maximum a few hundred nanometers are used as
  - protective layers (hardcoats),
  - anticorrosion layers,
  - broadband antireflection (AR) coatings,
  - adhesion and antiadhesion coatings,
  - decorative coatings,
  - absorbing layers,
  - photoresists, and
  - transparent conductive layers (TCF and TCO).

  Thicker single layers appear as, for example,
  - protective varnishes (hardcoats),

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

- finishes,
- anodized aluminum, and
- photoresists.

- Applications with two layers on a substrate.
  Frequent applications of two layers on a substrate are the following:
  - A hardcoat on a protective or decorative lacquer (primer) on a substrate.
  - A photoresist on silica on a wafer.
  - Bonded wafer: a semiconductor wafer (mostly silicon, several ten microns) bonded with a glue layer (several microns) on a thick semiconductor wafer.
  - SOI wafer: thin silicon layer on a thin oxide layer on a thick semiconductor wafer (SOI, silicon on insulator).
- Multilayer applications.
  Multilayer systems of several nanometers thick layers are often used for
  - high reflective (HR) and antireflective (AR) coatings with two–eight layers,
  - beam splitter coatings with 4–15 layers,
  - dielectric mirrors with layers 10 to up to more than 40,
  - optical filters with 40 to up to more than 100 layers, and
  - low-E coatings with 2-6 layers including thin Ag layers.

  Typically, in these stacks layers of low refractive index materials (e.g., MgF_2 and SiO_2) alternate with layers of high refractive index materials (e.g., Al_2O_3, Ta_2O_5, and TiO_2). The thickness of each layer can thoroughly be calculated, and the complete layer stack can be optimized.
  Other examples for multilayer systems are
  - thin-film solar cells and
  - OLEDs (organic light emitting diodes).
- Other applications

There are several applications where the assumption of thin layers or films is advantagous for the modeling of the optical properties of the systems. One example is the measurement of critical dimensions of vias and trenches.

In the following sections, we present some examples of modern applications in the field of thin films.

7.1 High-Reflection and Antireflection Coatings

High-reflection (HR) coatings are used to increase the reflectance of the substrate in a certain wavelength range. Vice versa, antireflection coatings are used to reduce the

reflectance of a substrate in a certain wavelength range. Therefore, the main application of HR coatings is the increase of the reflectance of metal mirrors. This will be demonstrated in Section 7.1.1. The main application of AR coatings is the reduction of the reflectance of optical components like lenses to increase the amount of transmitted light. This will be demonstrated with two examples in Section 7.1.2. In other applications also, a reduction of the reflectance is favorable. Particularly in photovoltaics the high reflectance of the semiconducting absorber material strongly reduces the amount of absorbed light and hence the efficiency of the solar cell. As will be seen in Section 7.1.3, single-layer AR coatings can dramatically reduce the reflectance of the shiny solar wafer in the most interesting visible spectral region.

7.1.1
HR Coatings on Metallic Mirrors

HR coatings are mostly used to increase the reflectance of metal mirrors. Particularly, aluminum is often used as mirror in the visible spectral region. With a layer stack of SiO_2–TiO_2 as depicted in Figure 7.1a and proposed by Glöß *et al.* [162], the reflectivity of an aluminum mirror can be increased in the visible spectral region from 92 to 99%. This is shown in Figure 7.1b with the calculated reflectance spectra of a pure aluminum mirror and a HR-coated aluminum mirror with the HR stack from Figure 7.1a. For wavelengths less than 450 nm, the absorption in the TiO_2 layers limits the effect. Also for wavelengths larger than 750 nm, the HR coating yields a less reflectance of the coated mirror than the reflectance of the uncoated mirror. This can clearly be recognized in the interband transition region of aluminum.

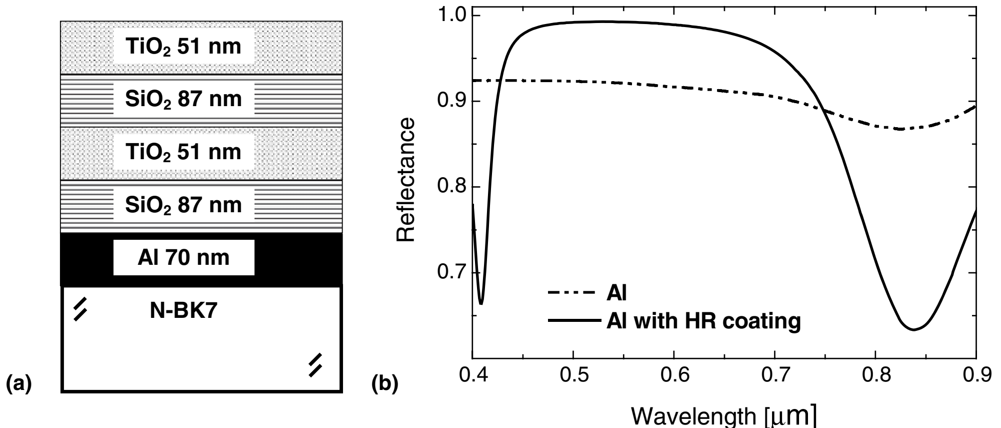


Figure 7.1 (a) Layer stack for HR coating of an aluminum mirror. (b) Reflectance of an aluminum mirror with HR coating.

7.1.2
AR Coatings on Glass

AR coatings act contrary to HR coatings: they reduce the reflectance of the substrate. This is important particularly for high-precision optics where a maximum of light transmission is desired. For the improvement of the quality of life, they can also contribute; for example, when wearing AR-coated glasses bothersome reflections at the glasses can be suppressed.

In the following, we present two examples of AR coatings that are useful to reduce the reflection of a glass. The first example is a combination of SiO_2–TiO_2 layers with different thickness, according to Glöß *et al.* [162], shown in Figure 7.2a.

If this layer stack gets deposited on a N-BK7 glass on one side or on both sides of the glass, the reflectance of a pure N-BK7 glass gets reduced in the visible spectral region from approximately 8.5% (including rear-side reflection) to approximately 4.5% for single-side coating and to approximately 1% for double-side coating. This can be followed from the reflectance curves shown in Figure 7.2b. The double-side coating exhibits a maximum in the remaining reflectance of 3% at approximately 580 nm wavelength, with a resulting yellow color. The reflectance is decreased in the wavelength range between 450 and 800 nm.

The second example is a commonly used three-layer AR coating with materials and thicknesses as given in Figure 7.3a. Again, we calculated the reflectance after single-sided and double-sided coating of a N-BK7 glass.

The reflectance of the N-BK7 glass gets reduced in the wavelength region between 400 and 800 nm from approximately 8.5% (including rear-side reflection) to approximately 5% for single-side coating and to less than 1% for double-side coating. This can be followed from the reflectance curves shown in Figure 7.3b. Both coatings exhibit a maximum at approximately 480 nm wavelength with a resulting cyan color.

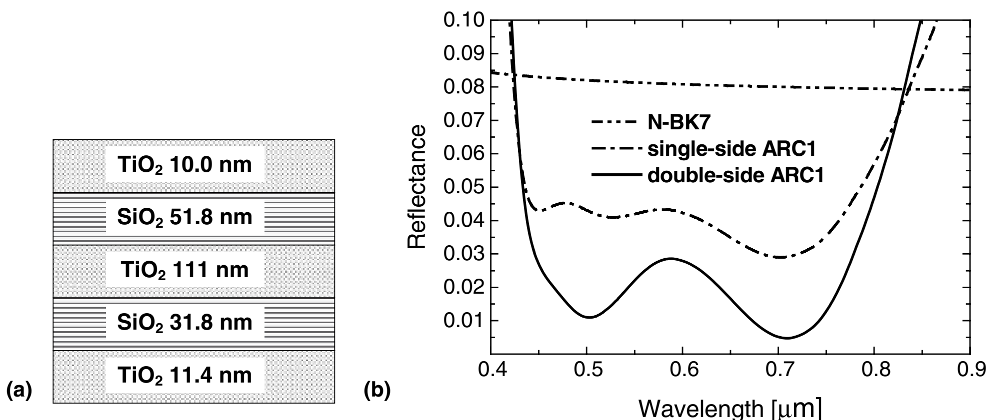


Figure 7.2 (a) Example 1 of an AR coating on N-BK7 glass. (b) Reflectance of an N-BK7 glass slide with this AR coating.

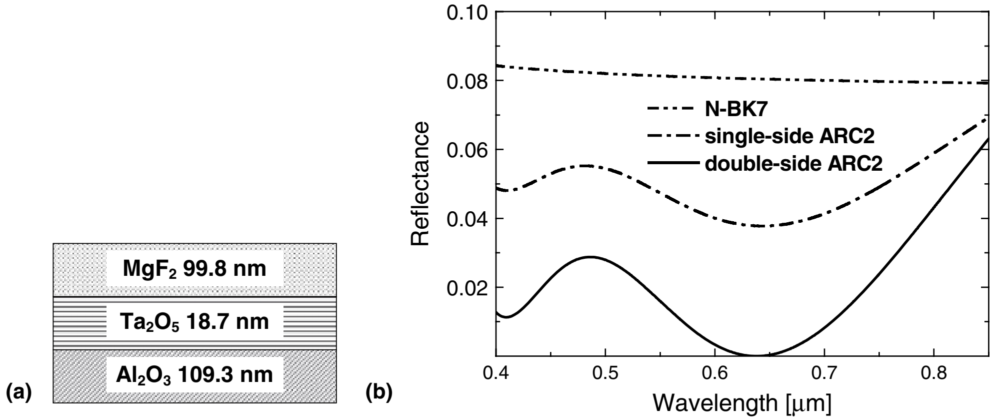


Figure 7.3 (a) Example 2 of an AR coating on N-BK7 glass. (b) Reflectance of an N-BK7 glass slide with this AR coating.

Although the reflectance measurement is here the simplest and fastest technique for an even automated quality control, there are some problems to solve for a 100% automated control. They arise by the transparency of the specimen and the fact that mostly coated lenses with curved surfaces are inspected. The transparency may cause disturbing reflections when measuring only the reflectance of the top surface. For curved surfaces as for lenses the reflectometric measurement is perpendicular only in the apex of the lens causing errors in the determination of the layer thicknesses or the reflectance spectrum. For both problems, a reflectometer has been developed by AudioDev GmbH [163] that automatically adjusts the measuring head perpendicular to the surface and suppresses the backside reflection.

7.1.3 AR Coatings on Solar Wafers

Photovoltaic solar cells are used to convert sunlight into electricity, based on the photoelectric effect. The silicon is doped with phosphorus and boron to make the semiconductor capable of conducting electricity and to establish a p–n junction.

As such a silicon wafer is shiny and reflects between 30 and 45% of the sunlight, it is necessary to coat the wafer with an antireflective coating to reduce the amount of sunlight lost. The most commonly used coatings are titanium dioxide (TiO_2) and silicon dioxide (SiO_2). Commercial solar cells are coated with silicon nitride (Si_3N_4).

In Chapter 6, we already showed representative calculated spectra of thin films of these three materials on a silicon substrate and discussed the apparent color of thin films. For an effective reduction of the reflectance of the solar wafer, it is important that it is significantly reduced over a large wavelength range in the visible spectral region. Therefore, only thin films of Si_3N_4 of less than about 100 nm are of interest as AR coating as can be recognized from Figure 7.4.

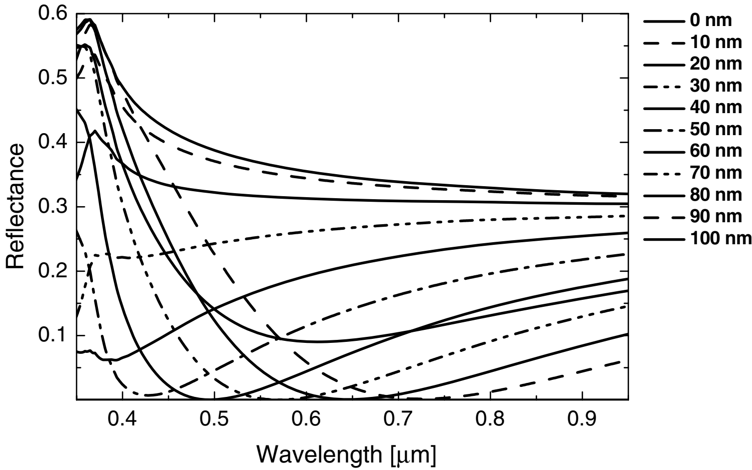


Figure 7.4 Calculated reflectance of AR coatings of Si_3N_4 on a silicon solar wafer.

The largest reduction in the reflectance of the silicon wafer is obtained in the visible spectral region for Si_3N_4 films with thickness d between 60 and 80 nm. However, these spectra are calculated for normal incidence. The effect of the AR coating becomes less with increasing angle of incidence. This is demonstrated by calculated reflectance spectra for a Si_3N_4 film with $d = 70$ nm on a Si wafer in Figure 7.5. As the reflectance is polarization dependent, in principle we must distinguish between the reflectance for p-polarized light and that for s-polarized light. The spectra shown in Figure 7.5 are for unpolarized incident light, that is, the sum of p- and s-polarized light.

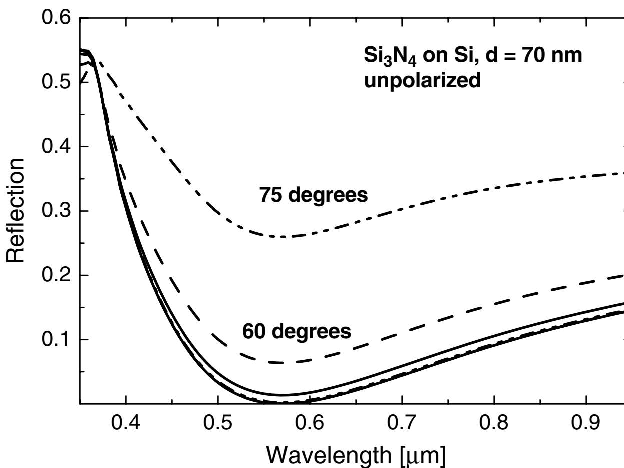


Figure 7.5 Calculated reflectance spectra of a silicon solar wafer with a Si_3N_4 AR coating with $d = 70$ nm in dependence upon the angle of incidence.

From these spectra, it becomes obvious that the Si_3N_4 coating is effective up to angles of incidence of 45°. For larger angles, the reduction in the reflectance due to the AR coating rapidly decreases with increasing angle.

The following Figure 7.6 displays exemplaric measurements of silicon nitride coatings as antireflective coatings on two solar wafers. At left a representative reflectance spectrum is shown and at right a full thickness map of a surface area of $156 \times 156\,mm^2$ is shown. For the full-size thickness mapping, a multisensor metrology tool FRT MicroProf® equipped with a thin-film thickness sensor FTR was used. Note that the z-direction in the 3D views in Figure 7.6 corresponds to the film thickness and not to a height. The mean thickness amounts to $d_{mean} = 79.5$ nm for wafer no. 1 and $d_{mean} = 80.4$ nm for wafer no. 2. The layer thickness is not homogeneous over the wafer area but exhibits a maximum in the middle of the wafers with thicknesses up to 82.4 nm. The smallest thickness is 76.3 nm.

The fabrication of silicon nitride layers by PECVD, sputtering, and thermal evaporation often leads to nonstochiometric materials Si_xN_y instead of Si_3N_4. Then, however, the optical constants are also different from tabulated (n, κ) values of stoichiometric Si_3N_4. The use of optical constants of stochiometric silicon nitride in a regression analysis of the film thickness may then lead to other thickness values. The

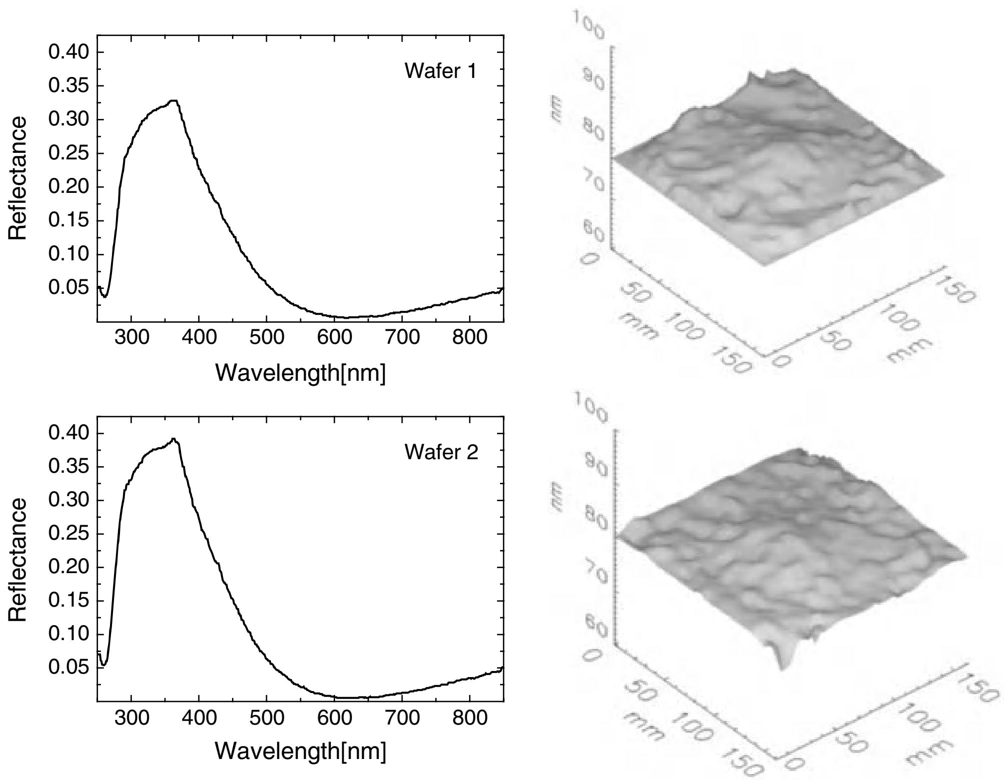


Figure 7.6 Measured spectra and thickness maps of silicon nitride AR coatings on solar wafers.

difference in the refractive index may be significant. Values ranging from $n = 1.8$ to $n = 2.2$ at wavelength 589 nm can be obtained. We studied the thickness and optical constants of a series of silicon nitride layers and determined the thickness and optical constants with a regression analysis with Levenberg–Marquardt algorithm using the self-made software MQNandK [8]. Figure 7.7 displays two measured reflectance spectra with the corresponding thickness fit and the result for the refractive index n and absorption index κ of this Si_xN_y layer material. The thickness of the film was obtained to $d = 83.1$ nm (left side) and $d = 81.1$ nm (right side). For determination of the optical constants, an exponential Cauchy ansatz (see Section 2.7.3) was used. Compared to optical constants from literature [38], the obtained refractive index is 0.15–0.25 higher than the values from literature. Moreover, the absorption index is significantly larger and contributes to the spectrum still at wavelengths $\lambda \approx 600$ nm. With these data, the thickness fit was significantly improved in all measurements in this sample series.

7.2
Thin Single- and Double-Layer Coatings

In the following, we give some application examples for thin single- and double-layer coatings and show representative reflectance spectra as well as full-size

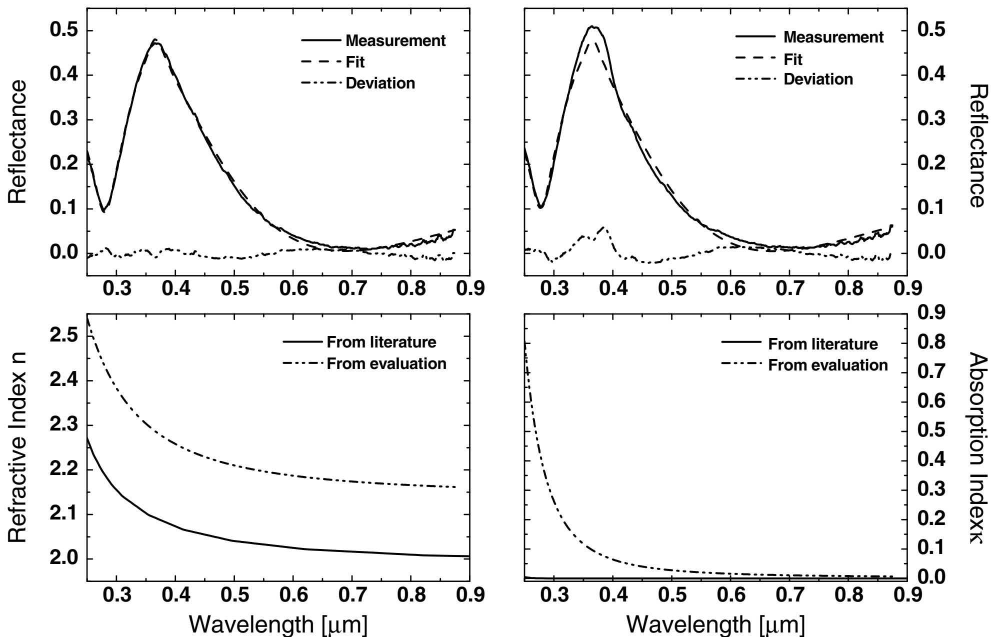


Figure 7.7 Reflectance of a Si_3N_4 antireflective coating on silicon solar wafers and optical constants of Si_xN_y determined from these reflectance measurements.

thickness mappings. The full-size thickness mappings were enabled using a multisensor metrology tool FRT MicroProf® equipped with a thin-film thickness sensor FTR.

7.2.1
SiO_2 on Silicon Wafers

Silicon dioxide, SiO_2, is often used in microelectronics as insulator in metal–oxide semiconductor (MOS) applications or in SOI structures (see Section 7.5), or as passivation layer. It can easily be generated by thermal oxidation of a silicon wafer on its surface. Wafer with such a coating can be bonded together with their oxide layers in contact. To guarantee the correct operation of the electronic devices later on, it is of great interest to check the homogeneity of the coating by measuring the film thickness variation on the wafer. Figure 7.8a shows two exemplaric reflectance spectra of a SiO_2 layer (thermal oxide) on a silicon wafer. They were measured at two positions: in the middle of the wafer and at the border. Figure 7.8b displays the full thickness map over the whole wafer surface. The thickness varies between 135 and 185 nm with a mean thickness $d_{\text{mean}} = 162$ nm for this layer.

7.2.2
Si_3N_4 Hardcoat

Another transparent material for passivation is silicon nitride, Si_3N_4, that also acts as hardcoat. The thickness of such a layer is bigger than that for Si_3N_4 films used as antireflective coating on silicon. Here, we show the result of a thickness mapping of a Si_3N_4 hardcoat film on a structured surface. The representative reflectance spectrum in Figure 7.9a exhibits the characteristic thickness interferences of a thicker film. In the wavelength range between 300 and 350 nm, these oscillations have vanished due

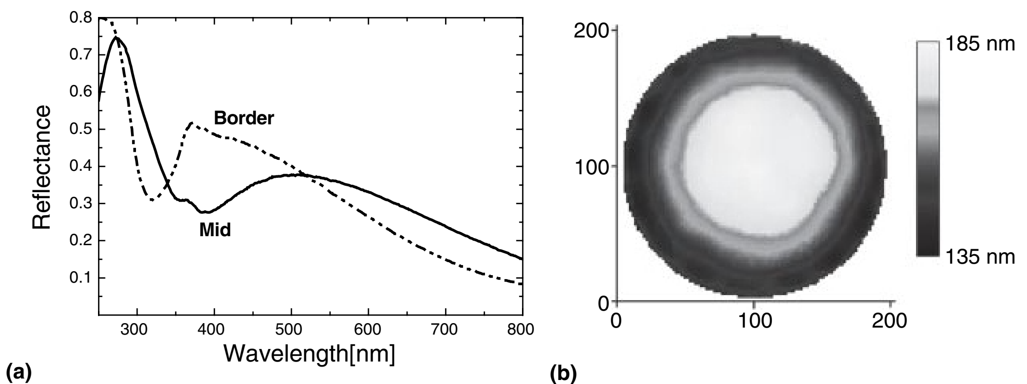


Figure 7.8 (a) Reflectance spectra of a SiO_2 film on a silicon wafer measured at two different positions: at the border and in the middle of the wafer. (b) Full thickness map of the SiO_2 film over the wafer surface.

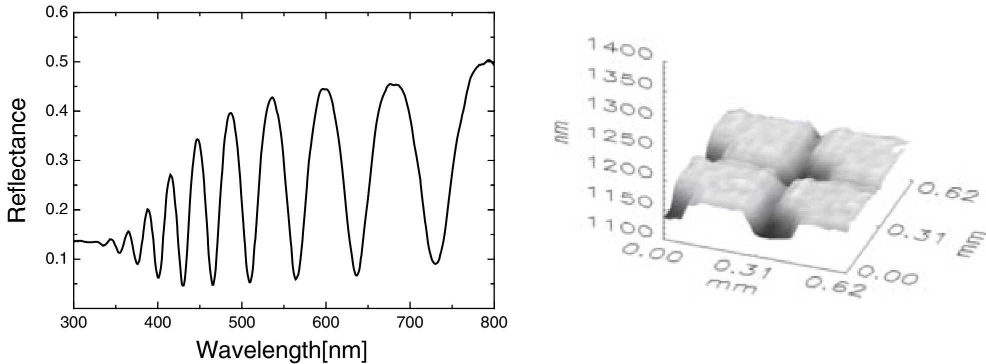


Figure 7.9 (a) Reflectance spectrum of a Si_3N_4 hardcoat film on a structured surface. (b) Full thickness map of the film over an area of $600 \times 600\,\mu m^2$.

to the absorption in Si_3N_4 at these wavelengths. From the mapping of an area of $600 \times 600\,\mu m^2$, it follows that the mean thickness is $d_{mean} = 1.17\,\mu m$ and the thickness varies between 1.16 and 1.18 μm. The film thickness is clearly thicker (a few nanometers) on the structures and thinner in between the structures.

7.2.3 Double-Layer System

The next example is for a double-layer system. Such double layer systems often consist of a lacquer on a substrate as protective or decorative layer and a hardcoat layer on top. The regression analysis of the double-layer system then yields two thicknesses and, hence, also two thickness maps, one map for each layer. Figure 7.10 depicts a typical reflectance spectrum and the thickness maps for the first layer and the hardcoat layer. The two thicknesses differ by approximately a factor of 3, leading to a reflectance spectrum with fast oscillations from the thicker layer overlayed by the oscillations from the thinner layer. In this example, the first layer has a mean thickness $d_{1,mean} = 2.18\,\mu m$ and the hardcoat has a mean thickness of $d_{2,mean} = 7.42\,\mu m$.

7.2.4 Porous Silicon on Silicon

The last example deals with porous silicon on silicon. As the production of pure silicon is expensive, much efforts are made to reduce the amount of silicon in photovoltaic applications. One way is to use porous silicon prepared, for example, by electrochemical processing or chemical etching. Nanoporous silicon is also a promising candidate to replace NaOH texturization of silicon solar cells followed by the deposition of a SiO_2:TiO_2 double layer as a passivation and antireflection coating [164].

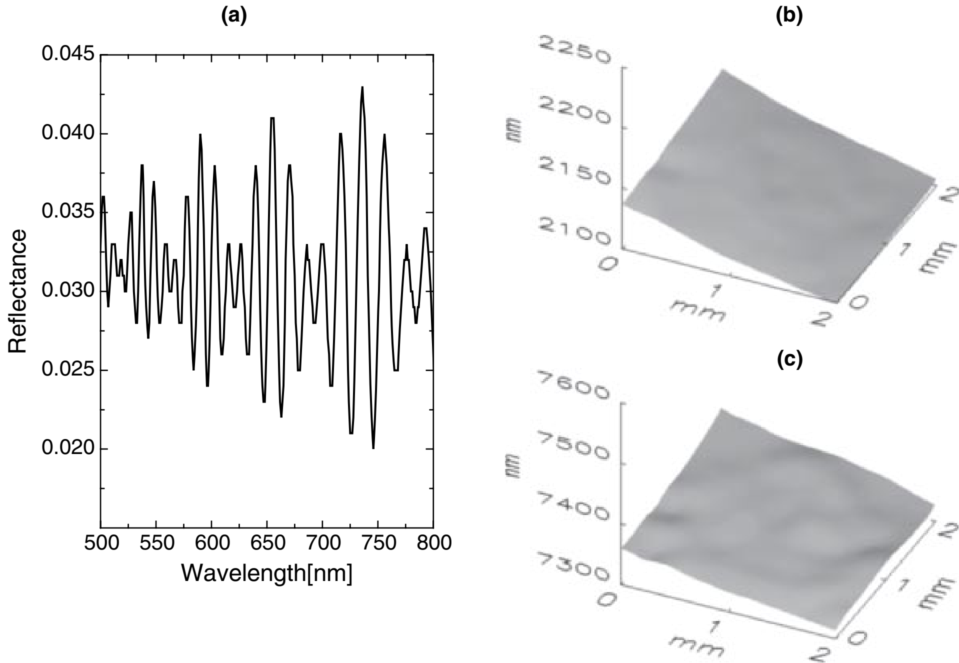


Figure 7.10 (a) Reflectance spectrum of the double layer stack protective film – hardcoat. (b) Full thickness map of the protective film over an area of $2 \times 2\,\text{mm}^2$. (c) Full thickness map of the hardcoat over an area of $2 \times 2\,\text{mm}^2$.

The porosity leads to a significant change in the optical properties compared to bulk silicon. This can be followed from Table 7.1 in which we have summarized values for the complex refractive index of porous silicon [165] and bulk silicon [37].

Table 7.1 Complex refractive index of porous silicon and bulk silicon at different wavelengths.

λ (nm)	$n + i \cdot \kappa$	
	Porous silicon	**Silicon**
300	$2.32324 + i \cdot 1.190440$	$5.00041 + i \cdot 4.169420$
400	$2.46192 + i \cdot 0.112458$	$5.58068 + i \cdot 0.365449$
500	$2.08917 + i \cdot 0.020921$	$4.29848 + i \cdot 0.069782$
600	$1.98428 + i \cdot 0.008176$	$3.94753 + i \cdot 0.025863$
700	$1.93476 + i \cdot 0.003666$	$3.78351 + i \cdot 0.009429$
800	$1.90751 + i \cdot 0.001924$	$3.69269 + i \cdot 0.003868$
900	$1.88638 + i \cdot 0.001132$	$3.62462 + i \cdot 0.001847$
1000	$1.87496 + i \cdot 0.000705$	$3.58608 + i \cdot 0.000963$
1100	$1.86556 + i \cdot 0.000600$	$3.55543 + i \cdot 0.000608$
1200	$1.85783 + i \cdot 0.000452$	$3.52992 + i \cdot 0.000461$

The values are from SOPRALAB [165] for porous silicon and from Humlicek *et al.* [37] for silicon.

The refractive index n is almost halved at all wavelengths compared to bulk silicon and even the absorption is strongly reduced in porous silicon. For wavelengths longer than 900 nm, the absorption index is comparable to bulk silicon. In this near-infrared wavelength range, it is now possible to clearly distinguish a thin film of porous silicon from a substrate of bulk silicon. Therefore, we used a spectrometer working in the near-infrared region from 900 to 1700 nm to measure the reflectance of porous silicon layers on silicon. The resulting spectra of the measurement (solid lines) and of the regression analysis (dashed lines) are summarized in Figure 7.11 for four differently thick layers. The fits are in excellent agreement with the measurement. The obtained film thicknesses are $d = 390$, 770, 896, and 2530 nm.

7.3
Photoresists and Photolithographic Structuring

The thickness and uniformity measurement of thin photoresist films or polymer films is of great interest in photolithographic structuring. The film gets exposed with radiation through a mask to copy certain structures into the film. After

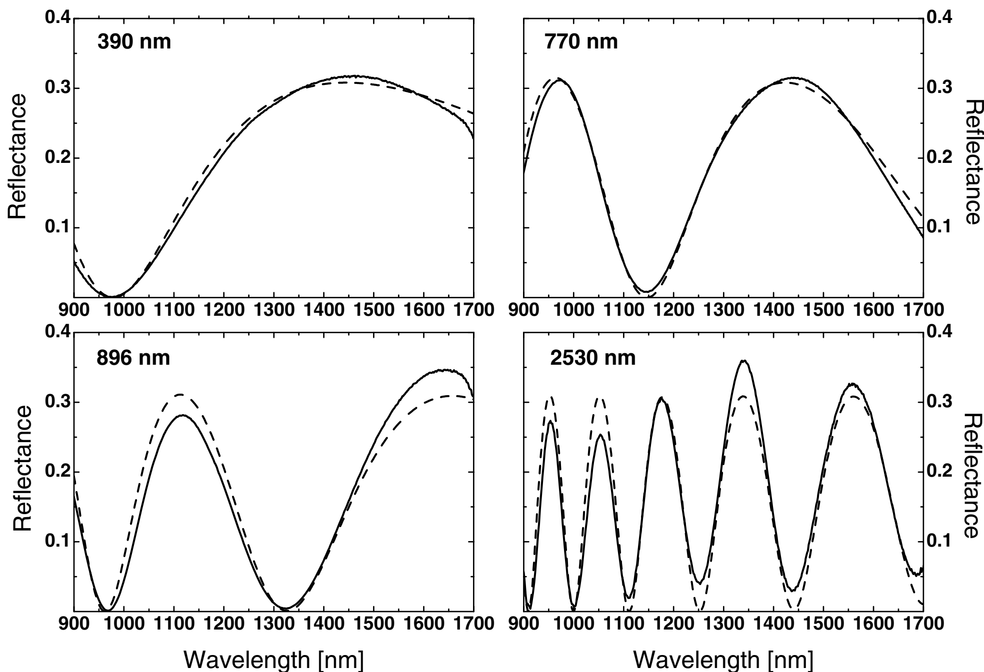


Figure 7.11 Measured (solid lines) and calculated (dashed lines) reflectance spectra of four layers of porous silicon on silicon substrate. The obtained thicknesses are $d = 390$, 770, 896, and 2530 nm.

development or dissolution of the film at the exposed sites (positive resist) or at the nonexposed sites (negative resist), the desired structure can be further etched into the underlying substrate. Thickness variations of the coating lead to nonuniform exposure and hence may result in a nonuniform mask for the further etching process. Therefore, it is of great interest to measure the film thickness variation either before or after exposure.

One example is that of the inorganic–organic hybrid polymers ORMOCER® (ORganical MOdified CERamics, a trademark of the Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V., Munich, Germany). They can be used for many applications, for example, as antisoiling, antistatic, or antireflective coating, in microelectronics or as dental materials.

A film of thickness $d = 13\,\mu m$ was prepared on a substrate and was photolithographically structured. After development of the ORMOCER® film, the thickness variation was measured. The interferometric full-size thickness mapping was enabled using a multisensor metrology tool FRT MicroProf® equipped with an interferometric thickness sensor CWL FT. Note that the z-direction in the 3D view in Figure 7.12a corresponds to the film thickness and not to a height. The film thickness variation along a line (profile) is shown in Figure 7.12c. It is taken out of the 3D measurement as indicated in Figure 7.12b with the black bar. The thickness of the film is rather homogeneous except in the vicinity of the rectangular holes. Obviously, here the film accumulates around the holes, leading to approximately 0.2 μm thicker films there. The small dips in between the holes are clearly to recognize both in the film thickness map and in the profile with a film thickness in the dips of $d = 12.75\,\mu m$.

The second example is the photoresist SU-8 from Microchem Corp. with high UV photosensitivity. It is mainly used in microsystem technology in the LIGA process. It belongs to the group of negative tone resists, meaning that when exposing with UV radiation, SU-8 becomes cross-linked/polymerized and more difficult to dissolve in developer. Therefore, the negative resist remains on the surface of the substrate where it is exposed, and the developer solution removes only the unexposed areas. SU-8 is available with different viscosities for generation of films with different thickness. The thickness obtained at 3000 rpm is given in the name of the photoresist, for example, SU-8 2, SU-8 10, and SU-8 100 for 2, 10, and 100 μm thickness.

The refractive index of SU-8 varies between $n = 1.615$ at $\lambda = 400$ nm and $n = 1.566$ at $\lambda = 800$ nm. Figure 7.13 depicts the measured reflectance spectrum of a SU-8 film of approximately 10 μm thickness and the corresponding power spectrum from applying an FFT analysis. The optical thickness $n{\cdot}d$ is 19.24 μm, from which a thickness $d = 12.16\,\mu m$ follows for a refractive index $n = 1.582$ (at $\lambda = 550$ nm).

The third example is a film of the positive tone photoresist AZ 7200 from Microchem. Corp. on a silica (SiO_2) layer on a silicon wafer. Figure 7.14 depicts a reflectance spectrum measured at the center of the wafer and the corresponding power spectrum of the layer stack photoresist plus silica. With the mean refractive indices $n_{PR} = 1.628$ for the photoresist and $n_{SiO2} = 1.456$ for the silica layer at wavelength $\lambda = 675$ nm, we obtain the thicknesses $d_{PR} = 2.44\,\mu m$ and $d_{SiO_2} = 8.19\,\mu m$.

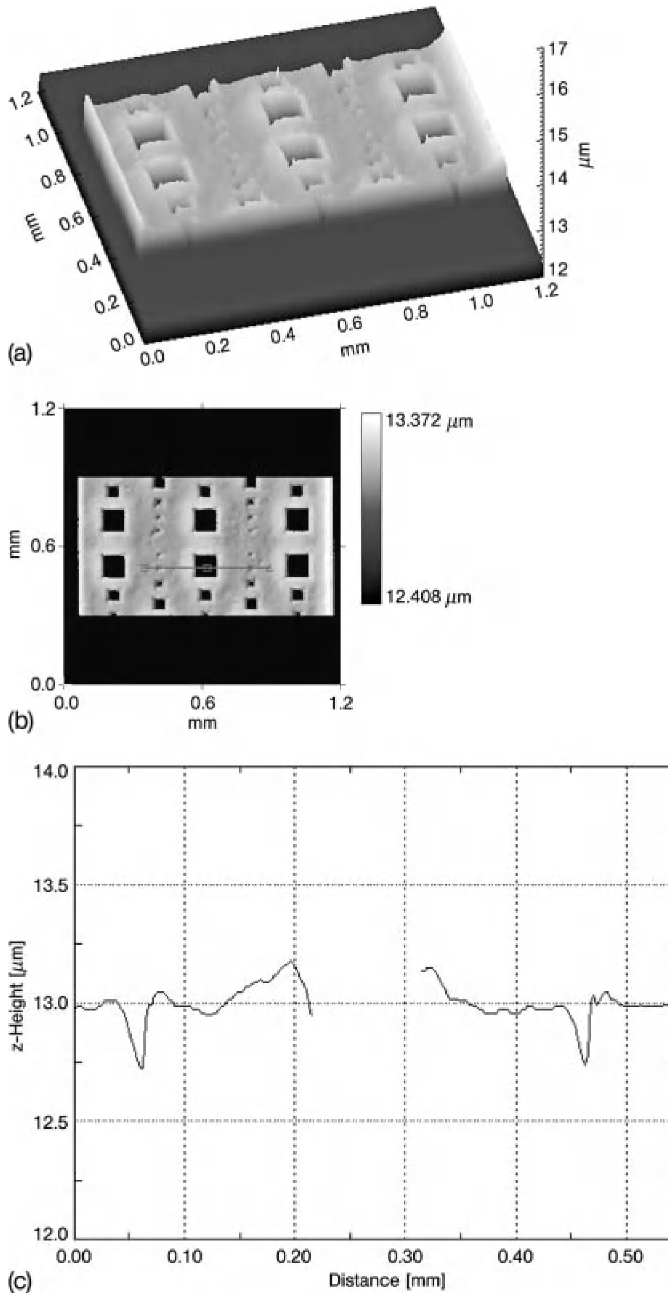


Figure 7.12 Film thickness map of an ORMOCER® film with thickness $d = 13\ \mu m$ on a wafer. (a) 3D view, (b) top view, and (c) thickness profile. Reproduced with permission from Ref. [166]. Copyright Wiley-VCH Verlag GmbH & Co. KGaA.

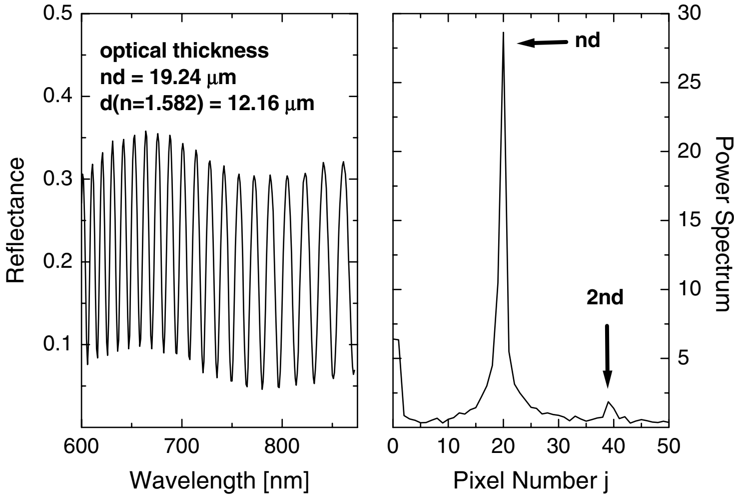


Figure 7.13 Reflectance spectrum and power spectrum of a SU-8 photoresist film on a silicon wafer.

7.4 Thickness of Wafers and Transparent Plastic Films

7.4.1 Thickness of Semiconductor, Glass, and Sapphire Wafers

Semiconducting wafers with thicknesses between 250 and 800 μm are usually opaque in the visible spectral range (see, for example, Table 3.2). Therefore, to

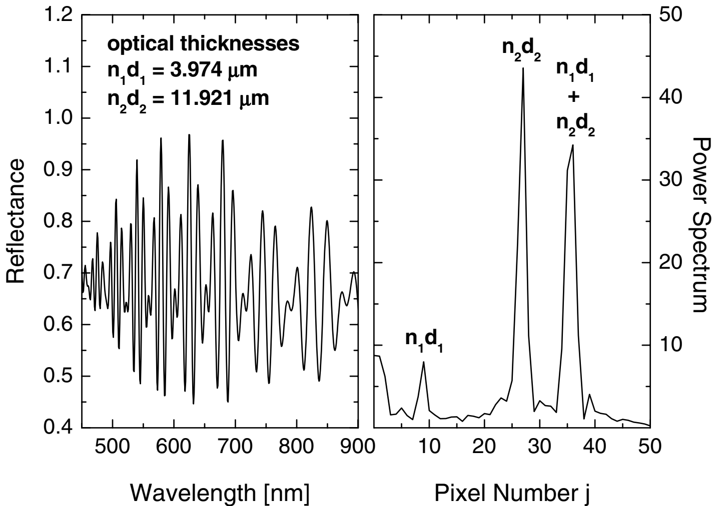


Figure 7.14 Reflectance spectrum and power spectrum of an AZ 7200 photoresist film with $d = 2.44\,\mu m$ on a silica layer with $d = 8.19\,\mu m$ on a silicon wafer.

determine their thickness by optical means, we need to use another spectral range where the wafer becomes transparent. For silicon, this is the near-infrared region between 1200 and 1700 nm. The commonly high refractive index of the semiconducting materials leads to also high values for the optical thicknesses $n{\cdot}d$ for these wafers. In consequence, for interferometric measurement of the thickness, the used spectrometer must have a high wavelength resolution that is achieved by a high number of pixels of the detector and a grating with a narrow spectral range. Typically, the spectrometers have a spectral bandwidth of 60–150 nm and 512 pixel of an InGaAs line detector.

Similar to the opaque semiconductor wafers, thick glass and sapphire wafers of approximately 400–1500 μm can be measured with the same interferometric devices. These materials are transparent both in the visible and in the near-infrared spectral region.

Interferometric full-size thickness mappings of semiconductor, glass, or sapphire can be carried out using the multisensor metrology tool FRT MicroProf® equipped with an interferometric thickness sensor CWL IR. This sensor is available with different overlapping thickness ranges so that a thickness range from 4 to 5000 μm can be covered in such a multisensor tool. These values are valid for refractive index $n = 1$ and must be divided by the refractive index of the corresponding material. Figure 7.15 shows a typical result for an interferometric thickness mapping of a silicon wafer.

Another example is the so-called step wafer. It is produced by step-wise grinding of the full wafer using the TAIKO process developed by DISCO Corp., Japan. The TAIKO process is characterized by a remaining border of a couple of millimeters, while the main part of the wafer gets grinded. By this method the resulting thin wafer keeps mechanically more stable than a wafer the surface of which gets grinded completely. Beginning in the middle of the wafer, first a circular area was grinded to a minimum thickness of approximately $d = 100$ μm. Then, two more segments of a circle were grinded from the inner to the border of the wafer with approximate thicknesses $d = 200$ and $d = 300$ μm. Measurement of reflectance spectra was carried out with a high-resolution spectrometer of AudioDev GmbH using a spectral range from 1490 to 1615 nm. The obtained spectra are summarized in Figure 7.16 together

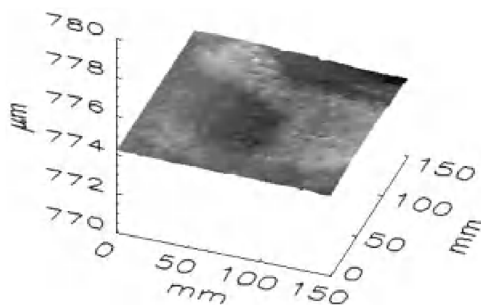


Figure 7.15 Interferometric thickness mapping of a silicon wafer with mean thickness $d = 774.6$ μm.

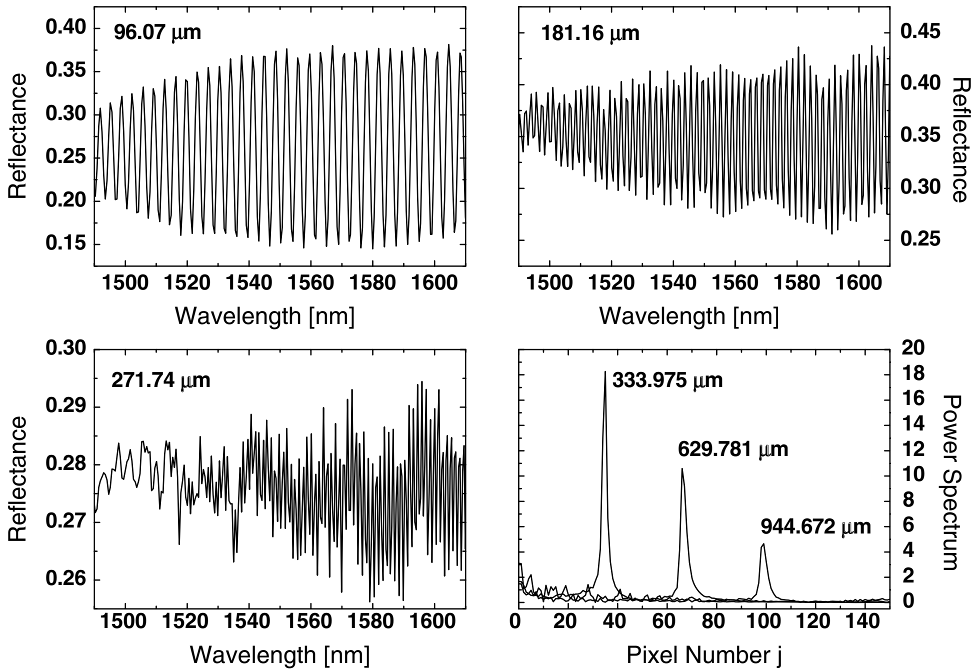


Figure 7.16 Reflectance spectra and power spectra of a step wafer with three grinded regions with $d = 96.07$, 181.16, and 271.74 μm. The wafer has been manufactured with the TAIKO process.

with the power spectra of all three regions. With a mean refractive index for silicon $n_{Si} = 3.4764$ at wavelength $\lambda = 1550$ nm, the thicknesses of the three grinded regions are $d = 96.07$, $d = 181.16$, and $d = 271.74$ μm.

Strongly connected to the measurement of the thickness of bare semiconductor wafers is the thickness measurement of compound wafers where a semiconductor wafer of certain thickness is bonded to a substrate of glass or sapphire. The measuring task is to determine the thickness of the semiconductor, the glue, and the substrate as well as the total thickness. As long as the components are transparent in the considered wavelength range, the interferometric measurement can be applied for determination of all three components. For that purpose, however, the refractive indices of the components must be known in the used spectral range, and the thickness range of the used spectrometer includes all desired thicknesses. In Figure 7.17, we show the result of the interferometric measurement on a compound wafer consisting of silicon with thickness $d_{Si} = 215.5$ μm ($n_{Si} = 3.503$) bonded on a glass wafer with thickness $d_{glass} = 701.1$ μm ($n_{glass} = 1.505$) with an epoxy of thickness $d_{epoxy} = 60.2$ μm ($n_{epoxy} = 1.41$). The analysis of the measured reflectance with FFT exhibits up to six prominent peaks that can be assigned to the optical thicknesses of silicon, glue, and glass, and to the sums of the optical thicknesses.

Beyond the interferometric measurement, another method has successfully been established for optical thickness measurement of thick opaque and transparent substrates: the thickness measurement with two opposite chromatic white light

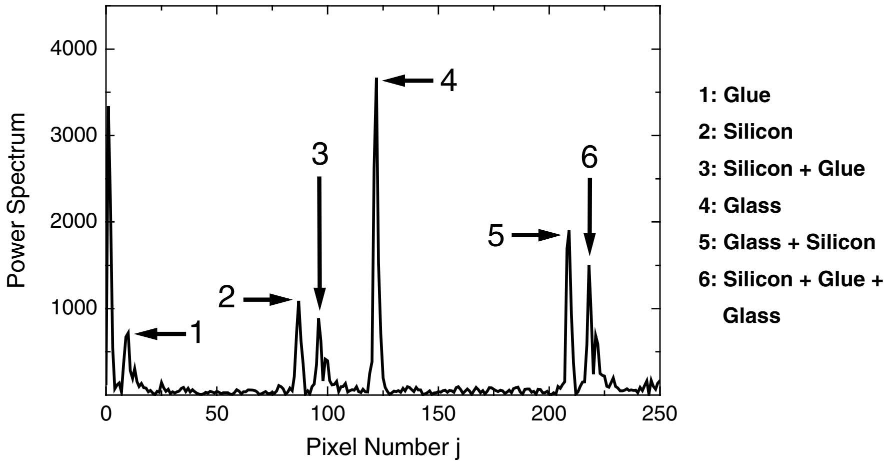


Figure 7.17 Power spectrum of a silicon wafer ($d_{Si} = 215.5\,\mu m$) bonded on a glass substrate ($d_{glass} = 701.1\,\mu m$). The thickness of the epoxy is $d_{epoxy} = 60.2\,\mu m$.

sensors (see Section 4.3.2 for the working principle of the chromatic sensor). A sketch of such a system is drawn in Figure 7.18. Both sensors measure the distance from sensor to the specimen. Calibration of the sensors on a gauge standard of certified thickness allows a rather precise determination of the total thickness and the total thickness variation (TTV) of the wafer.

As the used chromatic white light sensors can also be used both for contactless topography and roughness measurement and for measurement of bow and warp of the wafers, this combination has been established as a workhorse in semiconductor and solar industry. It is realized in, for example, the multisensor metrology tool FRT MicroProf® TTV and can be supplemented by more interferometric thickness

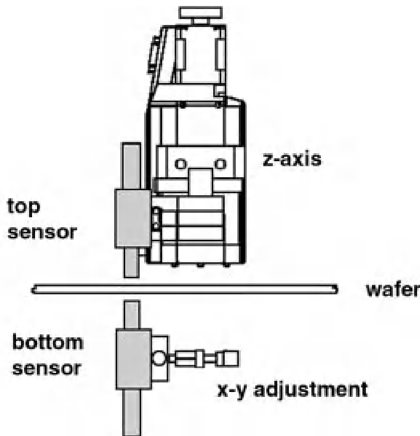


Figure 7.18 Sketch of a TTV measurement with two opposite chromatic white light sensors. Courtesy of FRT GmbH, Germany.

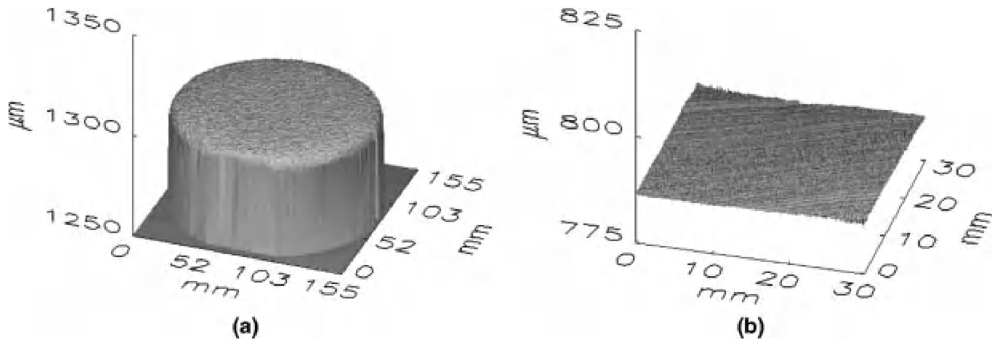


Figure 7.19 Thickness mapping of (a) a sapphire wafer with $d_{mean} = 1299\,\mu m$ and (b) a glass plate of 30×30 mm size with $d_{mean} = 786.2\,\mu m$ using a TTV system with two opposite chromatic sensors.

sensors for thin-film thickness determination. Figure 7.19 displays two typical results for the thickness mapping of transparent wafers with a TTV system.

The thickness of the sapphire wafer is very homogeneous. It varies only between 1.297 and 1.300 mm around the mean thickness of 1.299 mm except for the border where it partly falls down to 1.293 mm. The thickness of the glass plate is also homogeneous with a mean thickness of 786.2 μm and variations between 785 and 787 μm. The surfaces of the glass plate reveal grooves. They result from the polishing of the surfaces.

The advantage of the TTV measurement over the interferometric thickness measurement is that it can also be applied to compound wafers with at least one component being opaque, for example, if germanium is bonded to a glass or sapphire wafer.

7.4.2 Thickness of Transparent Plastic Films

Transparent foils often come across to us as food packaging, wrapping, foils, membranes, and lamination. The thickness typically ranges from approximately 10 to approximately 500 μm. Thickness determination is important mainly to save money because many foils are prepared by extrusion using several different more or less expensive precursors. Hence, mainly in-line thickness measurement is a fast and effective process control technique. In the following, we present two examples of optical thickness measurement on thin transparent foils.

The following example is a thin polypropylene (PP) sheet. Its thickness was measured with a spectrometer working in the near-infrared spectral region because the foil is cloudy in the visible spectral region. Natural PP is translucent because it contains crystals of PP that scatter light. The scattering strongly disturbs the interferences in the visible spectral region. The optical thickness of the foil was determined to $n{\cdot}d = 29.82\,\mu m$, corresponding to a thickness $d = 20.15\,\mu m$ for $n = 1.48$ (PP in the NIR). The measured reflectance spectrum and the corresponding power spectrum from FFT analysis are summarized in Figure 7.20.

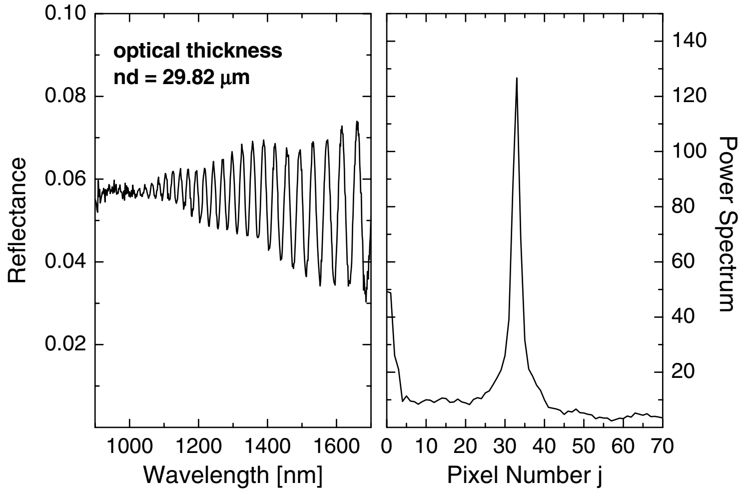


Figure 7.20 Near-infrared reflectance spectrum and power spectrum of a polypropylene foil of thickness $d = 20.15\,\mu m$.

In the next step, this plastic film was coated with a thin varnish. The measured reflectance spectrum in Figure 7.21 now exhibits fast oscillations from the foil overlayed by a slow oscillation that corresponds to the thin coating. The analysis with FFT yielded two peaks that could be assigned to the foil ($n_2 \cdot d_2 = 29.52$ or $d_2 = 19.95$ μm) and the varnish with $n_1 \cdot d_1 = 1.807\,\mu m$. This corresponds to a thickness of $d_1 = 1.203\,\mu m$ with a refractive index of $n_1 = 1.502$ (in the NIR). The sum of both

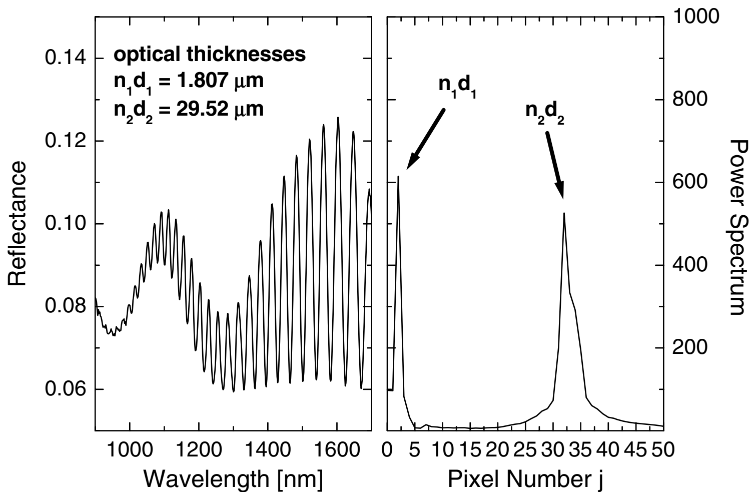


Figure 7.21 Near-infrared reflectance spectrum and power spectrum of a polypropylene foil of thickness $d = 19.95\,\mu m$ coated with a varnish of thickness $d = 1.203\,\mu m$.

optical thicknesses $n_1 \cdot d_1 + n_2 \cdot d_2$ can be recognized in the power spectrum as shoulder at higher pixel numbers. The difference in the thickness of the PP foil measured before and after the coating is mainly due to different measurement positions. Then, the thickness variation in the PP foil leads to slightly different thickness values in the measurements.

7.4.3 Thickness of Doped Silicon

Almost all silicon and solar wafers are doped. The reason is that the intrinsic concentration of free charge carriers in silicon is only on the order of $N_i \approx 10^{10}\,\mathrm{cm}^{-3}$. This is not sufficient to use silicon for electronic devices or solar wafers since with this low carrier concentration the conductivity is too low. It can be increased by a factor up to 10^6 by doping either with elements from the third group in the periodic table (boron, indium, aluminum, and gallium) as electron acceptors (p-doped Si) or with elements from the V group in the periodic table (phosphorus, arsen, and antimony) as electron donors (n-doped Si) to several $10^{20}\,\mathrm{cm}^{-3}$ free charge carriers. The doping elements are embedded in the lattice of the semiconductor crystal.

The common doping processes are

- adding dopants to the melt during the Czochralski process (CZ) or the float zone process (FZ) when forming the ingot,
- thermal diffusion into the wafer from one wafer surface,
- ion implantation of dopants,
- epitaxial growth, and
- alloying.

Epitaxial growth of a film of the dopant material is almost never used. Alloying is also sparsely used and often even undesired. Hence, mainly the first three doping methods above are commonly used.

Doping during the CZ or FZ when forming the ingot is the simplest doping process and results in a homogeneous distribution of the dopants in the complete wafer (basic doping) with charge carrier concentrations of $N \approx 10^{15}$–10^{16}. This basic p- or n-doping increases mainly the conductivity of the semiconductor. For a higher concentration of dopants in a smaller region of the wafer, doping by thermal diffusion or ion implantation is used.

Doping by thermal diffusion is a low-cost method, but has some disadvantages. Typically, the wafer surface gets overlayed with the dopant (infinite source) and the wafer gets heated. The diffusion of dopant material from the infinite source into the wafer leads to a characteristic profile shown in Figure 7.22a as solid line that depends on the diffusion constant of the dopant elements and the time, and can be approximated by an error function erf(x). In the second step – the drive-in – the wafer gets heated again to distribute better the diffused dopants. This diffusion from finite source results in a Gaussian profile as illustrated in Figure 7.22a with the dash–dot-dot line. The final profile is a convolution of error function and Gaussian function with maximum concentration of dopants at the wafer surface. The total

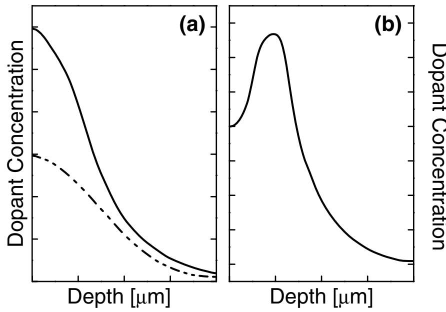


Figure 7.22 Dopant concentration profiles for (a) thermal diffusion and (b) ion implantation.

depth of the profiles from diffusion is typically on the order of 2–4 μm, and can approach about 20 μm. The disadvantages of the diffusion process are that preparation of homogeneously doped layers is not possible and that preparation of low-doped layers is problematic.

For ion implantation, the dopants get ionized and accelerated in an electric field. They hit the wafer with energies of typically 10–400 keV and penetrate the wafer. The penetration depth depends on the mass and energy of the dopant ions, the mass of the crystal atoms, the electronic interaction cross sections, and the crystalline orientation. The resulting concentration profiles are typically Gaussian profiles with the maximum inside a small layer in the wafer, as illustrated in Figure 7.22b. However, the total depth of the profiles is only on the order of 0.5–1.5 μm. As the ions have a high kinetic energy, they damage the crystalline structure of the wafer, which can be partly repaired again by annealing at 700–1000 °C.

From the point of view of optical thickness determination, we must consider that doping also affects the refractive index and the absorption index of the semiconducting material due to the shafting of free charge carriers. This change may not be uniform since the distribution of the dopants may not always be uniform but follows a profile as shown in Figure 7.22. We first discuss the changes in the complex refractive index caused by doping. Then, we consider in detail the thickness determination for wafers with uniform distribution of dopants (only n- or p-doped) and for wafers with nonuniform distribution (basic n- or p-doping during CZ or FZ plus heavy p- or n-doping by diffusion or ion implantation).

Dielectric function or complex refractive index of heavily doped silicon has already been subject of several experimental ellipsometric studies, for example, [167–170].

For calculation of the complex refractive index of doped silicon, we assume the dielectric function of doped silicon ε^* as sum of the dielectric function of undoped silicon ε from Humlicek *et al.* [37] plus a Drude susceptibility for the contribution of the dopants:

$$\varepsilon^*_{\mathrm{Si}}(\omega) = \varepsilon_{\mathrm{Si}}(\omega) - \frac{\omega_{\mathrm{P}}^2}{\omega^2 + i\omega\gamma}. \tag{7.1}$$

The Drude susceptibility changes in dependence on the concentration N of free charge carriers (electrons (n-doped) or holes (p-doped)) as the plasma frequency ω_P depends on the concentration:

$$\omega_P^2 = \frac{Ne_0^2}{m_{eff}\varepsilon_0}. \tag{7.2}$$

In a parabolic band structure, the effective mass m_{eff} of electrons is identical to the electron mass m_e, but in nonparabolic band structures m_{eff} may differ from m_e. For silicon the effective mass is $1.08 \cdot m_e$ for the electron and $0.56 \cdot m_e$ for the hole (derived from the density of states). For $m_{eff} = m_e$, we have approximately $\omega_P^2 = 3.18261 \times 10^9 \times N[\mathrm{cm}^{-3}]$.

The damping constant γ in the Drude susceptibility is also an important factor. This can be recognized best if we resolve (7.1) into real and imaginary part:

$$\varepsilon_{1,Si}^*(\omega) + i \cdot \varepsilon_{2,Si}^*(\omega) = \varepsilon_{1,Si}(\omega) - \frac{\omega_P^2}{\omega^2+\gamma^2} + i \cdot \left(\varepsilon_{2,Si}(\omega) + \frac{\gamma}{\omega}\frac{\omega_P^2}{\omega^2+\gamma^2}\right). \tag{7.3}$$

While the real part of the dielectric function gets decreased by $(\omega_P^2/\omega^2 + \gamma^2)$, the imaginary part increases by this amount, but multiplied by γ/ω. Hence, for a strong damping of the free charge carriers introduced by doping, mainly the absorption in the wafer gets affected.

The damping constant γ can be retrieved from the conductivity σ or the resistivity $\varrho = 1/\sigma$ of the wafer using the relation

$$\sigma = \varepsilon_0 \frac{\omega_P^2}{\gamma}, \tag{7.4}$$

presuming the conductivity or the resistivity has been measured for the whole wafer. Using Maxwell's relation

$$n + i\kappa = \sqrt{\varepsilon_1 + i\varepsilon_2} \tag{7.5}$$

the complex refractive index is obtained from the changed dielectric function.

To study the effect of doping on the refractive index, we carried out calculations on the optical constants of doped silicon assuming three damping constants $\gamma = 10^{14}$ s^{-1}, $10^{15}\,\mathrm{s}^{-1}$, and $10^{16}\,\mathrm{s}^{-1}$, and varied the carrier concentration N from $N = 10^{15}$ cm^{-3} to $N = 10^{20}\,\mathrm{cm}^{-3}$ by multiplying with the factor 10. The dielectric function of doped silicon was calculated for $m_{eff} = m_e$. In Figure 7.23, we summarized the results for $n(\lambda)$ and $\kappa(\lambda)$ for a damping constant $\gamma = 10^{15}\,\mathrm{s}^{-1}$ in comparison to the data from Humlicek *et al.* [37] and the ellipsometric data for heavily arsenic-doped silicon from Aspnes *et al.* [168]. The carrier concentration N was here $N = 3.3 \times 10^{20}\,\mathrm{cm}^{-3}$. We found a good agreement with the data from Aspnes for $N = 3.3 \times 10^{20}\,\mathrm{cm}^{-3}$ and a damping constant $\gamma = 2 \times 10^{15}\,\mathrm{s}^{-1}$ for the imaginary part κ of the complex refractive index. The agreement is not as good for the real part n. The refractive index decreases less than in the data from Aspnes compared to the intrinsic Si values. First influences of the doping on the refractive index can be recognized for $N \geq 10^{17}\,\mathrm{cm}^{-3}$ at wavelengths $\lambda > 1.0\,\mu\mathrm{m}$.

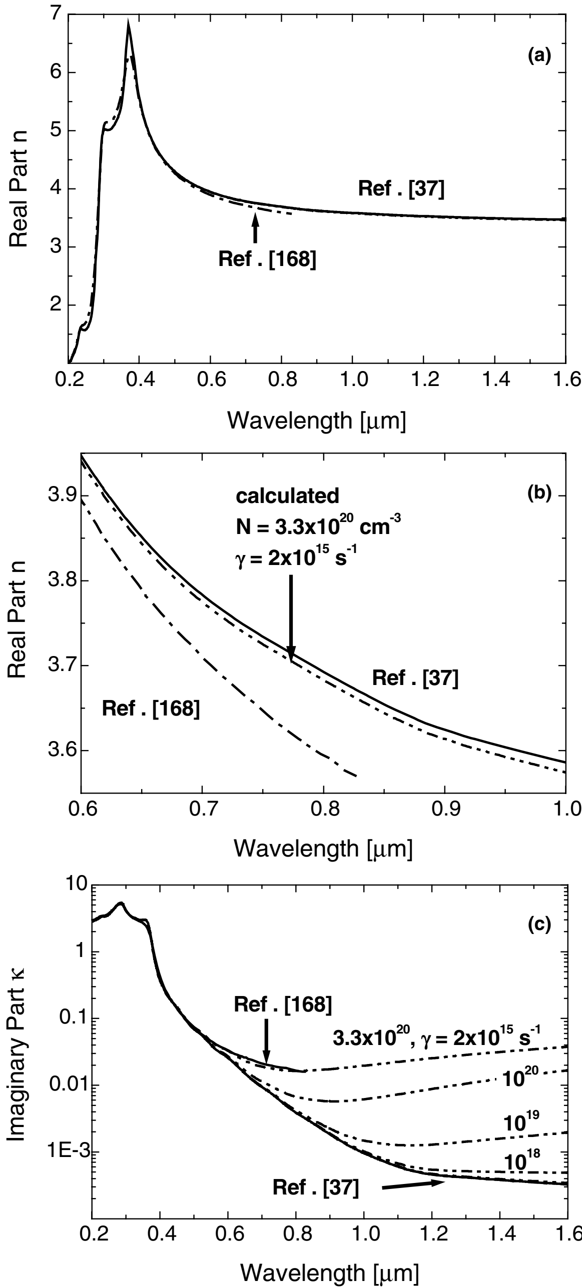


Figure 7.23 Comparison of calculated complex refractive indices of doped Si with measured data from Humlicek *et al.* [37] (intrinsic Si) and Aspnes *et al.* [168] (heavily doped) for $\gamma = 10^{15}\,s^{-1}$ and $N = 10^{15}$, 10^{16}, 10^{17}, 10^{18}, 10^{19}, and $10^{20}\,cm^{-3}$. The best fit to the imaginary part κ of the data of Aspnes is obtained for $N = 3.3 \times 10^{15}\,cm^{-3}$ and $\gamma = 2 \times 10^{15}\,s^{-1}$. (a) Real part n. (b) Real part n in the region 0.6–1.0 μm wavelength. (c) Imaginary part κ.

In the following, we assume a uniform distribution of dopants in the wafer as obtained when doping during the CZ or FZ process. Then, the thickness determination is in principle similar to that of an undoped silicon wafer, except that the dielectric function and the corresponding refractive index are changed by the doping. Therefore, we can expect a dependence of the maximum measurable thickness d_{max} upon the concentration of free charge carriers N.

We calculated from (3.11) the maximum thickness d_{max} for a bare silicon wafer assuming a signal-to-noise ratio SNR = 1000 and considered that also the reflectivity changes with the doping because the refractive index $n + i{\cdot}\kappa$ is affected by the doping. The results for d_{max} in dependence on N are summarized in Table 7.2.

Comparing the maximum thickness d_{max} for $N = 10^{15}$ and $10^{16}\,cm^{-3}$ with the results for d_{max} in Table 3.2, there are only slight differences. The differences increase with N and remarkable differences are obtained for $N = 10^{18}\,cm^{-3}$. Up to this concentration, it is, however, still possible to measure the thickness of uniformly doped silicon wafers of 1000 μm thickness or less, at wavelengths $\lambda > 1100$ nm, presuming the signal-to-noise ratio is SNR = 1000 or even higher. With a concentration of $N = 10^{19}\,cm^{-3}$, the maximum thickness gets almost halved or even cut into thirds compared to $N = 10^{15}\,cm^{-3}$ for wavelengths $\lambda < 1400$ nm. For longer wavelengths, the absorption caused by the free carriers reduces even more drastically the maximum measurable thickness. For $N = 10^{20}\,cm^{-3}$, the maximum thickness is finally less than 100 μm at all wavelengths. For the heavily doped silicon of Aspnes *et al.* [168], the maximum thickness approaches only approximately 20–25 μm.

Having a uniformly doped wafer, for example, a n-type Si-wafer, it will be p-doped in a thin surface region by thermal diffusion or ion implantation to establish a p–n junction. As we have seen above, both the diffusion process and the ion implantation process allow high concentration N of the dopants but with a nonuniform distribution described by either an error function or a Gaussian distribution.

Table 7.2 Maximum thickness d_{max} in micrometers for doped silicon with $N = 10^{15}\,cm^{-3}$ to $N = 10^{20}\,cm^{-3}$ and $\gamma = 10^{15}\,s^{-1}$.

λ (nm)	N (cm^{-3})						
	10^{15}	10^{16}	10^{17}	10^{18}	10^{19}	10^{20}	Aspnes
1000	590	590	587	560	384	92.3	26.2
1100	1028	1027	1017	927	491	85.8	24.3
1200	1475	1473	1450	1251	527	77.4	22.6
1300	1770	1766	1728	1425	516	69.7	21.3
1400	2075	2069	2012	1578	499	63.4	20.2
1500	2408	2399	2316	1719	480	58.3	19.3
1600	2759	2746	2629	1843	461	54.0	18.6
1700	3151	3134	2971	1957	443	50.4	18.0

The last column is for the data that correspond to the data of Aspnes *et al.* [168] with $N = 3.3 \times 10^{20}\,cm^{-3}$ and $\gamma = 2 \times 10^{15}\,s^{-1}$. All values are obtained for a signal-to-noise ratio SNR = 1000.

How can we now determine the thickness of the wafer and what is the influence of the doping profile? A possible way to determine the thickness and even the profile of the dopant concentration may be to divide the heavily doped region into M thin layers of identical thickness d_0. For each of these thin layers, the corresponding concentration N of free charge carriers is obtained from the concentration profile. Then, the refractive index of each thin layer can be modeled using (7.3) and Maxwell's relation (7.5). By this way, the wafer is composed by a layer stack of M layers with thickness d_0 and optical constants $(n + i\kappa)_m$ ($m = 1, 2, \ldots, M$) according to the concentration N_m, on a silicon substrate with thickness $d - M{\cdot}d_0$. The result for the total thickness d then depends on how many layers M are used and how thick they are assumed to simulate the concentration profile. The determination of d is, therefore, very complex and will need a comprehensive regression analysis rather than a single interferometric measurement.

In a theoretical approach, we assumed a doping profile as resulting from ion implantation that extends to 300 or 1500 nm. We divided this region into $M = 15$ layers of thickness $d_0 = 20$ nm or 100 nm and assigned them the corresponding carrier concentration. We assumed a heavy doping up to $N = 3.3 \times 10^{20}\,\mathrm{cm}^{-3}$. The basic doping was assumed $N = 10^{16}\,\mathrm{cm}^{-3}$. With these assumptions, we calculated the wavelength-dependent refractive index for each layer and finally the reflectance spectra of doped wafers of total thickness $d = 50$, 60, 70, 80, 90, and 100 μm. The resulting spectra were analyzed with a fast Fourier transform, and the wafers with the basic doping were analyzed with FFT for comparison. The result is that the profile did not really affect the thickness determination. One reason may be that the total thickness of 1500 nm for the layer stack is too low to introduce significant changes in the reflectance spectra.

7.5
Silicon on Insulator

SOI wafers are silicon wafers having an oxide layer buried below the surface of a crystalline silicon top layer. The wafer is a stack of material with silicon on top (SOI), resting on an oxide film (buried oxide = BO_x) on top of silicon substrate. SOI wafers are typically created by bonding two silicon wafers with SiO_2 layer (thermally oxidized silicon) and then thinning the top silicon layer to the desired thickness. Another manufacturing process is to implant oxygen ions into a thin layer in the silicon wafer followed by annealing to generate a homogeneous SiO_2 layer. This process is called SIMOX. According to the VSI specifications for SOI wafers, the thickness of the thermal SiO_2 layer is between 0.5 and 4 μm, the handle wafer thickness is 250–1000 μm, and the device wafer thickness is 10–525 μm. For a compendium on SOI technology, we refer to Ref. [171].

In Figure 7.24, we give an example of the reflectance spectrum and the corresponding power spectrum from FFT for a SOI wafer with a device wafer thickness of approximately 25 μm on SiO_2 with approximately 4 μm thickness. The silicon substrate thickness amounts to 725 μm. The reflectance spectrum is characteristically

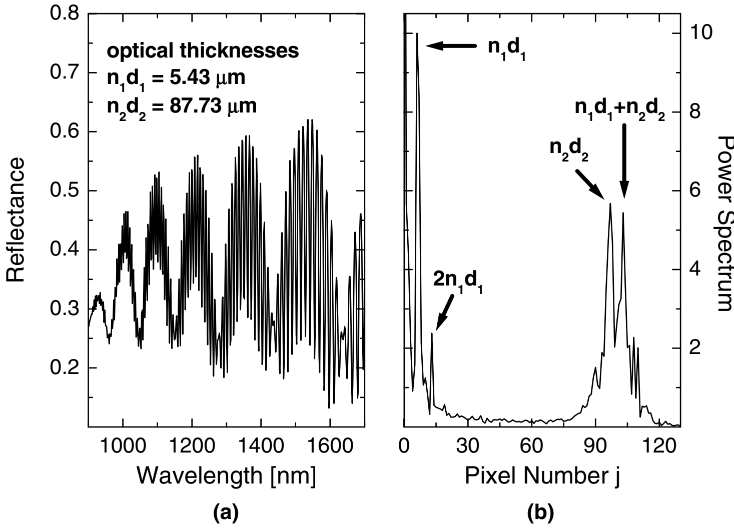


Figure 7.24 Example of a thick silicon top layer with $d_2 = 25.04\,\mu m$ on a buried silica layer with $d_1 = 3.75\,\mu m$. The optical thicknesses of the two layers are $n_1 d_1 = 5.43$ and $n_2 d_2 = 87.73\,\mu m$. (a) shows the reflectance and (b) the corresponding power spectrum from the FFT.

modified by the film thickness interference patterns. The optical thicknesses of the two layers are $n_1 d_1 = 5.43\,\mu m$ for the buried oxide layer and $n_2 d_2 = 87.73\,\mu m$ for the silicon layer. The silicon layer results in fast oscillations that are superimposed by the slower oscillations due to the thin SiO_2 film. The power spectrum exhibits several maxima at different pixel numbers. The most prominent can be assigned both to the single optical thicknesses and to the sum of the two optical thicknesses and higher harmonics. With the mean refractive indices $n_1 = 1.447$ for silica and $n_2 = 3.503$ for silicon at wavelength $\lambda = 1300\,nm$, we obtain thicknesses $d_1 = 3.75\,\mu m$ and $d_2 = 25.04\,\mu m$.

SOI wafers with ultrathin silicon top layer are of growing importance for efficient transistor miniaturization as fully depleted transistors. In this case, the starting substrate for planar ultrathin body transistors and three-dimensional transistors (FinFETand triGate) are SOI wafers with an extremely thin silicon layer of only a few nanometers (typically 10–30 nm) on silica (typically 50–150 nm). The trend is to approach 12 nm SOI/25 nm BO_x. Silicon-on-insulator technology gives many advantages over bulk silicon CMOS processing: higher speed, lower power dissipation, high radiation tolerance, lower parasitic capacitance, low short-channel effects, and high subthreshold voltage swing.

In a series of calculated spectra, we demonstrate in Figure 7.25 how the optical response of such a thin and transparent Si layer on an also transparent and thin SiO_2 layer ($d = 100\,nm$) evolves. Optical constants were taken from Refs [37, 38].

The spectra for $d \leq 25\,nm$ are determined by the buried oxide layer that exhibit a first interference minimum at 950 nm wavelength. However, with increasing Si layer

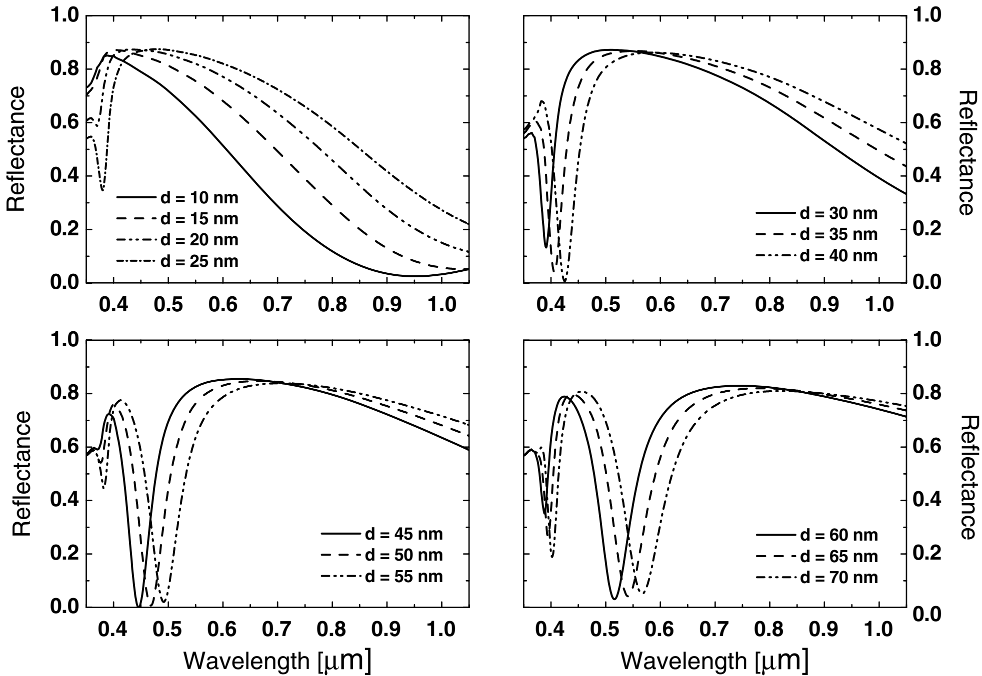


Figure 7.25 Evolution of the reflectance of the system silicon–insulator–silicon for a thin silicon layer with $d = 10$–70 nm on a silica layer with $d = 100$ nm on a thick silicon substrate.

thickness this minimum gets shifted to longer wavelengths and a first interference minimum from the high refractive silicon top layer appears. This minimum shifts to longer wavelengths with increasing thickness and for $d \geq 45$ nm even a second minimum appears that also shifts to longer wavelengths with increasing thickness. The thickness of the very thin Si top layer and the thickness of the buried SiO_2 layer can be evaluated from the spectra by applying a downhill simplex algorithm. Another possibility is to estimate the thickness of the Si top layer from the wavelength position of the first minimum. However, this is possible only if the thickness of the buried oxide layer is well known. For example, in our calculations the thickness of the SiO_2 layer was fixed to $d = 100$ nm and we obtained a dependence of the wavelength position of the first minimum upon the thickness of the layer that can be expressed by the formula

$$\begin{aligned} \mathrm{Pos}(d) = & 392.81219\,\mathrm{nm} - 3.53854\,\mathrm{nm}^{-1} \cdot d \\ & + 0.13888\,\mathrm{nm}^{-2} \cdot d^2 - 7.59907 \cdot 10^{-4}\,\mathrm{nm}^{-3} \cdot d^3 \end{aligned} \tag{7.6}$$

with the thickness d in nanometers.

7.6 Thin-Film Photovoltaics

7.6.1 Inorganic Thin-Film Solar Cells

All inorganic thin-film technologies applied in photovoltaics have the advantage of reducing the amount of material required in creating the active material of a solar cell. However, the majority of film panels has significantly lower conversion efficiencies and need therefore larger areas per watt production. This is the main reason why they have not yet become mainstream solar products due to their lower efficiency and corresponding larger area consumption per watt production. Mainly the following three thin-film technologies are often used for outdoor photovoltaic power production

- Amorphous silicon technology.
- Cadmium telluride (CdTe) technology.
- Copper indium gallium diselenide (CIGS) technology.

Among them, CIGS has the highest efficiency (approx. 20%). The principal setup of the used modules is sketched in Figure 7.26.

The material properties of amorphous silicon (a-Si) are significantly different from those of crystalline silicon (c-Si). Caused by the missing long-range order in a-Si, a large number of dangling bonds are present that must be passivated with hydrogen before the material can be used as solar cell material. The standard a-Si cell is made of multiple hydrogenated a-Si:H layers: the usual p- and n-type layers, plus an intrinsic (i-type) layer. The intrinsic or undoped layer is the active layer of the device. It consists either of a-Si:H or of microcrystalline μc-Si:H. This p–i–n stack of a few 10 micron thickness is sandwiched between a transparent conductive oxide (TCO) and a back

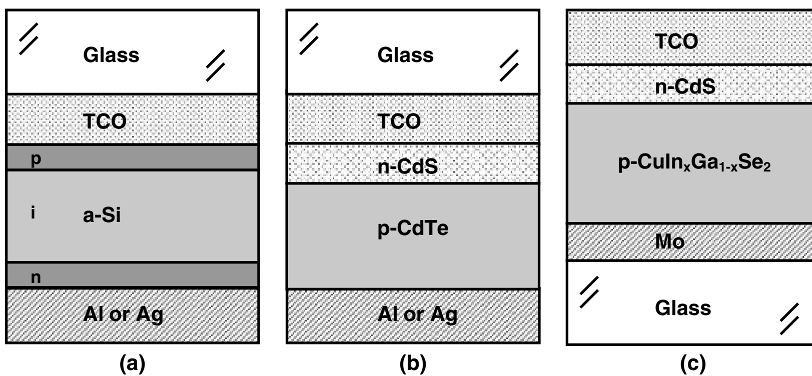


Figure 7.26 Thin-film solar modules: (a) amorphous silicon (a-Si), (b) CdTe, and (c) CIGS.

metallization (Al or Ag). All layers are deposited on a glass or other substrate (compare Figure 7.26)). To improve the performance, often tandem or triple junction p–i–n stacks are used.

Copper indium gallium diselenide is a polycrystalline material consisting of small crystallites of approximately 0.5–1.0 μm in size. The material is a solid solution of copper indium selenide and copper gallium selenide with the chemical formula of $CuIn_xGa_{(1-x)}Se_2$. CIGS has an absorption coefficient that is among the highest for semiconductor materials. Ninety-nine percent of the light incident on CIGS is absorbed in the first micrometer of the device for which the thickness of the CIGS is usually about 2 μm. Optical constants of CIGS can be found, for example, in Ref. [172]. Another favorable characteristic is that copper indium gallium diselenide has one of the highest current densities of any semiconductor material, with the potential to produce high current outputs. These films retain their performance properties better than most semiconductors. And finally, CIGS is amenable to large-area, automated production.

The CIGS is usually formed on a base electrode of molybdenum (Mo) of approximately 0.5–1 μm thickness (back contact) on glass. Thin-film CIGS is a p-type semiconductor. A junction is formed at the surface by deposition of a very thin layer of n-type CdS (approx. 50 nm). This creates an n–p homojunction just inside the CIGS, rather than a simple heterojunction. The device is completed by deposition of a transparent conductive oxide on top of the junction. It usually consists of a thin intrinsic ZnO layer that is capped by a thicker, Al doped ZnO (ZnO:Al) layer.

For a detailed review on thin-film photovoltaic devices, we refer the reader to Refs [173, 174]. From the viewpoint of optical film thickness determination, it is of interest to measure

- TCO and a-Si layer thickness for the a-Si cells,
- TCO and CdS layer thickness for the CdTe cells, and
- TCO and CdS layer thickness for the CIGS cellls.

In the following, we present results for TCO and CdS for CIGS cells.

TCOs typically are made of a layer of indium tin oxide (ITO). Tin-doped indium oxide (ITO, In_2O_3:SnO_2) is a solid solution of indium(III) oxide and tin(IV) oxide. Owing to the increased price of indium, today alternatives to ITO have been developed with aluminum-doped zinc oxide (AZO, ZnO:Al), fluorine-doped tin oxide (FTO, SnO_2:F), and antimony-doped tin oxide (ATO, SnO_2:Sb). The highest optical transparency and the highest electrical conductivity (specific resistance 110 μΩ cm), however, is still obtained for ITO. For solar cells, ZnO:Al layers are widely established.

In Figure 7.27, we present representative measured spectra of transparent conductive oxides ITO (In_2O_3:SnO_2), AZO (ZnO:Al), and FTO (SnO_2:F), and a spectrum of a CdS buffer layer on CIGS like they are used in thin-film solar cells. The indicated thicknesses are obtained from regression analysis. The AZO films exhibit a reflection maximum around 300 nm that is not caused by thickness interferences but results from the absorption in ZnO:Al in this wavelength range. The FTO layer exhibits thickness oscillations corresponding to its thickness of 500 nm and a mean

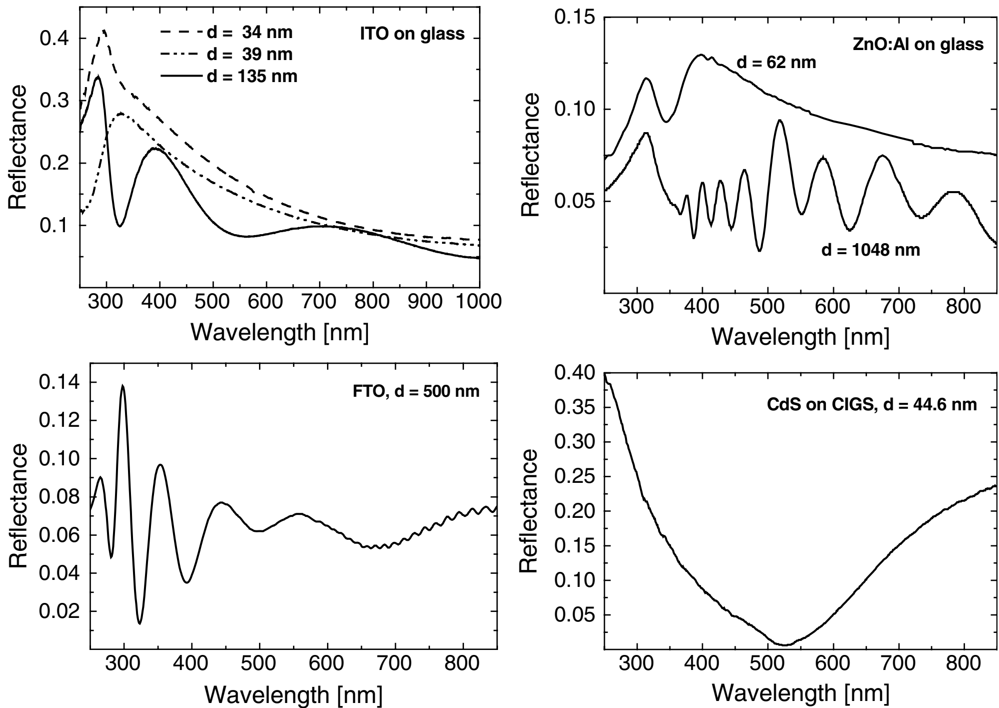


Figure 7.27 Measured spectra of ITO (In_2O_3:SnO_2), AZO (ZnO:Al), and FTO (SnO_2:F) transparent conductive oxides and cadmium sulfide on CIGS.

refractive index of $n = 1.82$. The thin CdS layer with $d = 44.6$ nm on the CIGS absorber acts similar to an antireflective layer. The reflectance exhibits a minimum around 525 nm wavelength with a reflectance of less than 1%. Considering only the visible spectral range (380–780 nm), the reflectance is higher at longer wavelengths than at shorter wavelengths. This behavior results in a yellow or an orange color of CdS films of comparable thickness.

Optical characterization of ITO films is challenging as the optical properties are highly dependent on the film deposition (DC/RF magnetron sputtering, reactive MF magnetron sputtering, CVD) and the annealing process. Mostly, the fabrication process leads to polycrystalline or amorphous microstructures. Hence, besides the thickness determination, the determination of the optical constants of the ITO film is of interest. In Figure 7.28, we present optical constants for ITO derived from various measurements of ITO films on glass in comparison to ellipsometric data from SOPRALAB [165] and Gerfin and Grätzel [39]. Gerfin and Grätzel used a harmonic oscillator model for the interband transition in the UV in ITO plus a Drude susceptibility for the free charge carriers in ITO. This approach was also used in our calculations with MQNandK [8].

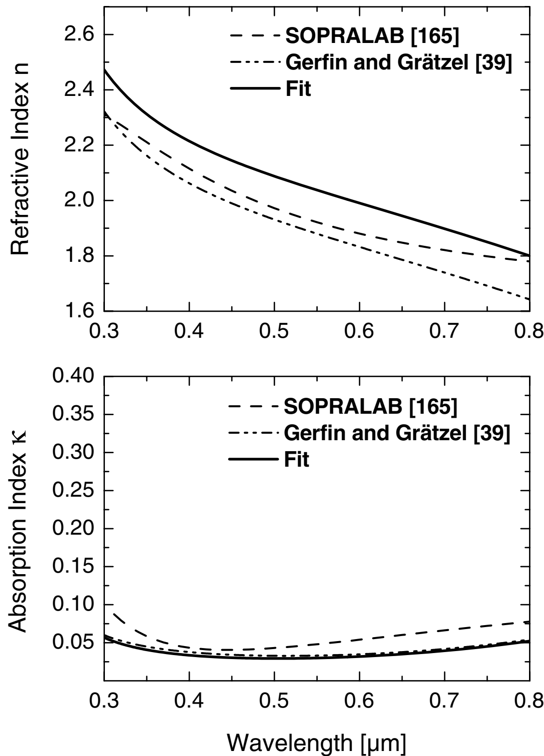


Figure 7.28 Optical constants of ITO films derived from thin-film measurements on glass in comparison to optical constants from SOPRALAB [165] and Gerfin and Graetzel [39].

7.6.2 Organic Thin-Film Solar Cells

Organic solar cells are a relatively novel technology. As these cells can be processed from solution, a simple roll-to-roll printing process leads to inexpensive, large-scale production. However, energy conversion efficiencies achieved to date are low compared to inorganic materials.

Organic solar cells and polymer solar cells are built from thin films (typically 100 nm) of organic semiconductors including polymers. The principal setup of an organic solar cell is sketched in Figure 7.29. ITO (or another TCO) is put on glass or PET substrate (thickness of ITO about 20–100 nm). Then, a layer of the organic conductive and transparent film PEDOT:PSS (poly(3,4-ethylenedioxy-thiophene): poly(styrene sulfonate)) follows that is used both for smoothing the ITO surface and for improving the hole conduction (typically 40–100 nm thickness of PEDOT). The active layer is formed by P3HT:PCBM (poly(3-hexylthiophene):[6,6]-phenyl-C61-butyric acid methyl ester) (typically 100–350 nm). The back contact is formed by aluminum (around 80–100 nm thickness). The use of PET as substrate (around

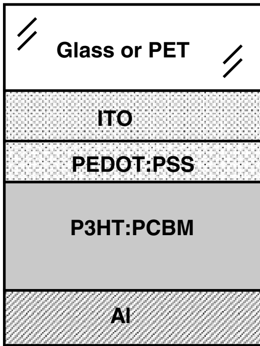


Figure 7.29 Organic thin-film solar module.

150–250 μm) allows to build a flexible solar cell. When a photon is converted into an electron hole pair in the P3HT:PCBM layer, the charges remain bound in the form of an exciton and are separated when the exciton diffuses to the interface to the PEDOT:PSS layer.

In Figure 7.30, we show representative reflectance measurements of a PEDOT film on ITO on glass and a P3HT film on ITO on glass. The thicknesses of the films were obtained by regression analysis.

The PEDOT film is too thin to exhibit oscillations in the reflectance. Only the first minimum appears around 310 nm wavelength. The thicker P3HT film already exhibits thickness oscillations with a maximum around 600 nm that is caused not only by the thickness interferences but also by the absorption in the film. The P3HT film appears greenish in the reflectance as it exhibits a maximum between 500 and 650 nm in the visible spectral range. In this range also the absorption is high in the P3HT:PCBM layer.

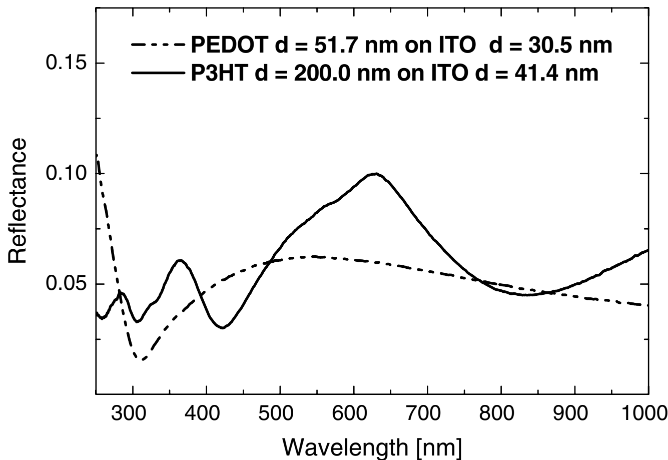


Figure 7.30 Reflectance spectra of a thin PEDOT film and a thin P3HT film on ITO on glass.

7.7 Measurement of Critical Dimensions

Advanced semiconductor devices like deep-trench capacitor DRAM devices are scaled to more and more smaller dimensions. For example, the TSV roadmap specifies 0.8 μm diameter TSVs with an aspect ratio of 10 : 1–20 : 1, that is, deep narrow structures. There are four main structures requiring advanced in-line automated metrology techniques.

(A) **Through silicon vias**: The critical dimensions of a through silicon vias (TSV) are the depth, the top diameter, the bottom diameter, and, if top and bottom diameter differ, the taper angle. Usually, the silicon wafer has a coating of SiO_2 on the order of 1–2 μm on top surface. For process control, TSV arrays with fixed pitch are used.
(B) **Deep straight trenches**: The critical dimensions of deep straight trenches are very similar to the TSV critical dimensions.
(C) **Bottle-shaped trenches**: At the 90 nm technology node, the trench capacitor design of DRAM devices has changed from a vertical pit to the "bottle" trench design, with a "neck" that is narrower than the "body" below.
(D) **Recessed trenches**: Following the fabrication of the deep-trench storage capacitor, the transistor and isolation structures are formed near the top of the deep trench. This process involves several cycles of trench filling, with polycrystalline silicon (polysilicon) or resist, and etchback to form recess structures. The recess depth, that is, the remaining depth of the trench after refill, is of interest for the fabrication processes illustrated in Figure 7.31.

Traditionally used methods in process control run into more and more problems when measuring critical dimensions of trenches, vias, TSVs, and recess parameters. Hence, a method has been established that uses infrared reflectometry or ellipsometry for optical measurement and modeling the measured reflectance using EMA (effective medium approximation) models (see Section 2.7.3). This method is called *model-based infrared reflectometry* (MBIR).

MBIR provides noncontact, rapid measurements from which details of the complex trench profiles can be extracted. It uses an infrared beam with wavelengths $\lambda > 1.4$ μm to probe the trench structure. Silicon microstructures are transparent at these wavelengths and one obtains an interference pattern in the reflectance spectrum that encodes details of the trench shape and depth. Light scattering is minimized since the measurement wavelengths are significantly longer than the DRAM array pitch. Then, the DRAM trench structures can be modeled as multi-layered film stacks with optical properties (n and k) of the various layers computed according to EMAs.

In Figure 7.32, we show schematically how the above different physical structures (A–D) are optically modeled. The general procedure is as follows:

- Replacing the hardcoat layer (Si_3N_4 or SiO_2) with TSVs or trenches by a layer of Si_3N_4 or SiO_2 with a certain amount of voids. The dielectric function of this layer gets calculated with a two-dimensional EMA model. It is necessary to have a

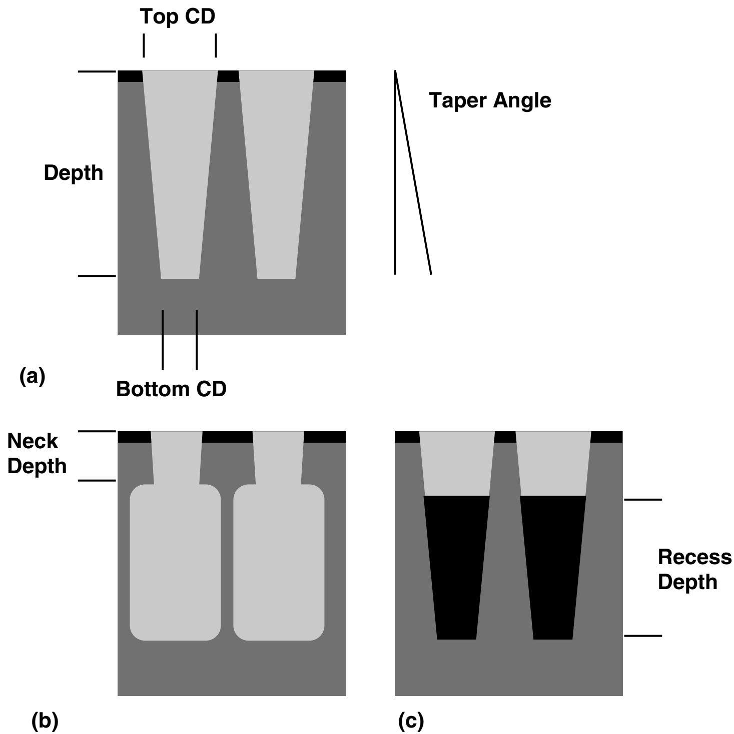


Figure 7.31 Illustration of the main structures with critical dimensions: (a) through silicon vias and deep straight trenches, (b) bottle-shaped trenches, and (c) recessed trenches.

periodic arrangement of trenches or vias because only then the filling factor *f* in the EMA model can be connected with the volume of the single via or trench and the periodicity of the vias or trenches in the periodic array.

- Replacing the silicon region with vias or trenches by a layer of Si with voids. The effective dielectric function is again calculated with 2D-EMA models. If necessary, this region can be divided into multiple layers to consider the shape of the via or trench in more detail. For the bottle-shaped trenches and the recessed trenches, the silicon region gets divided into two regions: a neck region with empty narrow trenches and a region either with wider trenches for the bottle-shaped trenches or with narrow trenches filled with polycrystalline silicon or resist.
- Replacing the bottom silicon region with vias or trenches by a multilayer stack of thin layers (graded layer) to consider different depths of the trenches. The optical constants of each layer are calculated from an EMA model.
- Calculation of the reflectance of this multilayer stack with a silicon substrate.
- Fit with Levenberg–Marquardt algorithm to determine the critical dimensions and the depth.

For further reading on MBIR and its application, we refer to Refs [175–181]. An overview of the state-of-the-art metrology of periodic trench structures based on

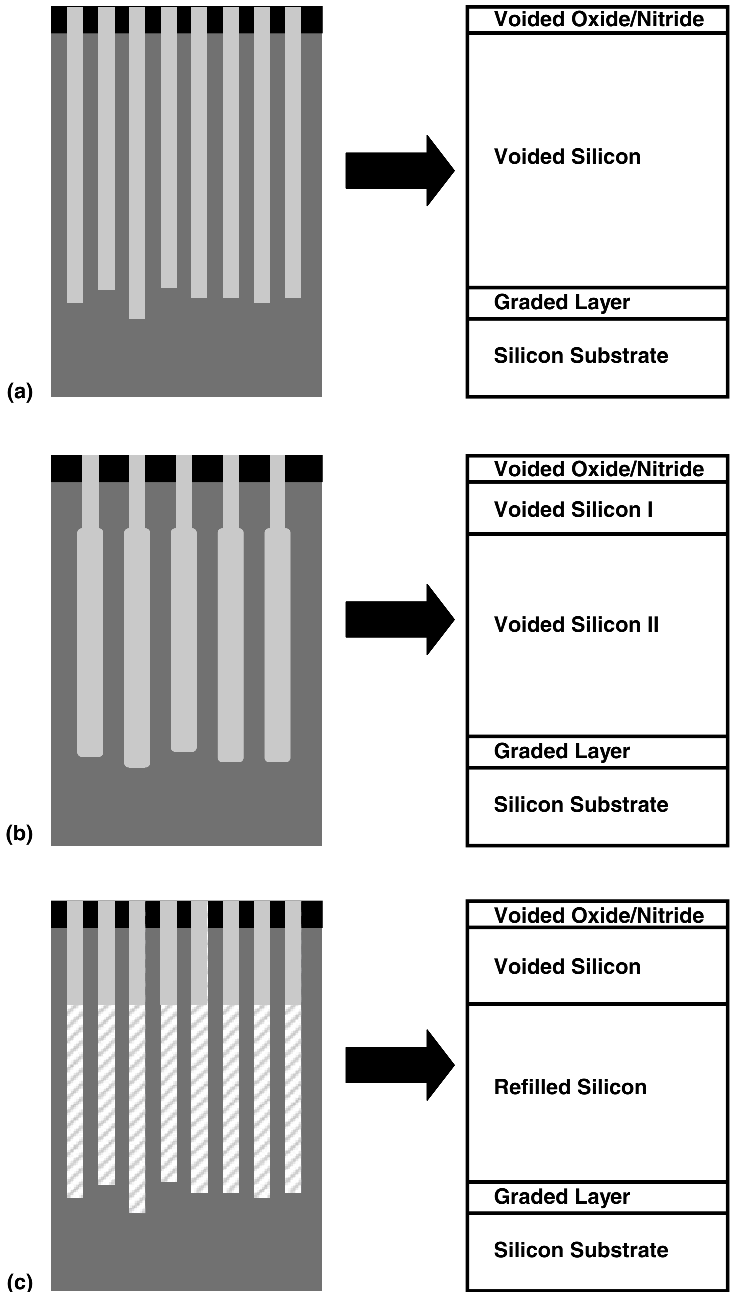


Figure 7.32 Optical multilayer systems for the modeling of the various trench structures. (a) Model for TSV and deep straight trenches, (b) model for bottle-shaped trenches, and (c) model for recessed trenches.

infrared reflectance spectroscopy is provided in the paper by Maznev *et al.* [182]. Exemplary experimental spectra are taken from Gostein *et al.* [179] and are shown in Figure 7.33. The newest MBIR technique – the small spot extended MBIR – is realized, for example, in the IR 2500 S from Semilab AMS, using a probe size of 50 μm and a wavelength range from 0.9 to 20 μm.

Recent results obtained applying the MBIR technique as an in-line monitoring technique for high aspect ratio structures (deep-trench isolation structures, TSVs) are presented in Ref. [183]. The technique is demonstrated to be a robust method for the in-line geometry control of etched structures.

The success of this technique relies heavily on accurate modeling of trench structures and fast extraction of trench parameters. In 2009, Zhang *et al.* [184] proposed a modeling method named corrected effective medium approximation (CEMA) for an accurate and fast reflectivity calculation of deep-trench structures. They also developed a method combining an artificial neural network with a Levenberg–Marquardt algorithm for robust and fast extraction of geometric parameters from the measured reflectance spectrum.

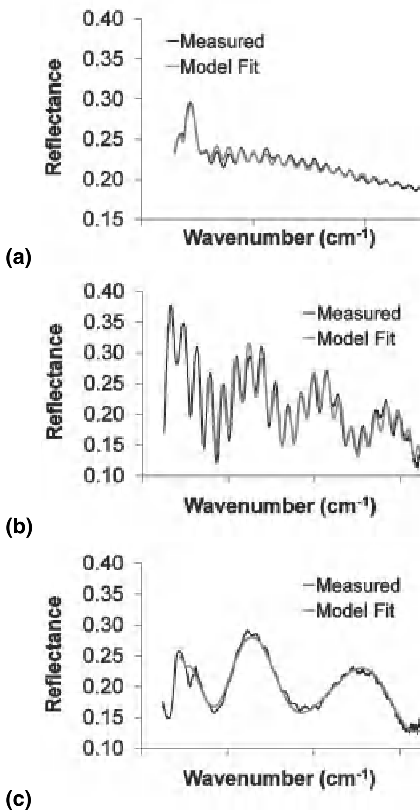


Figure 7.33 Experimental MBIR spectra of (a) deep trenches, (b) bottle-shaped trenches, and (c) recessed trenches.

Appendix A: Numerics with Complex Numbers

Complex numbers are useful abstract quantities that can be used in calculations and result in physically meaningful solutions. Complex numbers have been introduced to allow for solutions of certain equations that have no real solution. For example, the quadratic equation $x^2 + 1 = 0$ has no solution in the field of real numbers. Complex numbers are a solution to this problem. The complex numbers are the field $\mathbb{C}$ of numbers of the form $z = x + i \cdot y$, where $x, y \in \mathbb{R}$, the field of real numbers, and i is the *imaginary unit* $i = \sqrt{-1}$. They extend the idea of the one-dimensional number line to the two-dimensional complex plane by using the number line for the *real part* (x-values) and adding a vertical axis for the *imaginary part* (y-values). The graphical representation of the complex number z in the complex plane is sketched in Figure A.1.

If $z = x + i \cdot y$ is a complex number, then x is called *real part* of z, that is, $\text{Re}(z) = x$. Analogously, $\text{Im}(z) = y$ is called the *imaginary part* of z.

If $z = x + i \cdot y$ is a complex number, then $z^* = x - i \cdot y$ is the complex number that lies in the conjugated plane and is therefore called *complex conjugate number.*

In the field $\mathbb{C}$ of complex numbers, there are two operations defined: "+", meaning addition, and "•", meaning multiplication.

A.1 Addition

The addition of the complex numbers z_1 and z_2 is defined as

$$z_1 + z_2 = (x_1 + i \cdot y_1) + (x_2 + i \cdot y_2) = (x_1 + x_2) + i \cdot (y_1 + y_2). \tag{A.1}$$

The addition is commutative, that is, $z_1 + z_2 = z_2 + z_1$.
The neutral element of the addition is $n(+) = 0 + i \cdot 0 = 0$.
For the inverse element of the addition inv^+, it is $z + \text{inv}^+(z) = n(+)$, resulting in $\text{inv}^+(z) = -z$.

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

A.2 Multiplication

The multiplication of two complex numbers z_1 and z_2 is defined as

$$z_1 \cdot z_2 = (x_1 + i \cdot y_1) \cdot (x_2 + i \cdot y_2) = (x_1 \cdot x_2 - y_1 \cdot y_2) + i \cdot (x_1 \cdot y_2 + x_2 \cdot y_1). \tag{A.2}$$

The multiplication is commutative, that is, $z_1 \cdot z_2 = z_1 \cdot z_2$.
The neutral element of the multiplication is $n(\bullet) = 1 + i \cdot 0 = 1$.
For the inverse element of the multiplication $\mathrm{inv}^\bullet$, it is $z \cdot \mathrm{inv}^\bullet(z) = n(\bullet)$, resulting in $\mathrm{inv}^\bullet(z) = 1/z$ for all complex numbers $\neq 0$. With the help of the complex conjugate number z^*, it can be expressed as

$$\mathrm{inv}^\bullet(z) = 1/z = z^*/(z \cdot z^*). \tag{A.3}$$

A.3 Modulus

The modulus of a complex number $z = x + i \cdot y$ corresponds to the length of the pointer in Figure A.1. It is the hypotenuse of the triangle formed by the real part x and the imaginary part y as legs of a right-angled triangle. Therefore, the modulus $|z|$ follows as

$$|z| = \sqrt{x^2 + y^2}. \tag{A.4}$$

For all complex numbers with the same modulus, the corresponding pointer ends on the dash–dot circle in Figure A.1. From this graphical representation, we can deduce that

$$z = x + i \cdot y = |z| \cdot (\cos(\theta) + i \cdot \sin(\theta)) = |z| \cdot \exp(i\theta), \tag{A.5}$$

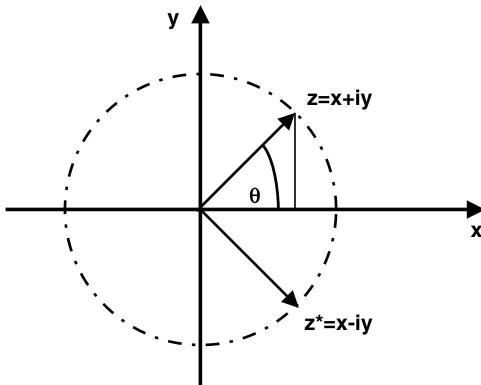


Figure A.1 Graphical representation of complex numbers.

with $\theta = \arg(z) = \tan^{-1}(y/x)$ being the *argument* of z. This is the polar representation of a complex number. The analytical identity $\cos(\theta) + i\cdot\sin(\theta) = \exp(i\cdot\theta)$ allows for the application of power laws when calculating with complex numbers. The multiplication of z with its complex conjugate number z^* yields $|z|^2 = z\cdot z^*$.

A.4 Division

The division of two complex numbers z_1/z_2 can be reformulated into a multiplication of two complex numbers z_1 and $(1/z_2) = \mathrm{inv}^\bullet(z_2)$. Hence, the division is defined as

$$\frac{z_1}{z_2} = z_1 \cdot \mathrm{inv}^\bullet(z_2) = \frac{z_1 \cdot z_2^*}{z_2 z_2^*}. \tag{A.6}$$

A.5 Power *n*

To calculate z^n with n being a real number $n \in \mathbb{R}$, the polar representation of a complex number (A.5) is useful. Then,

$$z^n = |z|^n \cdot (\cos(n\theta) + i \cdot \sin(n\theta)) = |z|^n \cdot \exp(in\theta). \tag{A.7}$$

For $n \in \mathbb{N}$, that is, a positive integer number, we can derive that z^n is given by

$$z^n = \left[x^n - \binom{n}{2} x^{n-2} y^2 + \binom{n}{4} x^{n-4} y^4 - \ldots\right] + i\cdot\left[\binom{n}{1} x^{n-1} y - \binom{n}{3} x^{n-3} y^3 + \ldots\right]. \tag{A.8}$$

A.6 Logarithm

The natural logarithm $\log(z)$ of a complex number z can easily be calculated using again the polar representation (A.5):

$$\log(z) = \log(|z|) + i\theta. \tag{A.9}$$

A.7 Exponentiation

For the complex exponentiation ${z_1}^{z_2}$, we can use the exponential function:

$${z_1}^{z_2} = (\exp(\log(z_1))^{z_2} = \exp(z_2 \cdot (\log(|z_1|) + i\theta_1)). \tag{A.10}$$

A.8
Trigonometric Functions

The trigonometric functions $\sin(z)$ and $\cos(z)$ can be calculated using the analytical identity $\cos(z) + i{\cdot}\sin(z) = \exp(iz)$. Then,

$$\sin(z) = \frac{\exp(iz) - \exp(-iz)}{2 \cdot i} = \sin(x) \cdot \cosh(y) + i \cdot \cos(x) \cdot \sinh(y), \tag{A.11}$$

$$\cos(z) = \frac{\exp(iz) + \exp(-iz)}{2} = \cos(x) \cdot \cosh(y) - i \cdot \sin(x) \cdot \sinh(y). \tag{A.12}$$

Appendix B: Fourier Transform

The Fourier transform is, in principle, a polynomial expansion of a function $f(x)$ that is known at N discrete, generally complex values x_n, with $n = 0, 1, \ldots, N-1$. Using the periodic functions $\exp(-i2\pi k \cdot n/N)$ and $\exp(i2\pi k \cdot n/N)$, with $k = 0, 1, \ldots, N-1$, which form an orthogonal basis in the field of complex numbers $\mathbb{C}$, we can approximate the function $f(x)$ with a polynomial $F(k)$ with these periodic functions as roots of unity of the polynomial. The expansion coefficients of the polynomial $F(k)$ result in

$$X_k = \sum_{n=0}^{N-1} x_n \cdot \exp\left(-i\frac{2\pi}{N}k \cdot n\right), \quad k = 0, 1, \ldots, N-1. \tag{B.1}$$

This expansion is called *discrete Fourier transform* (DFT). Vice versa, the *inverse discrete Fourier transform* (IDFT) gives the discrete values x_n:

$$x_n = \frac{1}{N}\sum_{k=0}^{N-1} X_k \cdot \exp\left(i\frac{2\pi}{N}k \cdot n\right), \quad n = 0, 1, \ldots, N-1. \tag{B.2}$$

It was the mathematician and physicist Jean Baptiste Joseph Fourier who claimed in 1822 in his *Théorie Analytique de la Chaleur* (*The Analytic Theory of Heat*) [185] that any function of a variable, whether continuous or discontinuous, can be expanded in a series of sines of multiples of the variable. A sine function $\sin(z)$ is a linear combination of $\exp(iz)$ and $\exp(-iz)$ (see (A.11)). Although his claim was not correct, he pioneered the above DFT and the following (integral) Fourier transform.
The DFT and IDFT are well suited for periodic functions that fulfill $f(x + T) = f(x)$. In the above equations, the period T is $T = 2\pi N$, so that $x_{n+N} = x_n$ and $X_{k+N} = X_k$. For $T \to \infty$, the discrete Fourier transforms become continuous:

$$F(k) = \int_{-\infty}^{\infty} f(x) \cdot \exp(-i2\pi k \cdot x) \cdot \mathrm{d}x, \tag{B.3}$$

$$f(x) = \int_{-\infty}^{\infty} F(k) \cdot \exp(i2\pi k \cdot x) \cdot \mathrm{d}k. \tag{B.4}$$

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

These *Fourier transform* and *inverse Fourier transform* do hold true both for periodic functions and for nonperiodic functions $f(x)$.

Looking in more detail at the argument of the exponential functions, it becomes clear that if x has a physical dimension (meter m, second s, etc.), the parameter k must have the reciprocal dimension (1/m, 1/s, etc.). The Fourier transform and the discrete Fourier transform always transform from the real space (length, time, etc.) into the reciprocal space (wavenumber, frequency, etc.) and vice versa. They are fundamental in signal processing (time → frequency) and solid-state physics (length → wavenumber), but have many other applications in physics, mathematics, and signal processing.

In the following, we describe some rules of the DFT and the Fourier transform (FT) [186]. For simplification, we introduce the abbrevation $\mathscr{F}$ for the transform of $f(x)$ to $F(k)$, that is, $\mathscr{F}(f(x)) = F(k)$.

B.1 Linearity

$$\mathscr{F}(\alpha \times f(x) + \beta \cdot g(x)) = \alpha \times F(k) + \beta \cdot G(k). \tag{B.5}$$

B.2 Scaling

$$\mathscr{F}(f(x/\alpha)) = |\alpha| \cdot F(\alpha \cdot k). \tag{B.6}$$

B.3 Shifting

$$\mathscr{F}(\alpha \cdot f(x) + \beta) = (1/\alpha) \cdot \exp(ik\beta/\alpha) \cdot F(k/\alpha), \quad \text{with } \alpha \neq 0. \tag{B.7}$$

B.4 Damping

$$\mathscr{F}(f(\alpha \cdot x) \cdot \exp(i\beta \cdot x)) = (1/\alpha) \cdot F((k-\beta)/\alpha), \quad \text{with } \alpha > 0. \tag{B.8}$$

B.5 Convolution

$$\mathscr{F}(f(x) \otimes g(x)) = \mathscr{F}(f(x)) \cdot \mathscr{F}(g(x)) = F(k) \cdot G(k). \tag{B.9}$$

Convolution in one domain is equivalent to multiplication in the other domain and vice versa.

B.6 Plancherel Theorem and Parseval's Theorem

For two functions $f(x)$ and $g(x)$ in real space, the Fourier transform of $f(x)\cdot g(x)^*$ yields the same result as the inverse Fourier transform of $F(k)\cdot G(k)^*$

$$\int_{-\infty}^{\infty} f(x)\cdot g(x)^* \mathrm{d}x = \int_{-\infty}^{\infty} F(k)\cdot G(k)^* \mathrm{d}k \tag{B.10}$$

or for the discrete Fourier transform

$$\sum_{n=0}^{N-1} x_n \cdot y_n^* = \frac{1}{N}\sum_{k=0}^{N-1} X_k \cdot Y_k^*, \tag{B.11}$$

where the asterisk denotes the complex conjugate. Parseval's theorem is a special case of the Plancherel theorem for $g(x)=f(x)$ or $y_n = x_n$

$$\int_{-\infty}^{\infty} |f(x)|^2 \mathrm{d}x = \int_{-\infty}^{\infty} |F(k)|^2 \mathrm{d}k \tag{B.12}$$

or for the discrete Fourier transform

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N}\sum_{k=0}^{N-1} |X_k|^2. \tag{B.13}$$

The exponential form of the discrete Fourier transform allows for negative frequency components. Negative frequencies follow the rule of symmetry: for real signals, negative frequency components are mirror images of the positive frequency components. That means if we have N sampling points in the real space, we get $N/2-1$ negative frequencies, $N/2-1$ positive frequencies, and a sample at frequency $k=0$ that is degenerated twice. This first sample X_0 is the Fourier coefficient of the DC component in the signal, more commonly known as the average of the input series.

The highest positive frequency sample $X_{N/2-1}$ is called *Nyquist frequency*. If the input signal of the DFT contains frequency components higher than this Nyquist frequency, they will encounter a "folding" about the Nyquist frequency, back into lower frequencies. This is called *aliasing* because the higher frequency appears to be a lower frequency. For example, if the Nyquist frequency is 10 kHz, an 11 kHz signal will fold, or alias, to 9 kHz. In that case, however, the original signal cannot be reconstructed from the frequencies in the frequency domain. If the original signal, however, does not contain any frequencies higher than the Nyquist frequency, it can be perfectly reproduced by sampling the signal at a rate of $\Delta x = 1/(2\cdot X_{N/2-1})$. This is the *Nyquist–Shannon sampling theorem*, after Harry Nyquist and Claude Shannon.

Another problem arising with the discrete Fourier transform is the *leakage effect*. An input signal can be transformed exactly with the DFT only if the input signal can

be continued periodically. Otherwise, it contains frequencies that do not belong to the frequencies calculated from the DFT. Then, these frequencies are approximated only by the frequencies of the DFT in the vicinity of this frequency. The contained energy gets distributed over these frequencies. This effect is called leakage. The limiting or clipping of the input data to a certain interval almost always results in leakage since this clipping acts like a convolution of the input data with the sinc function $\sin(x)/x$. Moreover, the probability for a nonperiodic signal within the clipped region is high. Leakage can be minimized by multiplying the input signal by a *window function*, for example, Hamming, von Hann (or Hanning or raised cosine), Blackman, Blackman–Harris, Welch, or Kaiser window, that tapers the amplitudes at each end of the data interval. For an overview on window functions, we refer, for example, to Ref. [135].

The *fast Fourier transform* (FFT) is also a discrete Fourier transform algorithm but which reduces the number of computations needed for N points from $2N^2$ to $2N{\cdot}\log_2(N)$, where $\log_2$ is the 2-based logarithm. The basic idea to achieve this reduction is to break up a transform of length N into N_1 transforms of length N_2 with $N_1 N_2 = N$. Cooley and Tukey [133] developed an algorithm for $N_1 = 2$ and $N_2 = N/2$ that allows to treat even-numbered points and odd-numbered points separately according to the Danielson–Lanczos lemma [134]. According to the way the reduction is done, fast Fourier transform algorithms generally fall into two classes: *decimation in frequency* (DIF, Sande–Tukey algorithm) and *decimation in time* (DIT, Cooley–Tukey algorithm).

Starting with the general expression for the discrete Fourier transformed coefficient X_k (B.1), the Sande–Tukey algorithm (DIF) considers the even-numbered terms X_{2k} and the odd-numbered terms X_{2k+1} as

$$\begin{aligned} X_{2k} &= \sum_{n=0}^{(N/2)-1} \left(x_n \cdot \exp\left(-i\frac{2\pi}{N}k \cdot n\right) + x_{n+N/2} \cdot \exp\left(-i\frac{2\pi}{N}k \cdot (n+N/2)\right)\right) \\ &= \sum_{n=0}^{(N/2)-1} \left(x_n + x_{n+N/2}\right) \cdot \exp\left(-i\frac{4\pi}{N}k \cdot n\right) \end{aligned} \tag{B.14}$$

and

$$X_{2k+1} = \sum_{n=0}^{(N/2)-1} \left(x_n - x_{n+N/2}\right) \cdot \exp\left(-i\frac{2\pi}{N}n\right)\exp\left(-i\frac{4\pi}{N}k \cdot n\right). \tag{B.15}$$

For both one obtains a reduction on $N/2$ terms in the sum. This reduction can be continued until only one term has left in the sum. As the number of data to be calculated in the frequency domain gets halved at each step, this method is called *decimation in frequency*. The Sande–Tukey algorithm first transforms and then rearranges the output values.

The Cooley–Tukey algorithm (DIT) first rearranges the input elements in bit-reversed order and then builds the output transform, for which it is called *decimation in time*. The Cooley–Tukey algorithm considers the term X_k as

$$X_k = \sum_{n=0}^{N-1} x_n \cdot \exp\left(-i\frac{2\pi}{N}k \cdot n\right)$$
$$= \sum_{n=0}^{(N/2)-1} x_{2n} \cdot \exp\left(-i\frac{4\pi}{N}k \cdot n\right) + \exp\left(-i\frac{2\pi}{N}k\right) \cdot \sum_{n=0}^{(N/2)-1} x_{2n+1} \cdot \exp\left(-i\frac{4\pi}{N}k \cdot n\right). \quad \text{(B.16)}$$

This division into even and odd coefficients can also be continued until only one term has left in the sums.
For further reading on Fourier series, discrete Fourier transforms, and Fourier transforms, we refer, for example, to Refs [136, 187–189].

Appendix C: Levenberg–Marquardt Algorithm

The primary application of the Levenberg–Marquardt algorithm is in the least squares curve fitting problem. Having data points (x_i, y_i) of a set of N measured data, the maximum likelihood estimate of the model parameters $\mathbf{a} = \{a_1, \ldots, a_M\}$ is obtained by minimizing the quantity chi-squared χ^2:

$$\chi^2(\boldsymbol{a}) = \sum_{i=1}^{N} (y_i - f(x_i, \boldsymbol{a}))^2. \tag{C.1}$$

Here, we omitted for simplicity the individual standard deviation of each measured data point.

Like other numeric minimization algorithms, the Levenberg–Marquardt algorithm is an iterative procedure. To start a minimization, the user has to provide an initial guess for the parameters vector $\mathbf{a} = \{a_1, \ldots, a_M\}$. At each iteration step, the parameter vector $\mathbf{a}$ is replaced by a new estimate $\mathbf{a} + \delta$. To determine δ, the function $f(x, \mathbf{a} + \delta)$ is approximated linearly:

$$f(x, \mathbf{a} + \delta) \approx f(x, \mathbf{a}) + \underline{\underline{\mathbf{J}}} \cdot \delta. \tag{C.2}$$

The matrix $\underline{\underline{\mathbf{J}}}$ is the *Jacobian matrix* containing the partial derivatives of the function f according to the parameter a_j:

$$J_{ij} = \frac{\partial f(x_i, \mathbf{a})}{\partial a_j}. \tag{C.3}$$

From the first-order approximation of $f(x, \mathbf{a} + \delta)$ in (C.2), we obtain for χ^2:

$$\chi^2(\mathbf{a} + \delta) = \sum_{i=1}^{N} (y_i - f(x_i, \mathbf{a}) - \mathbf{J_i} \cdot \delta)^2 \tag{C.4}$$

or in vector notation

$$\chi^2(\mathbf{a} + \delta) \approx \|\mathbf{y} - \mathbf{f}(\mathbf{a}) - \underline{\underline{\mathbf{J}}} \cdot \delta\|. \tag{C.5}$$

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

Taking the derivative with respect to δ and setting the result to zero to find the minimum gives

$$\left(\underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot \underline{\underline{\mathbf{J}}}\right) \cdot \delta = \underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot (\mathbf{y} - \mathbf{f}(\mathbf{a})). \tag{C.6}$$

This is the *Gauss–Newton algorithm* to solve a set of linear equations for δ. Levenberg [138] replaced this equation by

$$\left(\underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot \underline{\underline{\mathbf{J}}} - \lambda \cdot \underline{\underline{\Im}}\right) \cdot \delta = \underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot (\mathbf{y} - \mathbf{f}(\mathbf{a})), \tag{C.7}$$

where $\underline{\underline{\Im}}$ is the identity matrix. The (nonnegative) damping factor λ is adjusted at each iteration. If the reduction of χ^2 is rapid, a smaller value of λ can be used, and the algorithm is similar to the Gauss–Newton algorithm. Vice versa, if the iteration is low, λ can be increased and the step will be taken approximately in the direction of the gradient.

Marquardt [139] improved the algorithm by scaling each component of the gradient according to the curvature so that there is larger movement along the directions where the gradient is smaller. This avoids slow convergence in the direction of small gradient. For that purpose, Marquardt replaced the identity matrix $\underline{\underline{\Im}}$ with the diagonal matrix consisting of the diagonal elements of $\left(\underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot \underline{\underline{\mathbf{J}}}\right)$, resulting in the Levenberg–Marquardt algorithm:

$$\left(\underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot \underline{\underline{\mathbf{J}}} + \lambda \cdot \left(\underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot \underline{\underline{\mathbf{J}}}\right)\right) \cdot \delta = \underline{\underline{\mathbf{J}}}^{\mathrm{T}} \cdot (\mathbf{y} - \mathbf{f}(\mathbf{a})). \tag{C.8}$$

The choice of the damping factor λ is not obvious at all. Marquardt recommended starting with a value $\lambda = \lambda_0$ and a factor $\nu > 1$. Depending on the value of χ^2, the damping factor will be replaced after each step by either λ/ν or $\lambda \cdot \nu$.

The Levenberg–Marquardt algorithm is a very popular curve-fitting algorithm used in many software applications for solving generic curve-fitting problems. However, also beware of that this algorithm finds only a local minimum like all other iterative procedures, not a global minimum.

Appendix D: Downhill Simplex Algorithm

This algorithm is based on a simplex, the simplest volume in the N-dimensional parameter area, which is stretched from $N + 1$ points. Given a continuous function $y = f(x_1, \ldots, x_N)$ of N variables $\mathbf{x} = \{x_1, \ldots, x_N\}$. The goal is to find a local minimum y_m of this function with corresponding variables $\mathbf{x^m}$. For that purpose, we construct a simplex of $N + 1$ points with vectors $\mathbf{x^1}, \ldots, \mathbf{x^N}, \mathbf{x^{N+1}}$, with $\mathbf{x^i} = \mathbf{x^0} + \lambda \cdot \mathbf{e^i}$.

The procedure is now as follows. After having generated the start simplex, the best point ($y_{\min}$, $x_{\min}$), the worst point ($y_{\max}$, $x_{\max}$), and the second-worst point (y_v, x_v) are determined. Then, the mirror center

$$\mathbf{x^s} = \frac{1}{N} \sum_{\mathbf{x^i} \neq \mathbf{x^{max}}} \mathbf{x^i} \tag{D.1}$$

is determined from all points except the worst point. The first step to generate a new simplex with lower volume is the reflection of the worst point at the mirror center:

$$\mathbf{x^r} = \mathbf{x^s} - \alpha(\mathbf{x^{max}} - \mathbf{x^s}). \tag{D.2}$$

There are three other methods to construct a new simplex:

- the expansion to accelerate the reduction of the simplex to a simplex of smaller volume,
- the contraction to keep the simplex small, and
- the compression around the actual best point.

All four methods are used repeatedly until the best point is obtained. Figure D.1 illustrates all four steps for a three point simplex from $N = 2$ parameters.
After the first reflection, the expansion point

$$\mathbf{x^e} = \mathbf{x^s} - \gamma(\mathbf{x^r} - \mathbf{x^s}). \tag{D.3}$$

is determined and compared with (y_r, $\mathbf{x^r}$) to determine the next steps. The following flow chart in Figure D.2 illustrates the complete algorithm.

The coordinate changes of the parameters during the used steps are made using the Nelder–Mead parameters α, β, and γ, usually set to 1, 0.5, and 2. The iteration is as long resumed until a convergence criterion is fulfilled. The procedure converges approximately linear and is thus not extremely fast but durable.

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

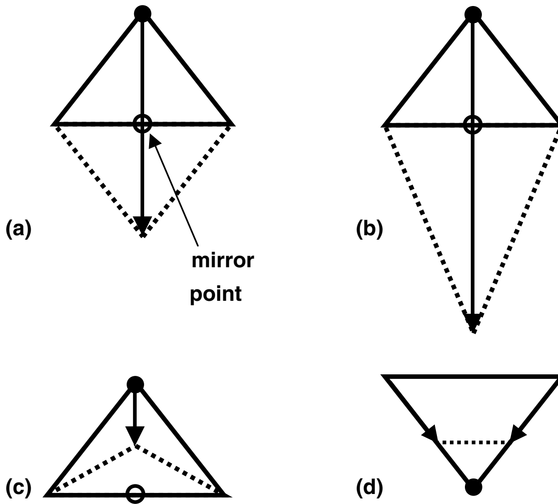


Figure D.1 Illustration of the four methods in the downhill simplex method to define new points of the simplex. (a) Reflection, (b) expansion, (c) contraction, and (d) compression.

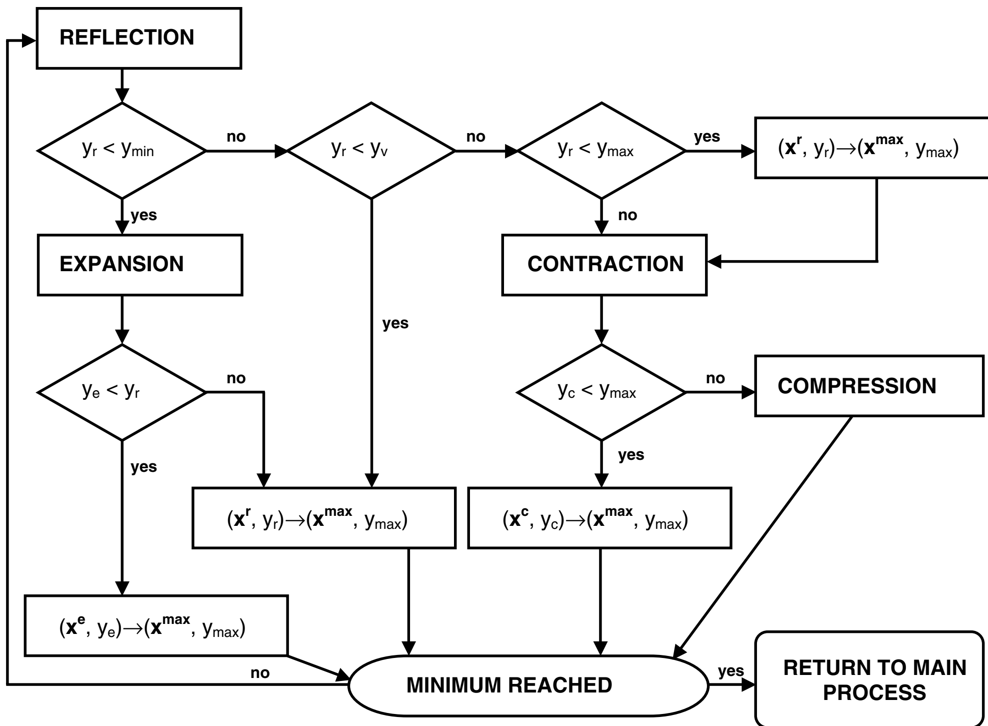


Figure D.2 Flowchart of the downhill simplex algorithm.

References

1. Tompkins, H.G. and McGahan, W.A. (1999) *Spectroscopic Ellipsometry and Reflectometry*, John Wiley & Sons, Inc., NewYork.
2. Drude, P. (1889) Ueber Oberflaechenschichten, I. Theil. *Ann. Phys. Chem.*, **36**, 532–560; II. Theil, 865–897.
3. Azzam, R.M.A. and Bashara, N.M. (1987) *Ellipsometry and Polarized Light*, 2nd edn, North-Holland, Amsterdam.
4. Tompkins, H.G. (1993) *A User's Guide to Ellipsometry*, Academic Press, San Diego; (2006) Dover Publications, Inc., New York.
5. Tompkins, H.G., Irene, E.A., and Haber, E.A. (2005) *Handbook of Ellipsometry (Materials Science and Process Technology)*, William Andrew Inc., New York.
6. Tompkins, H.G. and Irene, E.A. (eds) (2006) *Handbook of Ellipsometry*, Springer, Berlin.
7. Fujiwara, H. (2007) *Spectroscopic Ellipsometry: Principles and Applications*, 1st edn, John Wiley & Sons, Inc., New York.
8. For more information on the software MQLayer, MQColor, and MQNandK, please contact the author at ulmi.quinten@t-online.de.
9. Hertz, H. (1888) Die Kräfte elektrischer Schwingungen, behandelt nach der Maxwellschen Theorie. *Ann. Phys.*, **36**, 1–22.
10. Righi, A. (1901) Sui campi elettromagnetici e particolarmente su quelli creati da cariche elettriche o da poli magnetici in movimento. *Il Nuovo Cimento*, **2**, 104–121.
11. Pedrotti, F., Pedrotti, L., Bausch, W., and Schmidt, H. (2005) *Optik für Ingenieure, Grundlagen*, 3rd edn, Springer, Berlin.
12. Kogelnik, H. (1969) Coupled wave theory for thick hologram gratings. *Bell Syst. Tech. J.*, **48**, 2909–2947.
13. Gerritsen, H.J., Thornton, D.K., and Bolton, S.R. (1991) Application of Kogelnik's two-wave theory to deep, slanted, highly efficient, relief transmission gratings. *Appl. Opt.*, **30**, 807–814.
14. Magnusson, R. and Gaylord, T.K. (1977) Analysis of multiwave diffraction of thick gratings. *J. Opt. Soc. Am.*, **67**, 1165–1170.
15. Moharam, M.G. and Gaylord, T.K. (1982) Diffraction analysis of dielectric surface-relief gratings. *J. Opt. Soc. Am.*, **72**, 1385–1392.
16. Moharam, M.G. and Gaylord, T.K. (1983) Three-dimensional vector coupled-wave analysis of planar-grating diffraction. *J. Opt. Soc. Am.*, **73**, 1105–1112.
17. Chang, M. and George, N. (1970) Holographic dielectric grating: theory and practice. *Appl. Opt.*, **9**, 713–718.
18. Alferness, R. (1975) Analysis of optical propagation in thick holographic gratings. *Appl. Phys.*, **7**, 29–33.
19. Alferness, R. (1975) Equivalence of the thin-grating decomposition and coupled wave analysis of thick holographic gratings. *Opt. Comm.*, **15**, 209–212.
20. van de Hulst, H.C. (1981) *Light Scattering by Small Particles*, Dover Publications, Inc., New York.
21. Kerker, M. (1969) *The Scattering of Light*, Academic Press, San Diego.

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

22 Bohren, C.F. and Huffman, D.R. (1983) *Absorption and Scattering of Light by Small Particles*, John Wiley & Sons, Inc., New York.

23 Mishchenko, M.I., Hovenier, J.W. and Travis, L.D. (eds) (2000) *Light Scattering by Nonspherical Particles*, Academic Press, San Diego.

24 Mishchenko, M.I., Travis, L.D., and Lacis, A.A. (2002) *Scattering, Absorption, and Emission of Light by Small Particles*, Cambridge University Press, Cambridge.

25 Quinten, M. (2011) *Optical Properties of Nanoparticle Systems: Mie and Beyond*, Wiley-VCH, Berlin.

26 Lorentz, H.A. (1895) *Versuch einer Theorie der Electrischen und Optischen Erscheinungen in Bewegten Körpern*, E. J. Brill, Leiden.

27 Brendel, R. and Bormann, D. (1992) An infrared dielectric function model for amorphous solids. *J. Appl. Phys.*, **71**, 1–6.

28 Kim, C.C., Garland, J.W., Abad, H., and Raccah, P.M. (1992) Modeling the optical dielectric function of semiconductors: extension of the critical-point parabolic-band approximation. *Phys. Rev.*, **B45**, 11749–11767.

29 Drude, P. (1900) Zur Elektronentheorie der Metalle. Part 1. *Ann. Phys.*, **306**, 566–613.

30 Drude, P. (1900) Zur Elektronentheorie der Metalle. Part 2. *Ann. Phys.*, **308**, 369–402.

31 Kittel, C. (1995) *Introduction to Solid State Physics*, 7th edn, John Wiley & Sons, Inc., New York.

32 Allen, P.B. (1971) Electron–phonon effects in the infrared properties of metals. *Phys. Rev.*, **B3**, 305–320.

33 Tauc, J., Grigorovici, R., and Vancu, A. (1966) Optical properties and electronic structure of amorphous germanium. *Phys. Stat. Sol.*, **15**, 627–637.

34 Jellison, G.E., Jr. and Modine, F.A. (1996) Parametrization of the optical functions of amorphous materials in the interband region. *Appl. Phys. Lett.*, **69**, 371–373; Erratum, *Appl. Phys. Lett.*, **69**, 2137 (1996).

35 O'Leary, S.K., Johnson, S.R., and Lim, P.K. (1997) The relationship between the distribution of electronic states, the optical absorption spectrum of an amorphous semiconductor: an empirical analysis. *J. Appl. Phys.*, **82**, 3334–3340.

36 Forouhi, A.R. and Bloomer, I. (1986) Optical dispersion relations for amorphous semiconductors and amorphous dielectrics. *Phys. Rev.*, **B34**, 7018–7026.

37 Humlicek, J., Carriga, M., Alonso, M.I., and Cardona, M. (1989) Optical spectra of $Si_xGe_{(1-x)}$ alloys. *J. Appl. Phys.*, **65**, 2827–2832.

38 Palik, E.D. (ed.) (1985) *Handbook of Optical Constants of Solids I*, Academic Press, San Diego.

39 Gerfin, T. and Grätzel, M. (1996) Optical properties of tin-doped indium oxide determined by spectroscopic ellipsometry. *J. Appl. Phys.*, **79**, 1722–1729.

40 Kronig, R. (1926) On the theory of the dispersion of X-rays. *J. Opt. Soc. Am.*, **12**, 547–557.

41 Kramers, H.A. (1927) La diffusion de la lumière par les atomes. *Atti Cong. Intern. Fisica* (Transactions of Volta Centenary Congress, Como) **2**, 545–557.

42 Sellmeier, W. (1871) Zur Erklärung der abnormen Farbenfolge im Spectrum einiger Substanzen. *Ann. Phys. Chem.*, **219**, 272–282.

43 Cauchy, A.L. (1830) Sur la réfraction et la réflexion de la lumière. *Bulletin de Férussac, tomé*, **14**, 6–10.

44 Cauchy, A.L. (1836) *Mémoire sur la Dispersion de la Lumière*, J. G. Calve, Prague.

45 Urbach, F. (1953) The long wavelength edge of photographic sensitivity and of the electronic absorption of solids. *Phys. Rev.*, **92**, 1324.

46 Conrady, A.E. (1929) *Applied Optics and Optical Design*, Oxford University Press, London.

47 Conrady, A.E. (1958) *Applied Optics and Optical Design. Part I*, Dover Publications, Inc., NewYork.

48 Conrady, A.E. and Kingslake, R. (1960) *Applied Optics and Optical Design. Part II*, Dover Publications, Inc., NewYork.

49 Herzberger, M. (1959) Colour correction in optical systems and a new dispersion formula. *Opt. Acta (London)*, **6**, 197–215.

50 Herzberger, M. and Salzberg, C.D. (1962) Refractive indices of infrared optical

materials and color correction of infrared lenses. *J. Opt. Soc. Am.*, **52**, 420–424.

51 Hartmann, J. (1898) Ueber eine einfache Interpolationsformel für das prismatische Spectrum. *Publ. Astrophysik. Observat. Potsdam*, **42**, 4–29.

52 Bass, M. (ed.) (1994) *Handbook of Optics, Vol. 2: Devices, Measurements, and Properties*, 2nd edn, McGraw-Hill Professional, New York.

53 Lichtenecker, K. (1926) Die Dielektrizitätskonstante natürlicher und künstlicher Mischkörper. *Phys. Z.*, **27**, 115–158.

54 Beer, A. (1853) *Einleitung in die Höhere Optik*, Vieweg, Braunschweig.

55 Gladstone, J.H. and Dale, T.P. (1863) Researches on the refraction and dispersion and sensitiveness of liquids. *Phil. Trans. R. Soc. Lond.*, **153**, 317–343.

56 Landau, L.D. and Lifshitz, E.M. (1974) *Lehrbuch der Theoretischen Physik VIII: Elektrodynamik der Kontinua*, Akademie Verlag, Berlin.

57 Looyenga, H. (1965) Dielectric constants of heterogeneous mixtures. *Physica*, **31**, 401–406.

58 Garnett, J.C.M. (1904) Colours in metal glasses and in metallic films. *Phil. Trans. R. Soc. Lond.*, **A203**, 385–420.

59 Piredda, G., Smith, D.D., Wendling, B., and Boyd, R.W. (2008) Nonlinear optical properties of a gold–silica composite with high gold fill fraction and the sign change of its nonlinear absorption coefficient. *J. Opt. Soc. Am.*, **B25**, 945–950.

60 Bruggeman, D.A.G. (1935) Berechnung verschiedener physikalischer Konstanten von heterogenen Substanzen. I. Dielektrizitätskonstanten und Leitfähigkeiten der Mischkörper aus isotropen Substanzen. *Ann. Phys. (Leipzig)*, **24**, 636–679.

61 Aspnes, D.E. (1982) Optical properties of thin films. *Thin Solid Films*, **89**, 249–262.

62 Aspnes, D.E., Theeten, J.B., and Hottier, F. (1979) Investigation of effective-medium models of microscopic surface roughness by spectroscopic ellipsometry. *Phys. Rev.*, **B20**, 3292–3302.

63 Yeh, P. (1988) *Optical Waves in Layered Media*, John Wiley & Sons, Inc., New York.

64 Rancourt, J.D. (1996) *Optical Thin Films: User Handbook*, SPIE Optical Engineering Press, Bellingham.

65 Stenzel, O. (2005) *The Physics of Thin Film Optical Spectra: An Introduction, Springer Series in Surface Sciences*, Springer, Berlin.

66 Optical Glass Catalogue (2011) SCHOTT AG, Mainz, Germany, www.schott.com/advanced_optics.

67 Palik, E.D. (ed.) (1991) *Handbook of Optical Constants of Solids II*, Academic Press, San Diego.

68 Beckmann, P. and Spizzichino, A. (1963) *The Scattering of Electromagnetic Waves from Rough Surfaces*, Pergamon Press, New York.

69 Stover, J.C. (1995) *Optical Scattering: Measurement and Analysis*, 2nd edn, SPIE Optical Engineering Press, Bellingham.

70 Debye, P. (1913) Über die Intensitätsverteilung in den mit Röntgenstrahlen erzeugten Interferenzbildern. *Verh. Dtsch. Phys. Ges.*, **15**, 738–752.

71 Waller, I. (1923) Zur Frage der Einwirkung der Wärmebewegung auf die Interferenz von Röntgenstrahlen. *Z. Phys.*, **17**, 398–408.

72 Mitsas, C.L. and Siapkas, D.I. (1995) Generalized matrix method for analysis of coherent and incoherent reflectance and transmittance of multilayer structures with rough surfaces, interfaces, and finite substrates. *Appl. Opt.*, **34**, 1678–1683.

73 Katsidis, C.C. and Siapkas, D.I. (2002) General transfer-matrix method for optical multilayer systems with coherent, partially coherent, and incoherent interference. *Appl. Opt.*, **41**, 3978–3987.

74 Filinski, I. (1972) The effects of sample imperfections on optical spectra. *Phys. Stat. Sol. (b)*, **49**, 577–588.

75 Szczyrbowski, J. and Czapla, A. (1977) Optical absorption in d.c. sputtered InAs films. *Thin Solid Films*, **47**, 127–137.

76 Névot, L. and Croce, P. (1980) Characterisation des surfaces par reflection rasante de rayon X, application a etude du polissage de quelque verres silicates. *Rev. Phys. Appl. (Paris)*, **15**, 761–779.

77 Abelès, F. (1950) Recherches sur la propagation des ondes

electromagnetique sinusoidales dans les milieux stratifies. *Ann. Phys.*, **5**, 596–640; 706–782.

78 Abelès, F. (1950) La détermination de l'indice et de l'épaineur des couches minces transparentes. *J. Phys. Radium*, **11**, 310–314.

79 Harbecke, B. (1986) Coherent and incoherent reflection and transmission of multilayer structures. *Appl. Phys.*, **B39**, 165–170.

80 Ohta, K. and Ishida, H. (1990) Matrix formalism for calculation of electric field intensity of light in stratified multilayered films. *Appl. Opt.*, **29**, 1952–1959.

81 Ohta, K. and Ishida, H. (1990) Matrix formalism for calculation of the light beam intensity in stratified multilayered films, and its use in the analysis of emission spectra. *Appl. Opt.*, **29**, 2466–2473.

82 Prentice, J.S.C. (2000) Coherent, partially coherent and incoherent light absorption in thin-film multilayer structures. *J. Phys.*, **D33**, 3139–3145.

83 Carniglia, C.K. and Jensen, D.G. (2002) Single-layer model for surface roughness. *Appl. Opt.*, **41**, 3167–3171.

84 Carniglia, C.K. (1979) Scalar scattering theory for multilayer optical coatings. *Opt. Eng.*, **18**, 104–115.

85 Bennett, J.M. and Mattsson, L. (1999) *Introduction to Surface Roughness and Scattering*, 2nd edn, Optical Society of America, Washington, DC.

86 Berreman, D.W. (1972) Optics in stratified and anisotropic media: 4×4-matrix formulation. *J. Opt. Soc. Am.*, **62**, 502–510.

87 Lin-Chung, P.J. and Teitler, S. (1984) 4×4 matrix formalisms for optics in stratified anisotropic media. *J. Opt. Soc. Am.*, **A1**, 703–705.

88 Yeh, P. (1979) Electromagnetic propagation in birefringent layered media. *J. Opt. Soc. Am.*, **69**, 742–755.

89 Yeh, P. (1980) Optics of anisotropic layered media: a new 4×4 matrix algebra. *Surf. Sci.*, **96**, 41–53.

90 Visnovsky, S. (1986) Magneto-optical ellipsometry. *Czech. J. Phys.*, **B36**, 625–650.

91 Visnovsky, S. (1991) Optics of magnetic multilayers. *Czech. J. Phys.*, **B41**, 663–694.

92 Visnovsky, S., Lopusnik, R., Bauer, M., Bok, J., Fassbender, J., and Hillebrands, B. (2001) Magnetooptic ellipsometry in multilayers at arbitrary magnetization. *Opt. Expr.*, **9**, 121–135.

93 Schubert, M., Tiwald, T.E., and Woollam, J.A. (1999) Explicit solutions for the optical properties of arbitrary magneto-optic materials in generalized ellipsometry. *Appl. Opt.*, **38**, 177–187.

94 Postava, K., Pistora, J., and Visnovsky, S. (1999) Magneto-optical effects in ultrathin structures at transversal magnetization. *Czech. J. Phys.*, **B49**, 1185–1204.

95 Mansuripur, M. (1990) Analysis of multilayer thin-film structures containing magneto-optic and anisotropic media at oblique incidence using 2×2 matrices. *J. Appl. Phys.*, **67**, 6466–6475.

96 Yeh, P. (1982) Extended Jones matrix method. *J. Opt. Soc. Am.*, **72**, 507–513.

97 Gu, C. and Yeh, P. (1993) Extended Jones matrix method. II. *J. Opt. Soc. Am.*, **A10**, 966–973.

98 Cojocaru, E. (1997) Generalized Abelès relations for an anisotropic thin film of an arbitrary dielectric tensor. *Appl. Opt.*, **36**, 2825–2829.

99 Cojocaru, E. (2000) Simple recurrence matrix relations for multilayer anisotropic thin films. *Appl. Opt.*, **39**, 141–148.

100 Postava, K., Yamaguchi, T., and Kantor, R. (2002) Matrix description of coherent and incoherent light reflection and transmission by anisotropic multilayer structures. *Appl. Opt.*, **41**, 2521–2531.

101 Visnovsky, S. and Krishnan, R. (1981) Complex Faraday effect in multilayer structures. *J. Opt. Soc. Am.*, **71**, 315–319.

102 Röseler, A. (1990) *Infrared Spectroscopic Ellipsometry*, Akademie-Verlag, Berlin.

103 Woollam & Co. Inc., J.A., 645 M Street, Suite 102, Lincoln, NE 68508-2243 USA, www.jawoollam.com/tutorial_1.html.

104 Collins, R.W. (1990) Automatic rotating element ellipsometers: calibration, operation, and real-time applications. *Rev. Sci. Instrum.*, **61**, 2029–2062.

105 Jellison, G.E., Jr. (1993) Data analysis for spectroscopic ellipsometry. *Thin Solid Films*, **234**, 416–422.

106 Jellison, G.E., Jr. (1998) Spectroscopic ellipsometry data analysis: measured versus calculated quantities. *Thin Solid Films*, **313–314**, 33–39.

107 Decker, M.M. and Mueller, H. (1957) Transmitting data by light modulation. *Control Eng.*, **4**, 63–67.

108 Billardon, M. and Badoz, J. (1966) Birefringence modulator. *C. R. Acad. Sci.*, **B262**, 1672–1675.

109 Kemp, J.C. (1969) Piezo-optical birefringence modulators: new use for a long known effect. *J. Opt. Soc. Am.*, **59**, 950–954.

110 Jasperson, S.N. and Schnatterly, S.E. (1969) An improved method for high reflectivity ellipsometry based on a new polarization modulation technique. *Rev. Sci. Instrum.*, **40**, 761–767.

111 Jasperson, S.N., Burge, D.K., and O'Handley, R.C. (1973) A modulated ellipsometer for studying thin film optical properties and surface dynamics. *Surf. Sci.*, **37**, 548–558.

112 Jellison, G.E., Jr. and Modine, F.A. (1990) Two-channel polarization modulation ellipsometer. *Appl. Opt.*, **29**, 959–974.

113 Jellison, G.E., Jr. and Modine, F.A. (1997) Two-modulator generalized ellipsometry: experiment and calibration. *Appl. Opt.*, **36**, 8184–8189.

114 Jellison, G.E., Jr. and Modine, F.A. (1997) Two-modulator generalized ellipsometry: theory. *Appl. Opt.*, **36**, 8190–8198.

115 Woollam, J.A. and Snyder, P.G. (1992) Variable angle spectroscopic ellipsometry, in *Encyclopedia of Materials Characterization: Surfaces, Interfaces, Thin Films* (eds C.R. Brundle, C.A. Evans, and S. Wilson), Butterworth-Heinemann, Boston, pp. 401–411.

116 Woollam, J.A., Johs, B., Herzinger, C.M., Hilfiker, J., Synowicki, R., and Bungay, C.L. (1999) Overview of variable angle spectroscopic ellipsometry (VASE). Part I: basic theory and typical applications. *Crit. Rev. Opt. Sci. Technol.*, **CR72**, 2–28.

117 Johs, B., Woollam, J.A., Herzinger, C.M., Hilfiker, J., Synowicki, R., and Bungay, C.L. (1999) Overview of variable angle spectroscopic ellipsometry (VASE). Part II: advanced applications. *Crit. Rev. Opt. Sci. Technol.*, **CR72**, 29–58.

118 Bader, G., Ashrit, P.V., Girouard, F.E., and Truong, V.-V. (1995) Reflection–transmission photoellipsometry: theory and experiments. *Appl. Opt.*, **34**, 1684–1691.

119 Bader, G., Ashrit, P.V., and Truong, V.-V. (1998) Transmission and reflection ellipsometry of thin films and multilayer systems. *Appl. Opt.*, **37**, 1146–1151.

120 Otto, A. (1968) Excitation of nonradiative surface plasma waves in silver by the method of frustrated total reflection. *Z. Phys.*, **216**, 398–410.

121 Kretschmann, E. (1971) Die Bestimmung optischer Konstanten von Metallen durch Anregung von Oberflächenplasmaschwingungen. *Z. Phys.*, **241**, 313–324.

122 Raether, H. (1980) *Excitation of Plasmons and Interband Transitions by Electrons, Springer Tracts in Modern Physics*, vol. **88**, Springer, Berlin.

123 Raether, H. (1988) *Surface Plasmons on Smooth and Rough Surfaces and on Gratings, Springer Tracts in Modern Physics*, vol. **111**, Springer, Berlin.

124 Tien, P.K., Ulrich, R., and Martin, R.J. (1969) Modes of propagating light waves in thin deposited semiconductor films. *Appl. Phys. Lett.*, **14**, 291–294.

125 Harris, J.H., Shubert, R., and Polky, J.N. (1970) Beam coupling to films. *J. Opt. Soc. Am.*, **60**, 1007–1016.

126 Tien, P.K. and Ulrich, R. (1970) Theory of prism-film coupler and thin-film light guides. *J. Opt. Soc. Am.*, **60**, 1325–1337.

127 Ulrich, R. and Torge, R. (1973) Measurement of thin film parameters with a prism coupler. *Appl. Opt.*, **12**, 2901–2908.

128 Hunsperger, R.G. (1995) *Integrated Optics: Theory and Technology*, 4th edn, Springer, Berlin.

129 Nishihara, H., Haruna, M., and Sunhara, T. (1989) *Optical Integrated Circuits*, McGrawHill, New York.

130 Bludau, W. (1998) *Lichtwellenleiter in Sensorik und Optischer Nachrichtentechnik*, Springer, Berlin.

131 Palmer, C. and Loewen, E. (2005) *Diffraction Grating Handbook*, 6th edn, Newport Corp., Rochester.

132 HORIBA Jobin Yvon (1988) *Diffraction Gratings Ruled and Holographic Handbook*, HORIBA Jobin Yvon Div. d'Instruments S. A., France, and HORIBA Jobin Yvon, Inc., New Jersey, USA.

133 Cooley, J.W. and Tukey, J.W. (1965) An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, **19**, 297–301.

134 Danielson, G.C. and Lanczos, C. (1942) Some improvements in practical Fourier analysis and their application to X-ray scattering from liquids. *J. Franklin Inst.*, **233**, 365–380; 435–452.

135 http:\\en.wikipedia.org/wiki/window-function.

136 Ramirez, R.W. (1985) *The FFT, Fundamentals and Concepts*, Prentice-Hall, New Jersey.

137 Nelder, J.A. and Mead, R. (1965) A simplex method for function minimization. *Comp. J.*, **7**, 308–313.

138 Levenberg, K. (1944) A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, **2**, 164–168.

139 Marquardt, D. (1963) An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, **11**, 431–441.

140 Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2002) *Numerical Recipes in C++: The art of Scientifique Computing*, 2nd edn, Cambridge University Press, Cambridge.

141 Abelès, F. and Theye, M.L. (1966) Methode de calcul des constantes optiques des couches minces absorbantes a partir de mesures de réflexion et de transmission. *Surf. Sci.*, **5**, 325–331.

142 Bennett, J.M. and Booty, M.J. (1966) Computational method for determining n and k for a thin film from the measured reflectance, transmittance, and film thickness. *Appl. Opt.*, **5**, 41–43.

143 Nilsson, P.-O. (1968) Determination of optical constants for intensity measurements at normal incidence. *Appl. Opt.*, **7**, 435–442.

144 Manifacier, N.J.C., Gasiot, J., and Fillard, J.P. (1976) A simple method for the determination of the optical constants n, k, and the thickness of a weakly absorbing thin film. *J. Phys.*, **E9**, 1002–1004.

145 Hjortsberg, A. (1981) Determination of optical constants of absorbing materials using transmission and reflection of thin films on partially metallized substrates: analysis of the new (T, R) technique. *Appl. Opt.*, **20**, 1254–1263.

146 Case, W.E. (1983) Algebraic method for extracting thin-film optical parameters from spectrophotometer measurements. *Appl. Opt.*, **22**, 1832–1836.

147 Stenzel, O., Hopfe, V., and Klobes, P. (1991) Determination of optical parameters for amorphous thin-film materials on semitransparent substrates from transmittance and reflectance measurements. *J. Phys.*, **D24**, 2088–2094.

148 Tomlin, S.G. (1968) Optical reflection and transmission formulae for thin films. *J. Phys.*, **D2**, 1667–1671.

149 Garcia-Castafieda, M. and Safichez-Machet, H. (1989) An iterative and consistent method for the complex refraction index calculation of absorbent thin films. *Thin Solid Films*, **176**, 69–72.

150 McPhedran, R.C., Botten, L.C., McKenzie, D.R., and Netterfield, R.P. (1984) Unambiguous determination of optical constants of absorbing films by reflectance and transmittance measurements. *Appl. Opt.*, **23**, 1197–1205.

151 Nagendra, C.L. and Thutupalli, G.K.M. (1983) Determination of optical properties of absorbing materials: a generalized scheme. *Appl. Opt.*, **22**, 587–591.

152 Nestell, J.E., Jr. and Christy, R.W. (1972) Derivation of optical constants of metals from thin-film measurements at oblique incidence. *Appl. Opt.*, **11**, 643–651.

153 Borgogno, J.P., Lazarides, B., and Roche, P. (1983) An improved method for the determination of the extinction coefficient of thin film materials. *Thin Solids Films*, **102**, 209–220.

154 Elizalde, E. and Rueda, F. (1984) On the determination of the optical constants $n(\lambda)$ and $\kappa(\lambda)$ of thin supported films. *Thin Solid Films*, **122**, 45–57.

155 Ohlidal, I. (1988) Immersion spectroscopic reflectometry of multilayer systems. I. Theory. *J. Opt. Soc. Am.*, **A5**, 459–464.

156 Ohlidal, I., Navratil, K., and Holy, V. (1988) Immersion spectroscopic reflectometry of multilayer systems. II. Experimental results. *J. Opt. Soc. Am.*, **A5**, 465–470.

157 Kihara, T. and Yokomori, K. (1992) Simultaneous measurement of the refractive index and thickness of thin films by *S*-polarized reflectances. *Appl. Opt.*, **31**, 4482–4487.

158 Lamprecht, K., Papousek, W., and Leising, G. (1997) Problem of ambiguity in the determination of optical constants of thin absorbing films from spectroscopic reflectance and transmittance measurements. *Appl. Opt.*, **36**, 6364–6371.

159 Lamminpää, A., Nevas, S., Manoocheri, F., and Ikonen, E. (2006) Characterization of thin films based on reflectance and transmittance measurements at oblique angles of incidence. *Appl. Opt.*, **45**, 1392–1396.

160 Wyszecki, G. and Stiles, W.S. (1982) *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd edn, John Wiley & Sons, Inc., New York.

161 Berger-Schunn, A. (1983) *Practical Color Measurement: A Primer for the Beginner, a Reminder for the Expert*, John Wiley & Sons, Inc., New York.

162 Glöß, D., Frach, P., Gottfried, C., Klinkenberg, S., Liebig, J.-S., Hentsch, W., Liepack, H., and Krug, M. (2008) Multifunctional high-reflective and antireflective layer systems with easy-to-clean properties. *6th International Conference on Coatings on Glass and Plastics (ICCG6) Thin Solid Films*, **516**, 4487–4489.

163 Völkel, L., Schulz-Grosser, M., and Kubitzek, R. (2010) Non-destructive characterisation of AR coatings on curved surfaces. *Photonik Int.*, **1**, 37–39.

164 Strehlke, S., Bastide, S., Guillet, J., and Lévy-Clément, C. (2000) Design of porous silicon antireflection coatings for silicon solar cells. *Mat. Sci. Eng.*, **B69–70**, 81–86.

165 Database of complex refractive indices from SOPRALAB, 7, rue du Moulin des Bruyères, 92400 Courbevoie, France. www.sopra-sa.com.

166 Meyer, M., Koglin, J., and Fries, T. (2004) Bridging the gap between nanometer and meter, in *The Nano-Micro Interface: Bridging the Micro and Nano Worlds* (eds H.-J. Fecht and M. Werner), Wiley-VCH Verlag GmbH, Berlin, Fig. 8, pp. 89–105.

167 Jellison, G.E., Jr., Modine, F.A., White, C.W., Wood, R.F., and Young, R.T. (1981) Optical properties of heavily doped silicon between 1.5 and 4.1 eV. *Phys. Rev. Lett.*, **46**, 1414–1417.

168 Aspnes, D.E., Studna, A.A., and Kinsbron, E. (1984) Dielectric properties of heavily doped crystalline and amorphous silicon from 1.5 to 6.0 eV. *Phys. Rev.*, **B29**, 768–779.

169 Vina, L. and Cardona, M. (1984) Effect of heavy doping on the optical properties and the band structure of silicon. *Phys. Rev.*, **B12**, 6739–6751.

170 Borghesi, A., Chen-Jai, C., Guizzetti, G., Marabelli, F., Nosenzo, L., Regussoni, E., Stella, A., and Ostoja, P. (1985) Infra-red properties of bulk heavily doped silicon. *Il Nuovo Cimento*, **5**, 292–303.

171 Colinge, J.-P. (2004) *Silicon-On-Insulator Technology: Materials to VLSI*, 3rd edn, Springer Science + Business Media, Inc., New York.

172 Schleußner, S. (2003) Cu(In,Ga)Se2 Thin Film Solar Cells with ZrN as a Back Contact, diploma thesis, University of Stuttgart, IPE, Stuttgart.

173 Poortmans, J. and Arkhipov, V. (eds) (2006) *Thin Film Solar Cells: Fabrication, Characterization and Applications*, John Wiley & Sons Ltd., Chichester.

174 Hamakawa, Y. (ed.) (2010) *Thin-Film Solar Cells: Next Generation Photovoltaics and its Applications*, Springer, Berlin.

175 Parkinson, P.S., Settlemyer, K., McStay, I., Park, D.-G., Ramachandran, R., Chudzik, M., Cheng, K., Sung, C.-Y., Chen, F., Strong, A., Papworth, P., and Jammy, R. (2003) Novel techniques for scaling deep trench DRAM Capacitor technology to 0.11 μm and beyond, in *International Symposium on VLSI Technology, Systems, and Applications*, IEEE Press, Piscataway, pp. 21–24.

176 Zaidi, S., Stojakovic, G., Gutmann, A., Bozdog, C., Mantz, U., Charpenay, S.B., and Rosenthal, P.A. (2003) FTIR-based nondestructive method for metrology of depths in polysilicon-filled trenches, in *Proceedings of the SPIE*, vol. 5038, SPIE

Optical Engineering Press, Bellingham, pp. 185–190.

177 Mantz, U. and Kasic, A. (2005) Model-based infrared spectroscopy: new opportunities for in-line process control. *Future Fab*, **19**, 119–123.

178 Rosenthal, P.A., Duran, C., Tower, J., Mazurenko, A., Mantz, U., Weidner, P., and Kasic, A. (2005) Model-based infrared metrology for advanced technology nodes and 300 mm wafer processing, in *Characterization and Metrology for ULSI Technology*, vol. 788 (eds D.G. Seiler, A.C. Diebold, R. McDonald, C.M. Gamer, D. Herr, R.P. Khosla, and E.M. Secula), AIP Conf. Proc. AIP Press, Melville, pp. 620–624.

179 Gostein, M., Rosenthal, P.A., Maznev, A., Kasic, A., Weidner, P., and Guittet, P.-Y. (2006) Measuring deep-trench structures with model-based IR. *Solid State Technol.*, **49**, 38–42.

180 Guittet, P.-Y., Gostein, M., Weidner, P., and Kasic, A. (2006) Model-based infrared reflectometry: in-line applications for DRAM manufacturing. *Future Fab*, **21**, 131–135.

181 Gostein, M. and Maznev, A.A. (2009) Infrared metrology of 3D structures. *Future Fab*, **23**, 106–108.

182 Maznev, A.A., Mazurenko, A., Duran, C., and Gostein, M. (2007) Measuring trench structures for microelectronics with model-based infrared reflectometry, in *Frontiers of Characterization and Metrology for Nanoelectronics*, vol. **931** (eds D.G. Seiler, A.C. Diebold, R. McDonald, C.M. Gamer, D. Herr, R.P. Khosla, and E.M. Secula), AIP Conf. Proc. AIP Press, Melville, pp. 74–78.

183 Le Cunff, D., Höglund, L.J., and Laurent, N. (2011) In-line metrology of high aspect ratio structures with MBIR technique. *Proc. 22nd Annual IEEE/SEMI. Advanced Semiconductor Manufacturing Conference (ASMC 2011), Session 2.1*, IEEE Press, Piscataway

184 Zhang, C., Liu, S., Shi, T., and Tang, Z. (2009) Improved model-based infrared reflectometry for measuring deep trench structures. *J. Opt. Soc. Am.*, **A26**, 2327–2335.

185 Fourier, J.B.J. (1822) *Théorie Analytique de la Chaleur*, Firmin Didot Père et Fils, Paris.

186 Bronstein, I.N., Semendjajew, K.A., Musiol, G., and Mühlig, H. (2001) *Taschenbuch der Mathematik*, Verlag Harri Deutsch, Frankfurt.

187 Körner, T.W. (1988) *Fourier Analysis*, Cambridge University Press, Cambridge.

188 Walker, J.S. (1996) *Fast Fourier Transforms*, 2nd edn, CRC Press, Taylor & Francis Group, Boca Raton.

189 Chu, E. (2008) *Discrete and Continuous Fourier Transforms: Analysis, Applications and Fast Algorithms*, CRC Press and Taylor & Francis Group, Boca Raton.

Index

a
Abbe number 99
Abelès method 77
absorption 65
active pixel sensor 112
aluminum doped zinc oxide 178
angular dispersion 109
anomal dispersion 38
anti reflection coating 150
antimony doped tin oxide 178
AR coating 150
ATO 178
AZO 178

b
birefringent crystal 101
blackbody radiation 94
blaze wavelength 110
blazed gratings 33
Bragg grating 30
Brendel oscillator 38
Bruggeman 54

c
Cauchy formula 51
CCD 113
characteristic matrix 77
chi-squared 131
chromatic aberration 100
chromaticity coordinates 142
CMOS 111
coherent superposition 59
Commission Internationale de l'Éclairage 141
critical angle of total reflection 20
critical dimensions 182

d
Danielson-Lanczos lemma 124, 194
deuterium lamp 97
DFT 191
dielectric function 10, 35
diffraction 21
discrete Fourier transform 191
dispersion integrals 49
downhill simplex algorithm 132, 199
Drude susceptibility 39
dynamic resolution 118

e
electromagnetic waves 7
ellipsometric measurement 85
ellipsometric parameters 86
ellipsometry 85
EMA models 53
evanescent wave 21
exponential Cauchy formula 51
extended Drude model 40
extraordinary ray 101

f
Fast Fourier transform 121, 122, 194
FFT 121, 194
filling factor 53
fluorine doped tin oxide 178
Forouhi and Bloomer 43
Fourier transform 191, 192
free spectral range 26, 109
Fresnel coefficients 19
Fresnel equations 19
FTO 178
full width at half maximum 115

A Practical Guide to Optical Metrology for Thin Films, First Edition. Michael Quinten.
© 2013 Wiley-VCH Verlag GmbH & Co. KGaA. Published 2013 by Wiley-VCH Verlag GmbH & Co. KGaA.

g
ghosts 119
Glan-Taylor prism 101
grating efficiency 27, 110
grating equation 25, 109
grating function 24
grating period 23, 109
grating 107
groove density 23, 109

h
halogen lamps 94
harmonic oscillator model 36
Helmholtz equation 10
high reflection coating 150
holographic reflection gratings 34
holographic transmission gratings 29
HR coating 150
Huygens-Fresnel principle 14

i
IDFT 191
incoherent substrates 78
indium tin oxide 178
interference 15
inverse discrete Fourier transform 191
inverse Fourier transform 192
ITO 178

k
Kim oscillator 39
Kramers-Kronig relations 49

l
lamellar reflection gratings 31
lamellar transmission gratings 27
layer stack 75
leakage 126, 193
lenses 99
Levenberg-Marquardt algorithm 132, 197
linear superposition 15

m
Malus' law 102
Maxwell-Garnett 54
Maxwell's equations 8
MBIR 182
miniaturized spectrometer 107
mirrors 99
Model Based Infrared Reflectometry 182
MOSFET 111

n
normal dispersion 38
numerical aperture 104
Nyquist frequency 193
Nyquist-Shannon sampling theorem 193

o
OJL-model 42
optical fiber 103
optical retarder 102
ordinary ray 101
organic solar cells 180

p
P3HT 180
PEDOT 180
pixel dispersion 115
polarization 13
polarizer 101
power spectral distribution 124
power spectrum 124
propagating wave model 59, 75

r
Rayleigh criterion for rough surfaces 72
Rayleigh criterion 115
reciprocal linear dispersion 109
reflection law 17
reflectometric measurement 81
refractive index 35
regression analysis 121, 131
resolving power 110
roughness 72, 78
r-t-ϕ model 59, 79

s
scattering 34
Schott formula 51
Sellmeier formula 50
signal-to-noise ratio 118
silicon on insulator 174
Snell's law of refraction 18
SNR 118
SOI 174
spectral reflectance measurement 81
spectral resolution 110, 115
spectral transmittance measurement 81
stray light 118
super-luminescent diode 96

t
Tauc-Lorentz model 41
thick substrates 69

total reflection 20
transverse electric mode 11
transverse magnetic mode 10
tristimulus values 141

v
vector harmonics 11
vector wave equation 10
volume grating 30

w
wave equation 7
white light LED 95
Wien's law 94
Wollaston prism 102

x
Xenon high pressure arc lamp 97