SURFACE SCIENCES

O. Stenzel

The Physics of Thin Film Optical Spectra

An Introduction

Springer

SPRINGER SERIES IN SURFACE SCIENCES

Series Editors: G. Ertl, H. Lüth and D.L. Mills

This series covers the whole spectrum of surface sciences, including structure and dynamics of clean and adsorbate-covered surfaces, thin films, basic surface effects, analytical methods and also the physics and chemistry of interfaces. Written by leading researchers in the field, the books are intended primarily for researchers in academia and industry and for graduate students.

38 **Progress in Transmission Electron Microscopy 1**
Concepts and Techniques
Editors: X.-F. Zhang, Z. Zhang

39 **Progress in Transmission Electron Microscopy 2**
Applications in Materials Science
Editors: X.-F. Zhang, Z. Zhang

40 **Giant Magneto-Resistance Devices**
By E. Hirota, H. Sakakima, and K. Inomata

41 **The Physics of Ultra-High-Density Magnetic Recording**
Editors: M.L. Plumer, J. van Ek, and D. Weller

42 **Islands, Mounds and Atoms**
Patterns and Processes in Crystal Growth Far from Equilibrium
By T. Michely and J. Krug

43 **Electronic Properties of Semiconductor Interfaces**
By W. Mönch

44 **The Physics of Thin Film Optical Spectra**
An Introduction
By O. Stenzel

Volumes 1–37 are listed at the end of the book

O. Stenzel

The Physics of Thin Film Optical Spectra

An Introduction

With 86 Figures

Springer

Dr. habil. Olaf Stenzel

Fraunhofer Institut Angewandte Optik und Feinmechanik
Winzerlaer Str. 10, 07745 Jena, Germany
E-mail: stenzel@iof.fhg.de

Library of Congress Control Number: 2005925965

ISSN 0931-5195
ISBN-10 3-540-23147-1 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-23147-9 Springer Berlin Heidelberg New York

Typesetting and prodcution: PTP-Berlin, Protago-TEX-Production GmbH, Berlin
Cover concept: eStudio Calamar Steinen
Cover production: *design & production* GmbH, Heidelberg

Printed on acid-free paper SPIN: 10918548 57/3141/YU 5 4 3 2 1 0

To Gabi

Preface

The present monograph represents itself as a tutorial to the field of optical properties of thin solid films. It is neither a handbook for the thin film practitioner, nor an introduction to interference coatings design, nor a review on the latest developments in the field. Instead, it is a textbook which shall bridge the gap between ground level knowledge on optics, electrodynamics, quantum mechanics, and solid state physics on one hand, and the more specialized level of knowledge presumed in typical thin film optical research papers on the other hand.

In writing this preface, I feel it makes sense to comment on three points, which all seem to me equally important. They arise from the following (mutually interconnected) three questions:

1. Who can benefit from reading this book?
2. What is the origin of the particular material selection in this book?
3. Who encouraged and supported me in writing this book?

Let me start with the first question, the intended readership of this book. It should be of use for anybody, who is involved into the analysis of optical spectra of a thin film sample, no matter whether the sample has been prepared for optical or other applications. Thin film spectroscopy may be relevant in semiconductor physics, solar cell development, physical chemistry, optoelectronics, and optical coatings development, to give just a few examples. The book supplies the reader with the necessary theoretical apparatus for understanding and modelling the features of the recorded transmission and reflection spectra.

Concerning the presumed level of knowledge one should have before reading this book, so the reader should have some idea on Maxwell's equations and boundary conditions, should know what a Hamiltonian is and for what it is good to solve Schrödinger's equation. Finally, basic knowledge on the band structure of crystalline solids is presumed. The book should thus be understandable to anybody who listened to basic courses in physics at any university.

The material selection was strongly influenced by the always individual experience on working with and supervising physics students as well as PhD-students. To a large extent, it stems from teaching activities at Chemnitz University of Technology, Institute of Physics, where I was involved in uni-

versity research on thin film properties, and gave several courses on applied spectroscopy topics as a lecturer. This university time stands for the more "academic" features of the book. It must be mentioned, that in that time I authored a textbook on thin film optics in German "Das Dünnschichtspektrum" with emphasis on the formal treatment of the optical response of thin solid films. But the present monograph is by no means a translation of that German book. The reason is, that in fall 2001, I changed to the Optical Coating Department at the Fraunhofer Institute of Applied Optics and Precision Engineering (IOF) in Jena, Germany. From that time, my working field shifted to more applied research projects on the development of optical coatings, primarily for the visible or near infrared spectral regions. It is the *combination* of university teaching until 2001 with more applied research work at the Fraunhofer Institute, which defines the individual content and style of the present monograph.

Finally, let me acknowledge the support of colleagues, co-workers, and friends in writing this book. First of all, I acknowledge Dr. Claus Ascheron and Dr. Norbert Kaiser for encouraging me to write it. Thanks are due to Dr. Norbert Kaiser for critical reading of several parts of the manuscript. The book could never have been written without the technical assistance of Ellen Kämpfer, who took the task of writing plenty of equations, formatting graphics and finally the whole text to make the manuscript publishable. Further technical support was supplied by Martin Bischoff.

Concerning the practical examples integrated into this book, e.g. the measured optical spectra of organic and inorganic thin solid films, it should be emphasized that all of them have been obtained in the course of research work at Chemnitz University (until summer 2001) and the Fraunhofer IOF (from fall 2001). Therefore, thanks are to the former members of the (unfortunately no more existing) research group on thin film spectroscopy (at Chemnitz University of Technology, Institute of Physics, Department of Optical Spectroscopy and Molecular Physics), and to the researchers in the Optical Coatings Department of the Fraunhofer IOF in Jena. The book much benefited from the stimulating research atmosphere in these facilities.

Jena, March 2005 *Olaf Stenzel*

Contents

Symbols and Abbreviations

A	absorptance
A_j, a_j	arbitrary expansion coefficients (in Chaps. 3 and 10)
$\boldsymbol{A}$	operator in quantum mechanics
A_{21}	Einstein's coefficient for spontaneous emission
a	sometimes used for geometrical dimensions (for example lattice constant, interatomic spacing, or others, as follows from the text)
a_0	Bohr's radius
α	absorption coefficient
B_j, b_j	arbitrary expansion coefficients (in Chap. 4)
$\boldsymbol{B}$	magnetic induction
B_{21}	Einstein's coefficient for stimulated emission
B_{12}	Einstein's coefficient for absorption
β	linear microscopic polarizability
β_h	linear polarizability of the host
$\beta^{(j)}$	polarizability of j-th order
C	constant
c	velocity of light in vacuum
$\boldsymbol{D}$	electric displacement
D	density /joint density of quantum states
d	physical (film) thickness
d_s	physical substrate thickness
δ	phase, phase shift
$\boldsymbol{E}$	electric field strength (vector)
E	electric field strength (scalar)
E_0	field amplitude
E_0	band gap (in Chap. 12)
E_g	direct band gap
E_n	energy level in quantum mechanics
$\boldsymbol{e}$	unit vector
e	basis of natural logarithm
ε_0	permittivity of free space

ε	dielectric function
ε'	real part of the dielectric function
ε''	imaginary part of the dielectric function
$\varepsilon_{\mathrm{xx}}$	diagonal element of the dielectric tensor
ε_h	dielectric function of the host
$\varepsilon_{\mathrm{stat}}$	static value of the dielectric function
ε_∞	'background' dielectric function (in Chap. 4)
F	error function
f_{j}	relative strength of the absorption lines
f_{ij}	oscillator strength in quantum mechanics
ϕ	incident angle
ϕ_{B}	Brewsters angle
γ	damping constant
Γ	homogeneous linewidth
$\boldsymbol{H}$	magnetic-field strength (vector)
H	magnetic-field strength (scalar)
$\boldsymbol{H}$	Hamilton operator, Hamiltonian
H	layer with high refractive index (in Chap. 9)
I	intensity
IR	infrared spectral region
i	counting index (in sums, in quantum mechanics)
i	imaginary unit
$\boldsymbol{j}$	electric current density
j	counting index (in sums, in quantum mechanics)
K	extinction coefficient
$\boldsymbol{k}$	wave vector
k	counting index (in sums, in quantum mechanics)
κ	response function
k_{B}	Boltzmann's constant
L	depolarisation factor (Chaps. 3 and 4)
L	optical loss (in Chap. 6)
L	layer with low refractive index (in Chap. 9)
l	counting index (in sums, in quantum mechanics)
l, L	sometimes used for geometrical dimensions
LO	linear optics
λ	wavelength
Λ	period of a diffraction grating

$\boldsymbol{M}$	magnetization
M	number
$\hat{\boldsymbol{M}}$	characteristic matrix
m_{ij}	matrix elements
m	mass
m	counting index (in sums, in quantum mechanics)
μ_0	permeability of free space
MIR	middle infrared spectral region
N	concentration
N	number
n	counting index (in sums, in quantum mechanics)
n, n_0	refractive index
n_s	substrate refractive index
$n_{(e,\mathrm{o})}$	extraordinary or ordinary refractive index in Chap. 6
n_ν	refractive index of the void material
$\hat{\mathrm{n}}$	complex index of refraction
NIR	near infrared spectral region
NLO	nonlinear optics
θ	step function
p	dipole moment
p_j	filling factor
$\boldsymbol{P}$	polarization (vector)
P	polarization (scalar)
$P^{(j)}$	polarization of j-th order
$\boldsymbol{p}_{ml}$	matrix element of the dipole operator
ψ	refraction angle
ψ	time-independent wavefunction in quantum mechanics
Ψ	time-dependent wavefunction in quantum mechanics
q	charge
R	radius
R	reflectance
R_p	reflectance of p-polarized light
R_s	reflectance of s-polarized light
$\boldsymbol{r}$	position vector with $\boldsymbol{r} = (x, y, z)^{\mathrm{T}}$
$r_{(s,p)}$	field reflection coefficient (for s- or p-polarized light)
ρ	mass density
$\boldsymbol{\rho}$	density matrix in a mixed quantum state
ρ_{nm}	elements of the density matrix

S	scatter
σ	conductivity
σ_{stat}	static value for the conductivity
$\boldsymbol{\sigma}$	density matrix of a pure quantum state
σ_{mn}	elements of the density matrix
T	absolute temperature
T	transmittance
t	field transmission coefficient
t	time
t_{coh}	coherence time
τ	time constant, relaxation time
u	spectral density
UV	ultraviolet spectral region
$\boldsymbol{V}$	perturbation operator
V	volume
V_j	fraction volume
V_{ij}	matrix element of the perturbation operator
v_{phase}	phase velocity
v_{z}	z-component of the velocity
ν	wavenumber
VIS	visible spectral region
VP	Cauchy's principal value of the integral
W	probability
w	relative weight function
w	Boltzmann's factor in Chap. 10
ω	angular frequency
ω_0	eigenfrequency, resonance frequency
ω_D	Doppler-shifted frequency
ω_p	plasma frequency
$\tilde{\omega}_0$	shifted with respect to local field effects resonance frequency
ω_{nm}	transition frequency, resonance frequency in quantum mechanics
$\Delta\omega$	spectral bandwidth
x	position
χ	linear dielectric susceptibility
χ_h	linear dielectric susceptibility of the host
χ_{stat}	static value of the susceptibility
χ_{res}	resonant contribution to the susceptibility
χ_{nr}	nonresonant contribution to the susceptibility
$\chi^{(j)}$	susceptibility of j-th order
Z	number of quantum states

1 Introduction

1.1 General Remarks

Whenever one is involved in spectroscopic experiments with electromagnetic waves, knowledge of the interaction of electromagnetic irradiation with matter is fundamental to the theoretical understanding of the experimental results. This is true, for example, in molecular as well as in solid state optical spectroscopy. The light-with-matter interaction is the basis of numerous analytical measurement methods, which are applied in physics as well as in chemistry and biology. There are a tremendous amount of scientific publications and textbooks which deal with this subject. So what is the reason for writing this new book?

The main reason is, that in the present monograph the subject is described from the viewpoint of the thin-film spectroscopist. Caused by the special geometry of a thin film sample, in thin film spectroscopy one needs a substantially modified mathematical description compared to the spectroscopy of other objects. The reason is, that a thin film has a thickness that is usually in the nanometer- or micrometer region, while it may be considered to extend to infinity in the other two (lateral) dimensions. Of course, there also exist monographs on thin film optics (and particularly on optical coatings design). It is nevertheless the experience of the author that there appears to be a discrepancy between the typical reader's knowledge on the subject and the scientific level that is presumed in the highly specialized scientific literature. Moreover, the interaction of light with matter is usually not taught as a separate university course. An interested student must therefore complete his knowledge by referring to different courses or textbooks, such as those on general optics, classical continuum electrodynamics, quantum mechanics and solid state physics.

It is therefore the authors aim to provide the reader with a short and compact treatment of the interaction of light with matter (particularly with thin solid films), and thus to bridge the gap between the readers basic knowledge on electrodynamics and quantum mechanics and the highly specialized literature on thin film optics and spectroscopy.

1.2 About the Content of the Book

In most practical cases, a thin film is built from a solid material. Therefore, the particular treatment in this book will mostly concern the specifics of the spectroscopy of solid matter. However, there appear situations where a general spectroscopic principle is easier to be explained referring to other states of matter. Inhomogeneous broadening of spectral lines is a typical example, as it is most easily explained in terms of the Doppler broadening as observed in gases. In such cases, we will happily leave the solid state specifics and turn to gases, in order to make the general principle more transparent.

Crystalline solids may be optically anisotropic. It is absolutely clear that a general and strong treatment of solid state spectroscopy must consider anisotropy. Nevertheless, in this book we will mostly restrict on optically isotropic materials. There are several reasons for this. First of all, many physical principles relevant in spectroscopy may be understood basing on the mathematically more simple treatment of isotropic materials. This is particularly true for many optical coatings, in fact, in optical coatings practice it is usually sufficient to work with isotropic layers models. There are exclusions from this rule, and in these situations anisotropy will be taken into account. This concerns, for example, the Giant Birefringent Optics (GBO) effects treated in connection with Fresnel's equations (Chap. 6). We will also refer to material anisotropy when discussing nonlinear optical effects at the end of this book (Chap. 13). By the way, the depolarization factors introduced in the first part of this book allow to a certain extent to calculate the anisotropy in optical material constants as caused by the materials morphology (Chaps. 3 and 4). However, this book does definitely not deal with *wave propagation* in anisotropic materials.

Having clarified these general points, let us turn to the overall structure of this book. First of all it should be clear, that the reader is presumed to have a certain knowledge on general optics, electrodynamics and quantum mechanics. It is not the purpose of this book to discuss the transversality of electromagnetic waves, nor to introduce the terms of linear or elliptical light polarization. The reader should be familiar with such kind of basic knowledge, as well as simple fundamentals of thermodynamics such as Boltzmann's and Maxwell's statistics.

Basing on this knowledge, the first part of the book (Chaps. 2–5) deals with the classical treatment of optical constants. In the classical treatment, both the electromagnetic field and the material systems will be described in terms of classical (non-quantum mechanical) models. Basing on Maxwell's equations, we will start with a rather formal introduction of optical constants and their frequency dependence (dispersion). We will have to introduce such important terms like the susceptibility, the polarizability, the dielectric function and the complex refractive index. We will then derive the main classical dispersion models (Debye-, Drude-, and the Lorentzian oscillator model). Starting from the Lorentz-Lorenz-formula, there will be a broad discussion of

the optical properties of material mixtures. The first part of this book will be finished by the derivation of the Kramers–Kronig-relations for the dielectric function.

The second part (Chaps. 6–9) describes wave propagation in thin film systems. We start from Fresnel's equations for transmission and reflection at a single interface. This is an utmost important matter in thin film optics. For that reason, the discussion of these equations will fill the full Chap. 6. In order to emphasize the physical value of these equations, we will derive a variety of optical and spectroscopic effects from them. Namely, this chapter will discuss Brewster's angle, total and attenuated total reflection of light, metallic reflection, propagating surface plasmon polaritons and the already mentioned GBO effects. In Chap. 7, the reader becomes familiar with the optical properties of thick slabs and single thin films. Chapter 8 deals with gradient index layers and film stacks, in particular, the matrix method for calculating transmittance and reflectance of an optical coating is introduced. In Chap. 9, some special cases are discussed, such as simple quarterwave stacks and the so-called grating waveguide structures.

The third part of the book (Chaps. 10–12) deals with the semiclassical treatment of optical constants. In this approach, the electromagnetic field is still described by Maxwell's equations, while the material system is described in terms of Schrödinger's equation. The goal is to obtain a semiclassical expression for the dielectric function, and consequently for the optical constants. Again, the reader is presumed to be familiar with basic knowledge on quantum mechanics and solid state physics, such as general properties of the wavefunction, simple models like the harmonic oscillator, perturbation theory, and Bloch waves. We start with the derivation of Einstein coefficients (Chap. 10). In this derivation, we become familiar with quantum mechanical selection rules and Planck's formula for blackbody irradiation. By the way, we get the necessary knowledge to understand how a laser works. In Chap. 11, a density matrix approach will be presented to derive a general semiclassical expression for the dielectric polarizability of a quantum system with discrete energy levels. In Chap. 12, the derived apparatus will be generalized to the description of the optical constants of solids.

Finally, Chap. 13 (which forms the very short fourth part of the book) will deal with simple effects of nonlinear optics.

1.3 The General Problem

The basic problem we have to regard is the interaction of electromagnetic irradiation (light) with a specific kind of matter (a thin film system). In order to keep the treatment compact and 'simple', we will restrict our discussion to the electric dipole interaction. We will assume throughout this book, that among all terms in the multipole expansion of the electromagnetic field, the

electric dipole contribution is the dominant one, and that other (higher order electric and all magnetic) terms may be neglected.

It is also worth emphasizing, that this book does definitely not deal with optical coatings design. It rather pursues the physical understanding of the information that may be drawn from a thin film spectrum as obtained from the experiment. We will therefore start from the experimental situation a thin film spectroscopist is confronted with.

In the frames of classical electrodynamics, any kind of light (which is used in optics) may be regarded as a superposition of electromagnetic waves. The idea of optical spectroscopy (or in more general optical characterization) is quite simple: If we have an object to be investigated (we will call it a *sample*), we have to bring it into interaction with electromagnetic waves (light). As the result of the interaction with the sample, certain properties of the light will be modified. The specific modification of the properties of electromagnetic waves resulting from the interaction with the sample shall give us information about the nature of the sample of interest.

For sufficiently low light intensities, the interaction process does not result in sample damage. Therefore, the majority of optical characterization techniques belongs to the non-destructive analytical tools in materials science. This is one of the advantages of optical methods.

Although the main idea of optical characterization is quite simple, it may be an involved task to turn it into practice. In fact one has to solve two problems. The first one is of entirely experimental nature: The modifications in the light properties (which represent our *signal*) must be detected. For standard tasks, this part of the problem may be solved with the help of commercially available equipment. The second part is more closely related with mathematics: From the signal (which may be simply a curve in a diagram) one has to conclude on concrete quantities characteristic for the sample. Despite of the researchers intuition, this part may include severe computational efforts. Thus, the solution of the full problem requires the researcher to be skilled in experiment and theory alike.

Let us now have a look at Fig. 1.1. Imagine the very simplest case – a monochromatic light wave impinging on a sample which is to be investigated. Due to the restriction on electric dipole interaction, we will only discuss the electric field of the light wave. It may be written according to:

$$\boldsymbol{E} = \boldsymbol{E}_0 \, \mathrm{e}^{-\mathrm{i}(\omega t - \boldsymbol{k}\boldsymbol{r})} \tag{1.1}$$

The parameters characterizing the incoming light (angular frequency ω, intensity (depends on the amplitude $\boldsymbol{E}_0$), polarization of the light (direction of $\boldsymbol{E}_0$), propagation direction (direction of $\boldsymbol{k}$) are supposed to be known. Imagine further, that as the result of the interaction with the sample, we are able to detect an electromagnetic wave with modified properties. Which properties of the electromagnetic wave may have changed as the result of the interaction with the sample?

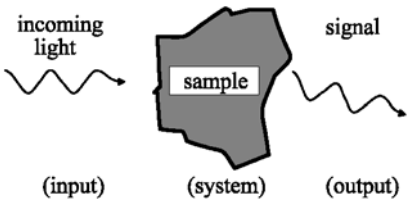

Fig. 1.1. Optical signal as the result of interaction of an electromagnetic wave with the sample

In principle, all of them may have changed. It is absolutely possible, that the interaction with the sample leads to changes in the frequency of the light. Typical examples are provided by Raman Scattering, or by several nonlinear optical processes. The polarization direction of the light may change as well. Ellipsometric techniques detect polarization changes and use them to judge the sample properties. Clearly, the light intensity may change (in most cases the light will be attenuated). This gives rise to numerous photometric methods analysing the sample properties basing on the measurement of intensity changes. And finally, anybody knows that the refraction of light may lead to changes in the propagation direction. Any refractometer makes use of this effect to determine the refractive index of a sample.

So we see, that the diversity of parameters characterizing electromagnetic radiation (in practice they are more than those mentioned here) may give rise to quite diverse optical characterization techniques.

We have now formulated our task: Starting from the analysis of certain parameters of the electromagnetic irradiation after having interacted with the sample, we want to obtain knowledge about the properties of the sample himself. Which kind of sample properties may be accessible to us?

Shortly spoken, the electromagnetic wave coming from the sample carries information about both the *sample material* and *sample geometry* (and the experimental geometry, but the latter is usually known to us). And if one is interested in the pure material properties, the geometrical influences on the signal have to be eliminated – experimentally or by calculations. In worse cases (and thin film spectra belong to these worse cases), geometrical and material informations are intermixed in the spectrum in a very complicated manner. In thin film systems, this is caused by the multiple internal reflections of light at the individual film interfaces. An experimental elimination of the geometrical sample contributions is then usually impossible, so that the derivation of material properties often becomes impossible without the instantaneous derivation of the geometrical properties by a corresponding mathematical treatment. As the result, we obtain information about both the sample material properties (for example the refractive index) and the geometry (for example the film thickness).

In order to make the theoretical treatment of thin film spectra more understandable, we will therefore develop the theory in two subsequent steps.

The first step deals with the description of pure material parameters, such as the refractive index, the absorption coefficient, the static dielectric constant and so on. We will present several models that describe these parameters in different physical systems.

The second step will be to solve Maxwell's equations in a system with given material parameters and a given geometry. In our particular case, we will do that for thin film systems. As the result, we obtain the electric field of the wave when it has leaved the system. Its properties will depend on the systems material *and* geometry. Having calculated the electric field, all the signal characteristics mentioned before may be theoretically derived. In the present book, the treatment will follow this philosophy.

In spectroscopy practice, one will proceed in a similar manner. The theoretical analysis of a measured spectrum starts from a hypothesis on the sample properties, including its material properties and geometry. Then, Maxwell's equations are solved, and the calculated characteristics are compared to the experimental values. From that, one may judge whether or not the assumptions previously made on the system were reasonable. If not, the assumed sample properties have to be altered, until a satisfying agreement between experiment and theory is achieved.

Having clarified the general features of our approach, let us now turn to the introduction of the linear optical susceptibility.

Part I

Classical Description of the Interaction of Light with Matter

2 The Linear Dielectric Susceptibility

2.1 Maxwell's Equations

Any optical phenomenon is connected with the interaction of electromagnetic radiation with matter. This interaction may be theoretically treated at different levels of difficulty. For example, one may use the purely classical description. It is on the other hand possible to build a strong quantum mechanical theory. In practice, a large number of practically important problems may be solved working with classical models only. We will therefore start our treatment with the classical description of the radiation-with-matter interaction.

A purely classical description utilizes Maxwell's Equations for the description of the electrical and magnetic fields and classical models (for example Newton's equations of motion) for the dynamics of the charge carriers present in any terrestrial matter. On the contrary, a quantum mechanical treatment is possible within the framework of the quantization of the electromagnetic field (so-called second quantization) and a quantum theoretical treatment of matter. This description is necessary, when spontaneous optical effects have to be described (spontaneous emission, spontaneous Raman scattering, or spontaneous paramagnetic interactions in nonlinear optics). In applied spectroscopy, the accurate quantum mechanical description is often omitted due to the rather complicated mathematics and replaced by the so-called semiclassical treatment. Here, the properties of matter are described in terms of quantum mechanical models, while the fields are treated within the framework of Maxwell's theory. Maxwell's equations are therefore used in both classical and semiclassical approaches, and for that reason we start our discussion from these equations, which are given below:

$$
\begin{aligned}
&1. \qquad div\ \boldsymbol{B} = 0, \\
&\qquad\quad \boldsymbol{B} = \mu_0\,(\boldsymbol{H} + \boldsymbol{M}), \\
&2. \qquad curl\ \boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial t}, \\
&3. \qquad div\ \boldsymbol{D} = 0, \\
&4. \qquad curl\ \boldsymbol{H} = \frac{\partial \boldsymbol{D}}{\partial t}, \\
&\qquad\quad \boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}.
\end{aligned}
\tag{2.1}
$$

Here, $\boldsymbol{E}$ and $\boldsymbol{H}$ represent the vectors of the electric and magnetic fields, while $\boldsymbol{D}$ and $\boldsymbol{B}$ stand for the electric displacement and the magnetic induction. $\boldsymbol{P}$ is the polarization, and $\boldsymbol{M}$ the magnetization. In (2.1), neither the free charge carrier density nor their current density are present. Keeping in mind, that optics deal with rapidly oscillating electric and magnetic fields, there is really no need to treat "free" charges separately – due to the short periods, they will only oscillate around their equilibrium position quite similar to bound charges. So in our description, the displacement vector contains information on both free and bound charges. The very few cases, where the static response of matter with free electrons becomes important in the frames of this book, cannot be treated within (2.1) and will need separate discussion.

In the following, we will assume that the media are generally non-magnetic ($\boldsymbol{M}$ is a zero-vector) and isotropic. Optically anisotropic materials will be treated in a special chapter later, but here we will assume isotropy for simplicity. Neglecting magnetism, from (2.1) one obtains straightforwardly:

$$curlcurl\boldsymbol{E} = graddiv\boldsymbol{E} - \Delta\boldsymbol{E} = -\mu_0 \frac{\partial^2 \boldsymbol{D}}{\partial t^2} \tag{2.2}$$

At this point, we need to establish a relationship between the vectors $\boldsymbol{E}$ and $\boldsymbol{D}$, which will be done in the next section.

2.2 The Dielectric Susceptibility

Let us assume, that a rapidly changing electric field with a completely arbitrary time-dependence interacts with a matter. One would naturally expect, that the electric field tends to displace, in general, both negative and positive charges, thus creating a macroscopic dipole moment of the system. The polarization $\boldsymbol{P}$ is per definition the dipole moment per unit volume, and it will be, of course, time-dependent in a manner that is determined by the time dependence of $\boldsymbol{E}$. For the moment, we neglect the spatial dependence of $\boldsymbol{E}$ and $\boldsymbol{P}$, because it is not essential for the further derivation. Generally, the polarization is thus a possibly very involved functional $\boldsymbol{F}$ of the field $\boldsymbol{E}$:

$$\boldsymbol{P}(t) = \boldsymbol{F}\big[\boldsymbol{E}(t' \leq t)\big] \tag{2.3}$$

Of course, the polarization of the medium is an action that is caused by the field (here and in the following, we do *not* regard ferroelectrics!). Due to the causality principle, the polarization at a given time t can depend on the field at the same moment as well as at previous moments t', but not on the field behaviour in the future. That is the meaning of the condition $t' \leq t$. We therefore postulate the following general relationship for the polarization as a functional of the electric field:

$$\boldsymbol{P}(t) = \varepsilon_0 \int_{-\infty}^{t} \kappa(t, t')\boldsymbol{E}(t')\mathrm{d}t' \tag{2.4}$$

Equation (2.4) postulates that the polarization at any time t may principally depend on the first power of the field at the current and all previous moments, as follows from the integration interval that is chosen in correspondence with the mentioned causality principle. The specific way, in which the system "remembers" the field strength at previous moments, is hidden in the response function $\kappa(t, t')$, which must be specific for any material. Equation (2.4) is in fact the first (linear) term of an expansion of (2.3) into a Taylor power series of $\boldsymbol{E}$. As we hold only the linear term of the series, all optical effects that arise from (2.4) form the field of *linear optics*.

In general, when the materials are anisotropic, $\kappa(t, t')$ is a tensor. As we restrict our attention here to optically isotropic materials, $\boldsymbol{P}$ will always be parallel to $\boldsymbol{E}$, so that $\kappa(t, t')$ becomes a scalar function.

A further facilitation is possible. Due to the homogeneity of time, $\kappa(t, t')$ will in fact not depend on both times t and t' separately, but only on their difference $\xi \equiv t - t'$. Substituting t' by ξ, we obtain:

$$\boldsymbol{P}(t) = \varepsilon_0 \int_0^\infty \kappa(\xi) \boldsymbol{E}(t - \xi) \mathrm{d}\xi \tag{2.5}$$

Let us now come to the utmost important case of a harmonic time dependence. Let us assume, that the electric field performs rapid oscillations according to:

$$\boldsymbol{E}(t) = \boldsymbol{E}_0 \, \mathrm{e}^{-\mathrm{i}\omega t}$$

and correspondingly

$$\boldsymbol{E}(t - \xi) = \boldsymbol{E}_0 \, \mathrm{e}^{-\mathrm{i}\omega t} \, \mathrm{e}^{\mathrm{i}\omega \xi}.$$

Note, that we assume a completely monochromatic field. It is then obtained:

$$\begin{aligned} \boldsymbol{P}(t) &= \boldsymbol{E}_0 \, \mathrm{e}^{-\mathrm{i}\omega t} \varepsilon_0 \int_0^\infty \kappa(\xi) \, \mathrm{e}^{\mathrm{i}\omega \xi} \, \mathrm{d}\xi \\ &= \boldsymbol{E}(t) \varepsilon_0 \int_0^\infty \kappa(\xi) \, \mathrm{e}^{\mathrm{i}\omega \xi} \, \mathrm{d}\xi \, . \end{aligned} \tag{2.6}$$

We define the *linear dielectric susceptibility* χ according to:

$$\chi = \int_0^\infty \kappa(\xi) \, \mathrm{e}^{\mathrm{i}\omega \xi} \, \mathrm{d}\xi = \chi(\omega) \, . \tag{2.7}$$

The thus defined susceptibility must be complex (it has both real and imaginary parts), and it depends on the frequency of the field even after performing

the integration in (2.7). Both circumstances arise mathematically from (2.5) and physically from finite inertness of any material system. Clearly, the charge carriers cannot instantaneously react on rapidly changing fields, so that their positions at a given time t depend on the history of the system, which is in fact the reason for the complicated temporal behaviour of the polarization. The information on the specific material properties is now carried by $\chi(\omega)$.

We are now able to formulate the relationship between $\boldsymbol{E}$ and $\boldsymbol{D}$ for monochromatic electric fields. Indeed, from (2.6) and (2.7) it follows, that

$$\boldsymbol{P} = \varepsilon_0 \chi \boldsymbol{E}\,. \tag{2.8}$$

In combination with the definition of $\boldsymbol{D}$ we have:

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} = \varepsilon_0 \big[1 + \chi(\omega)\big] \boldsymbol{E} \equiv \varepsilon_0 \varepsilon(\omega) \boldsymbol{E}\,, \tag{2.9}$$

where we defined the *dielectric function* $\varepsilon(\omega)$

$$\varepsilon(\omega) \equiv 1 + \chi(\omega)\,.$$

Equation (2.9) is completely analogous to what is known from the electrostatics of dielectrics, with the only difference that ε is complex and frequency dependent. So that we come to the conclusion, that in optics we have a similar relationship between field and displacement vectors as in electrostatics, with the difference that in optics the dielectric constant has to be replaced by the dielectric function.

2.3 Linear Optical Constants

We may now turn back to (2.2). Keeping in mind that our discussion is restricted to harmonic oscillations of the fields only, the second derivative with respect to time in (2.2) may be replaced by multiplying with $-\omega^2$. Replacing moreover $\boldsymbol{D}$ by (2.9), we obtain:

$$curl curl \boldsymbol{E} - \frac{\omega^2 \varepsilon(\omega)}{c^2} \boldsymbol{E} = 0\,. \tag{2.10}$$

Here we used the identity:

$$\varepsilon_0 \mu_0 = c^{-2}\,,$$

where c is the velocity of light in vacuum. For polychromatic fields, the single Fourier-components have to be treated separately in an analogous manner.

We now remember the vector identity:

$$curl curl \boldsymbol{E} \equiv grad div \boldsymbol{E} - \Delta \boldsymbol{E}.$$

In the case that $\varepsilon \neq 0$, from $div \boldsymbol{D} = 0$ it follows that $div \boldsymbol{E} = 0$. Thus we finally have:

$$\Delta E + \frac{\omega^2 \varepsilon(\omega)}{c^2} E = 0\,, \tag{2.11}$$

where the field vector has been replaced by a scalar field due to the assumed isotropy. A completely identical equation may be obtained for the magnetic field.

Let us remark at this point, that due to the assumed optical isotropy, we will often turn from the vectorial to the scalar mathematical description. Throughout this book, in these cases we will simply refrain from bold symbols without further notice.

Assuming that the dielectric function does not depend on the coordinates itself (homogeneous media), we are looking for a solution in the form:

$$E(\boldsymbol{r}, t) = E_0 \, \mathrm{e}^{-\mathrm{i}(\omega t - \boldsymbol{k}\boldsymbol{r})} \tag{2.12}$$

with $\boldsymbol{k}$ being the wavevector. Nontrivial solutions of (2.11) exist when

$$k = \pm \frac{\omega}{c} \sqrt{\varepsilon(\omega)} \tag{2.13}$$

is fulfilled. Assuming for simplicity, that $\boldsymbol{k}$ is parallel to the z-axis of a Cartesian coordinate system, (2.12) describes a planar wave travelling along the z-axis. It depends on the sign in (2.13) whether the wave is running into the positive or negative direction. We choose a wave running into the positive direction, and obtain:

$$E = E_0 \, \mathrm{e}^{-\mathrm{i}\left(\omega t - \frac{\omega}{c}\sqrt{\varepsilon(\omega)} z\right)} \tag{2.14}$$

where E_0 is the field amplitude at $z = 0$. Let us look on (2.14) in some more detail.

As we obtained in Sect. 2.2, the dielectric function may be complex, hence it may have an imaginary part. Of course, the square root will also be a complex function. We therefore have:

$$\sqrt{\varepsilon(\omega)} = \mathrm{Re}\sqrt{\varepsilon(\omega)} + \mathrm{i}\mathrm{Im}\sqrt{\varepsilon(\omega)}$$

Equation (2.14) therefore describes a damped wave according to:

$$E = E_0 \, \mathrm{e}^{-\frac{\omega}{c}\mathrm{Im}\sqrt{\varepsilon(\omega)} z} \, \mathrm{e}^{-\mathrm{i}\left(\omega t - \frac{\omega}{c}\mathrm{Re}\sqrt{\varepsilon(\omega)} z\right)} \tag{2.15}$$

with a z-dependent amplitude

$$E = E_0 \, \mathrm{e}^{-\frac{\omega}{c}\mathrm{Im}\sqrt{\varepsilon(\omega)} z} \tag{2.16}$$

and a phase:

$$\text{'phase'} = \omega t - \frac{\omega}{c} \mathrm{Re}\sqrt{\varepsilon(\omega)} z.$$

Let us calculate the velocity $\mathrm{d}z/\mathrm{d}t$ of any point at the surface of constant phase (which is a plane in our case). Regarding the phase as constant and differentiating the last equation with respect to time, we obtain the so called *phase velocity* of the wave according to:

$$\frac{\mathrm{d}z}{\mathrm{d}t} \equiv v_{\mathrm{phase}} = \frac{c}{\mathrm{Re}\sqrt{\varepsilon(\omega)}} \equiv \frac{c}{n(\omega)} \tag{2.17}$$

Here we introduced the *refractive index* $n(\omega)$ as the real part of the square root of the complex dielectric function. Naturally, the refractive index appears to be frequency dependent (so-called *dispersion* of the refractive index). In a medium with refractive index n, the phase velocity of an electromagnetic wave changes with respect to vacuum according to (2.17).

As a generalization to (2.17), one often defines the *complex index of refraction* as:

$$\hat{n}(\omega) = n(\omega) + \mathrm{i}K(\omega) \equiv \sqrt{\varepsilon(\omega)} \tag{2.18}$$

Its real part is identical with the ordinary refractive index as defined in (2.17), while its imaginary part (the so-called *extinction coefficient*) K is responsible for the damping of a wave. Indeed, returning to (2.16), we obtain for the amplitude of the wave:

$$E = E_0\,\mathrm{e}^{-\frac{\omega}{c}Kz}$$

Because the intensity I of the wave is proportional to the square of the field amplitude modulus, the intensity damps inside the medium as:

$$I = I(z=0)\,\mathrm{e}^{-2\frac{\omega}{c}Kz} \equiv I(z=0)\,\mathrm{e}^{-\alpha z}. \tag{2.19}$$

This exponential decay of light intensity for a wave travelling in a lossy medium is well known as Lambert's law of absorption with a frequency-dependent *absorption coefficient* α defined as:

$$\alpha(\omega) = 2\tfrac{\omega}{c}K(\omega) \tag{2.20a}$$

In terms of the identities:

$$\nu \equiv \frac{1}{\lambda} = \frac{\omega}{2\pi c}$$

where ν is the wavenumber and λ the wavelength in vacuum, we come to a more familiar expression:

$$\alpha(\nu) = 4\pi\nu K(\nu) \tag{2.20b}$$

Although the refractive index n and the extinction coefficient K are dimensionless, the absorption coefficient is given in reciprocal length units, usually in reciprocal centimetres. The reciprocal value of the absorption coefficient is sometimes called *penetration depth*. The pair of n and K forms the pair of *linear optical constants* of a material.

2.4 Some General Remarks

In practice, one often has to perform calculations of different spectra with the purpose to compare them with experimentally measured ones. One of the simplest tasks is the calculation of an absorption spectrum. Although we have not yet defined what may be meant by the term "absorption spectrum", it is intuitively clear that (at least in simple cases) such an absorption spectrum should resemble the wavelength dependence of the absorption coefficient of the material investigated. From the theoretical material described so far, we find however that the calculation of any absorption spectrum will contain at least two different parts: First of all, one has to find a suitable model for the dielectric function that contains the information about the *material*. After that, the optical constants may be calculated. Secondly, having this model in hands, one has to solve the wave equation (2.11) to account for the particular *geometry* valid for the (given or assumed) experiment. Having solved the wave equation with realistic boundary conditions, we obtain electric and/or magnetic fields that may be converted into light intensities, which in turn may be compared with experimental data. Changing the systems geometry will change the intensities obtained at the output, although the material might be the same. For example, in Sect. 2.3 we solved the wave equation, assuming however that the dielectric function is the same at any point. In other words, we assumed there a completely homogeneous medium, particularly without any interfaces. That resulted in Lambert's law (2.19), but the latter cannot be applied in other geometries, for example in thin film spectroscopy (although it is often done!). So that *both* material and geometry specifics must be considered in any spectra calculation.

There is a further complication in real live. What we have described so far is the philosophy of the forward search: We start from a model, calculate the optical constants, solve the wave equation and finally calculate the intensities. In practice, one is much more often confronted with reverse search tasks: The absorption (or any other) spectrum has been measured, and the optical constants have to be calculated. In several geometries (and particularly in thin film spectroscopy), the reverse search procedures are much more complicated than the forward search. The next section will exemplify a part of a *forward* search, namely the calculation of the dielectric function of a material consisting of microscopic dipoles.

2.5 Example: Orientation Polarization and Debye's Equations

Let us assume a material that is built from microscopic permanent electric dipoles. The dipoles are allowed to rotate freely with some damping. This is the typical situation in a liquid built from polar molecules (for example water). When no external electric field is applied, the statistical thermally

activated movement of the dipoles will not be able to create a macroscopic polarization. However, in an external electric field, the dipoles will more or less align with the field, creating a resulting macroscopic polarization. We shall find the frequency dependence of the dielectric function (and consequently of the optical constants) of such a material.

We will solve this task by means of (2.7). Because we still do not know the response function $\kappa(\xi)$, we start from the following thought experiment:

Let us assume, that a static electric field has been applied to the system for a sufficiently long time, so that a static polarization of the liquid has been well established. Let us further assume that the field is switched off at the moment $t = 0$. We model this situation by means of the electric field:

$$E(t) = E_0\big[1 - \theta(t)\big]$$

where $\theta(t)$ is a step function that has the value one for $t \geq 0$ and zero elsewhere. It makes no sense to assume that the polarization will vanish instantaneously with a vanishing external field. On the contrary, we shall assume, that due to the thermal movement of the particles, the macroscopic polarization decreases smoothly and asymptotically approaches the value of zero. This situation may be described by an exponentially descending behaviour with a time constant τ according to:

$$P(t) = P_0\,\mathrm{e}^{-\frac{t}{\tau}}\;; \quad t > 0$$

Furthermore, from (2.5) we have:

$$P(t) = P_0\,\mathrm{e}^{-\frac{t}{\tau}} = \varepsilon_0 \int_0^\infty \kappa(\xi) E_0\big[1 - \theta(t-\xi)\big]\,\mathrm{d}\xi$$

The only action of the step function is to reduce the integration interval:

$$P_0\,\mathrm{e}^{-\frac{t}{\tau}} = -\varepsilon_0 E_0 \int_\infty^t \kappa(\xi)\,\mathrm{d}\xi$$

We differentiate with respect to time and make use of the identity:

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left[\int_a^x f(\xi)\,\mathrm{d}\xi\right]$$

That leads us to the following expression for the response function $\kappa(t)$:

$$\kappa(t) = \frac{P_0}{\varepsilon_0 E_0 \tau}\,\mathrm{e}^{-\frac{t}{\tau}} \equiv \kappa_0\,\mathrm{e}^{-\frac{t}{\tau}} \tag{2.21}$$

Having found the response function, the further treatment is straightforward. Equations (2.7) and (2.9) yield the dielectric function:

$$\varepsilon(\omega) = 1 + \chi(\omega) = 1 + \int_0^\infty \kappa(\xi)\, \mathrm{e}^{\mathrm{i}\omega\xi}\, \mathrm{d}\xi$$

$$= 1 + \kappa_0 \int_0^\infty \mathrm{e}^{(\mathrm{i}\omega - \frac{1}{\tau})\xi} \mathrm{d}\xi = 1 + \frac{\kappa_0 \tau}{1 - \mathrm{i}\omega\tau}$$

or

$$\varepsilon(\omega) = 1 + \frac{\chi_{\mathrm{stat}}}{1 - \mathrm{i}\omega\tau} \tag{2.22}$$

where χ_{stat} is the static ($\omega = 0$) value of the susceptibility. The real and imaginary parts of the dielectric function may be written as follows:

$$\begin{aligned} \mathrm{Re}\varepsilon &= 1 + \frac{\chi_{\mathrm{stat}}}{1 + \omega^2\tau^2} \\ \mathrm{Im}\varepsilon &= \frac{\chi_{\mathrm{stat}}\omega\tau}{1 + \omega^2\tau^2} \end{aligned} \tag{2.23}$$

The thus obtained dielectric function represents a simplified version of Debye's equations valid for the dielectric function in polar media. In Fig. 2.1, the spectral shapes of real and imaginary parts of this particular dielectric function are presented. Figure 2.2 shows the corresponding optical constants. In these figures, a static susceptibility of $\chi_{\mathrm{stat}} = 80$ has been assumed, similar to what is valid in ordinary water. Obviously, the presence of permanent dipoles in the medium is connected with a high static dielectric constant, while for higher frequencies, the real part of the dielectric function may be essentially lower. Thus, in the visible spectral range, water has a dielectric function with a real part of approximately 1.77 and a refractive index of 1.33. This behaviour is consistent with the predictions from Debye's equations, where the refractive index is expected to steadily decrease with increasing frequency.

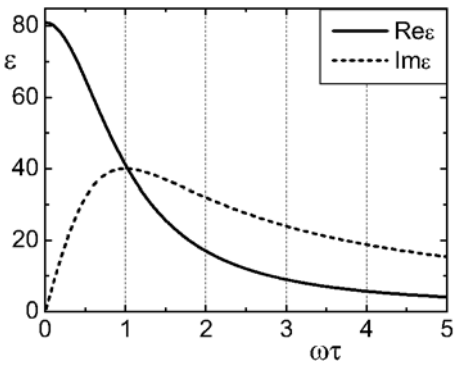

Fig. 2.1. Real and imaginary parts of the dielectric function according to (2.23)

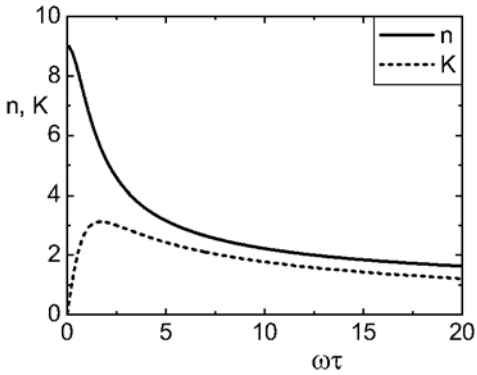

Fig. 2.2. Optical constants n and K for the dielectric function presented in Fig. 2.1, but in a broader spectral region

A more interesting fact is seen from Fig. 2.1. The imaginary part of the dielectric function has its maximum value exactly at the angular frequency $\omega = \tau^{-1}$. Consequently, the result of a *spectral* measurement (determining the peak position of $\mathrm{Im}\varepsilon$) reveals information about the *dynamic* behaviour of the system (the decay time of polarization). This is one example for the validity of a more general fundamental principle, that in optics the spectral ($\chi(\omega)$) and temporal ($\kappa(t)$) representations embody the same information and may be transferred into each other. Indeed, (2.7) is in fact a Fourier transformation of the response function, performed however only over a semi-infinite interval for reasons of causality. One may formally multiply the response function with a step-function:

$$\tilde{\kappa}(\xi) = \kappa(\xi)\theta(\xi) \tag{2.24}$$

The thus obtained modified (truncated) response function may be integrated over the full time interval, so that we have:

$$\chi(\omega) = \int_{-\infty}^{\infty} \tilde{\kappa}(\xi)\, \mathrm{e}^{\mathrm{i}\omega\xi} \mathrm{d}\xi \tag{2.25}$$

In (2.25), the susceptibility appears to be the Fourier-transform of the truncated response function.

Let us make two final remarks concerning the conclusions from Chap. 2:

We supposed the time dependence of the fields according to $\mathrm{e}^{-\mathrm{i}\omega t}$. As a consequence, we defined the complex index of refraction as $n + \mathrm{i}K$. The same kind of theory may be built postulating a time dependence of the fields as $\mathrm{e}^{+\mathrm{i}\omega t}$. However, in this case the index of refraction will be $n - \mathrm{i}K$. Both approaches are equally correct, however, they shouldn't be confused with each other.

A high extinction coefficient (high damping) is not necessarily connected with a high $\mathrm{Im}\varepsilon$. For example, a real but negative dielectric function will result in a purely imaginary refractive index. This seemingly exotic assumption is in fact important in metal optics and will be treated in the section about total internal reflection. Here the penetrating wave is indeed damped, but the light is rather reflected than absorbed. Therefore, the generally accepted terminus "absorption coefficient" may be misleading in special cases. In fact, for light absorption it is essential that $\mathrm{Im}\varepsilon \neq 0$. We will return to these questions later in more detail.

3 The Classical Treatment of Free and Bound Charge Carriers

3.1 Free Charge Carriers

3.1.1 Derivation of Drude's Formula

In this section, we come to the discussion of an important problem in solid state optics, namely the optical response of the free charge carrier fraction (in many cases electrons) in condensed matter. This is of utmost significance in metal optics, but of course, the optical properties of highly doped semiconductors may be influenced by free charge carriers as well.

Let us start with a more general statement. In Sect. 2.5, we derived equations that describe the optical response of permanent dipoles. In this chapter, we consider free electrons. The next step will be to discuss the contribution of bound electrons. As the result, we will have at least three models in hand each being tailored for a very special case. But real matter is more complicated. Thus, for example, metals have free *and* bound electrons. Analogously, the optical properties of water are not only determined by the permanent dipole moment of the water molecules. The relative movements of bound electrons is important as well, and once water is a conductor for electrical current, it must have a certain concentration of free charge carriers. Intramolecular vibrations of the cores will also add their contribution.

Fortunately, as charges are additive, all the degrees of freedom present in real matter will contribute their dipole moments to the final polarization that is obtained as a sum over all dipole moments in the medium. Consequently, the susceptibilities that correspond to different degrees of freedom (numbered by j) add up to the full susceptibility, so that the dielectric function will be:

$$\varepsilon(\omega) = 1 + \sum_j \chi_j(\omega)\,, \tag{3.1}$$

where the χ_j are the susceptibilities obtained for the corresponding group of dipoles.

After this remark, let us turn to the discussion of the role of free electrons in optics. The simplest derivation of the susceptibility of free electrons moving around positive atomic cores is based on Newton's equation of motion. As the cores are much heavier than the electrons, the cores will be considered

as fixed, so that only electrons are in motion when a harmonic electric field is applied.

Assuming, that the motion of electrons is confined to a region much smaller than the wavelength, we may write for the movement of a single electron:

$$qE = qE_0\,\mathrm{e}^{-\mathrm{i}\omega t} = m\ddot{x} + 2\gamma m\dot{x} \tag{3.2}$$

m and q are the mass and charge of the electron, and γ is a damping constant necessary to consider the damping of the electrons movement. We assume, that the electric field is polarized along the x-axis, hence we consider only movements of the electron along the x-axis. For non-relativistic velocities, the Lorentz-force may be neglected compared to the Coulomb-force, so that only the latter is apparent in (3.2).

Assuming $x(t) = x_o\,\mathrm{e}^{-\mathrm{i}\omega t}$, we obtain from (3.2):

$$\frac{qE}{m} = -\omega^2 x - 2\mathrm{i}\gamma\omega x\,.$$

The oscillation of the electron around its equilibrium position thus induces an oscillating dipole moment according to:

$$p = qx = -\frac{q^2E}{m}\,\frac{1}{\omega^2 + 2\mathrm{i}\gamma\omega}.$$

If N is the number of electrons per unit volume (we will call it the *concentration* of electrons), then the polarization P is given by

$$P = Np = -\frac{q^2NE}{m}\,\frac{1}{\omega^2 + 2\mathrm{i}\gamma\omega}$$

so that, according to (2.8), the susceptibility is:

$$\chi(\omega) = -\frac{Nq^2}{\varepsilon_0 m}\,\frac{1}{\omega^2 + 2\mathrm{i}\gamma\omega} \tag{3.3}$$

where the term $\dfrac{Nq^2}{\varepsilon_0 m}$ represents the square of the plasma frequency defined as:

$$\omega_p = \sqrt{\frac{Nq^2}{\varepsilon_0 m}} \tag{3.4}$$

As in Sect. 2.5, we get a complex and frequency-dependent susceptibility. The dielectric function is then given by:

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + 2\mathrm{i}\gamma\omega} \tag{3.5}$$

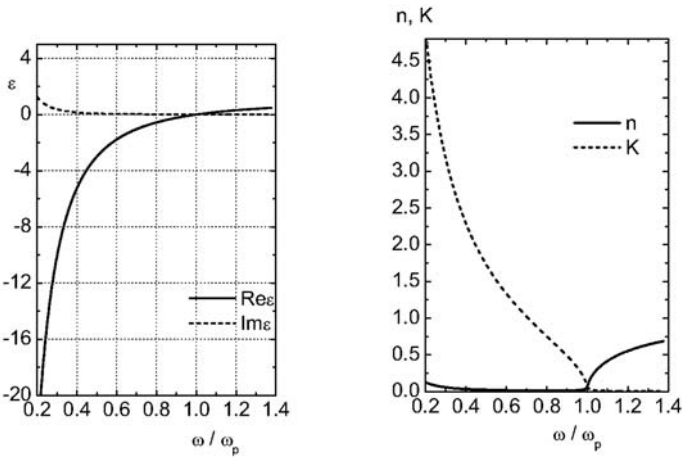

Fig. 3.1. Dielectric function and optical constants according to (3.5)

Figure 3.1 displays the principal shape of the real and imaginary parts of the dielectric function from (3.5), as well as the optical constants. The most striking feature appears in the refractive index, which is expected to be less than one in broad spectral regions. In fact, the imaginary part of the complex refractive index may be much larger than the real one. This is typical for metals, and as it will be seen in Chap. 6, it causes the well-known metallic brightness. Due to $n < 1$, the phase velocity of light in metals may be higher than in vacuum. This does not conflict with relativity, because light signals (for example wave packets) travel in space with the group velocity, and not with the phase velocity.

A simple discussion of (3.5) (which is sometimes called Drude's function) confirms the following asymptotic behaviour:

$$\omega \to \infty : \quad \mathrm{Re}\varepsilon \to 1; \quad \mathrm{Im}\varepsilon \to 0; \quad n \to 1; K \to 0 \tag{3.6}$$

Note that this is exactly the same behaviour as it would follow from Debye's function (equation (2.23)). The reason is simple: Due to the finite inertness of the electrons, they will not be able to comply with field oscillations that are too rapid. Hence, for $\omega \to \infty$, the electrons will not interact with the field, so that the field does not "feel" the electrons. For that reason, the optical constants of the system approach those of vacuum ($n = 1$; $K = 0$). The permanent dipoles, which are responsible for the dispersion described by Debye's formula (2.23), are much heavier than electrons and therefore even more inert. For high frequencies, they will give no optical response as well.

The static case is more difficult to handle. Drude's function (3.5) yields the following behaviour:

$$\omega \to 0 : \quad \mathrm{Re}\varepsilon \to 1 - \frac{\omega_p^2}{4\gamma^2}; \quad \mathrm{Im}\varepsilon \to \frac{\omega_p^2}{2\omega\gamma}; \quad n \approx K \to \frac{\omega_p}{2\sqrt{\omega\gamma}} \tag{3.7}$$

In the static case, only the real part of the dielectric function approaches a finite value, the other functions become infinitively large. This is intuitively clear, because in a static electric field, the free electrons do not oscillate, but move away from the cores, causing a finite electrical current, but infinitively large dipole moments.

3.1.2 Extendted Detail: Another Evaluation of Drude's Formula

As we have mentioned in Sect. 2.1, in optics it makes no sense to separate free and bound electrons in Maxwell's equations, because both types of electrons perform oscillations around their equilibrium positions. At the same time, we remarked that the static case ($\omega = 0$) cannot be treated this way. To comply with this particular situation, it is more convenient to discuss the current density $\boldsymbol{j}$ than the induced dipole moments.

The definition of the polarization vector (*induced* polarization only) may be written as:

$$\boldsymbol{P} = \frac{1}{V} \sum_l q_l (\boldsymbol{r}_l - \boldsymbol{r}_{0l})$$

where V is the volume, and l counts all charge carriers contained in the volume. $\boldsymbol{r}_o$ is the equilibrium position of a charge carrier, and $\boldsymbol{r}$ its actual position. Differentiating with respect to time leads to:

$$\frac{\partial \boldsymbol{P}}{\partial t} = \frac{1}{V} \sum_l q_l \dot{\boldsymbol{r}}_l = \boldsymbol{j} \tag{3.8}$$

so that $\boldsymbol{j} = \partial \boldsymbol{P}/\partial t$. Comparing equations (2.6) and (3.8), we conclude that for harmonic fields, the relation between $\boldsymbol{j}$ and $\boldsymbol{E}$ must have the same structure as between $\boldsymbol{P}$ and $\boldsymbol{E}$. We therefore write in full analogy to statics:

$$j = \sigma E \tag{3.9}$$

σ is the conductivity. For a static field, that has been switched on for a long time ago, one would expect a constant current density in the medium. After switching off the field at the moment $t = 0$, the current will not instantaneously drop to zero, because of the inertness of the charge carriers. Instead, the current density is expected to decay according to:

$$j = j_0 \, \mathrm{e}^{-\frac{t}{\tau}}$$

This situation is completely analogous to that discussed in Sect. 2.5, with the only difference that we deal with current densities here and not with dipole moments. We will therefore get an expression for the frequency-dependent conductivity as:

$$\sigma(\omega) = \frac{\sigma_{\mathrm{stat}}}{1 - \mathrm{i}\omega\tau}$$

analogous to (2.22). σ_{stat} is the familiar static value for the conductivity.

This offers the possibility to derive Drude's formula starting from the conductivity. For harmonic fields, the derivative with respect to time may be calculated according to the recipe:

$$\frac{\partial}{\partial t} \to *(-\mathrm{i}\omega)$$

From (3.8) we get for non-zero frequencies:

$$P = \int j \mathrm{d}t = \frac{j}{-\mathrm{i}\omega} = \frac{\sigma E}{-\mathrm{i}\omega} = -\frac{\sigma_{\mathrm{stat}}}{\omega^2\tau + \mathrm{i}\omega} E$$

and thus for the susceptibility:

$$\chi(\omega) = -\frac{\sigma_{\mathrm{stat}}/\varepsilon_0}{\omega^2\tau + \mathrm{i}\omega} \tag{3.10}$$

Comparing (3.3) and (3.10) yields the following relationships:

$$\frac{Nq^2}{\varepsilon_0 m} = \omega_p^2 = \frac{\sigma_{\mathrm{stat}}}{\varepsilon_0 \tau} \tag{3.11}$$

$$2\gamma = \tau^{-1} \tag{3.12}$$

Hence, electrical and optical properties of a "classical" metal are directly related to each other. For typical metals, ω_p is of the order $10^{15}\mathrm{s}^{-1}$, and $\tau \sim 10^{-13}\mathrm{s}$.

We thus found another version of Drude's formula, derived in a similar way as we have derived Debye's equations in Chap. 2. Nevertheless, there remains the question: Why didn't we use (2.7) or (2.25) directly in order to obtain Drude's formula?

The answer is given by these equations themselves. Evaluating the exponential function in (2.25), we get:

$$\begin{aligned} \chi(\omega) &= \int_{-\infty}^{\infty} \tilde{\kappa}(\xi)\, \mathrm{e}^{\mathrm{i}\omega\xi} \mathrm{d}\xi \\ &= \int_{-\infty}^{\infty} \tilde{\kappa}(\xi)\, \mathrm{d}\xi + \mathrm{i}\omega \int_{-\infty}^{\infty} \tilde{\kappa}(\xi)\xi \mathrm{d}\xi + \left(-\frac{\omega^2}{2}\right) \int_{-\infty}^{\infty} \tilde{\kappa}(\xi)\xi^2 \mathrm{d}\xi + \ldots \end{aligned} \tag{3.13}$$

That means, our approach for the optical (high frequency) susceptibility corresponds to an infinite series according to:

$$\chi(\omega) = a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3 + \ldots \tag{3.14}$$

where the even orders in ω correspond to the real part, while the odd orders determine the imaginary part of the susceptibility or the dielectric function.

The a_j-values are constants. For $\omega \to 0$, one has $\chi \to a_0$. Therefore, as we see from (3.7), Drude's function cannot be described this way. For $\omega \to 0$ it behaves like:

$$\chi^{\text{Drude}}(\omega)\,|_{\omega\to 0} \approx \mathrm{i}\,\frac{\sigma_{\text{stat}}}{\varepsilon_0\omega}$$

or

$$\chi^{\text{Drude}}(\omega)\,|_{\omega\to 0} \approx \mathrm{i}\,\frac{Nq^2}{\varepsilon_0 m}\,\frac{1}{2\gamma\omega}$$

so that for a conductor, one would expect:

$$\chi^{\text{conductor}}(\omega) = \mathrm{i}\,\frac{\sigma_{\text{stat}}}{\varepsilon_0\omega} + a_0 + a_1\omega + a_2\omega^2 + a_3\omega^3 + \ldots \tag{3.15}$$

This is generally incompatible with (2.25), but for sufficiently high (optical) frequencies, the first term in (3.15) has no significance, so that (2.25) or (2.7) remain valid. It is seen from (3.7), that already for $\omega > 2\gamma$, a_0 becomes larger by modulus than the first term in (3.15).

The rather formal discussion performed in this section might seem not relevant for applied spectroscopy practice. However, (3.15) will become important when the Kramers–Kronig-relationships will be evaluated (this will be done in Chap. 5), so that we will have to return to this question anyway.

3.2 The Oscillator Model for Bound Charge Carriers

3.2.1 General Idea

Even in metals, most of the electrons are bound, although the free electrons are utmost important for the specific optical behaviour of metals. As everybody knows, metals like silver, gold and copper have quite a different optical appearance, and this is a consequence of the response of the *bound* electron fraction. Of course, the optical properties of dielectrics are exclusively determined by the motion of bound charge carriers.

There is a more general question concerning the different role of negative electrons and positively charged cores. Generally, both electrons and cores may perform movements when being excited by external electric fields. But the cores are much heavier. In terms of classical physics, the vibrational eigenfrequencies of a system are determined by the restoring forces and the masses of the systems constituents. Assuming a typical core being 10^4 times heavier than an electron, one would expect the eigenfrequencies of the core motion approximately 100 times lower than that of electrons that are equally tight bound (in terms of quantum mechanics, these are the valence electrons). Therefore, at high frequencies, the movement of the cores may be neglected. At lower frequencies (and this is usually the infrared spectral region), the movements of the cores determine the optical properties of the material.

On the other hand, not all electrons are equally tight bound. Although this is again rather a quantum mechanical matter, we may formally assume, that there are groups of electrons (*core* electrons) that suffer much higher restoring forces than the other (the *valence*) electrons. Consequently, there are different groups of electrons with different eigenfrequencies.

The oscillator model derived in the following is very general. It may be applied to the intramolecular motion of cores (in infrared spectroscopy) as well as to bound electrons. So that we will simply speak in the following on induced dipole moments, no matter which is their physical origin.

So let us regard the motion of a charge carrier, which is bound to its equilibrium position ($x = 0$) by an elastic restoring force. An oscillating field may lead to small ($x \ll \lambda$) movements of the charge carriers, thus inducing dipoles that interact with the field. In contrast to (3.2), the equation of motion of a single charge carrier is now:

$$qE = qE_0\,\mathrm{e}^{-\mathrm{i}\omega t} = m\ddot{x} + 2\gamma m\dot{x} + m\omega_0^2 x \tag{3.16}$$

This is the equation for forced oscillations of a damped harmonic oscillator with the eigenfrequency ω_0, all other symbols have the same meaning as before. Proceeding exactly in the same way as in Sect. 3.1, we obtain for a single induced dipole moment $p = qx$:

$$p = \frac{q^2 E}{m}\,\frac{1}{\omega_0^2 - \omega^2 - 2\mathrm{i}\omega\gamma}$$

Thus, the electric field induces plenty of microscopic dipoles, which form a macroscopic polarization of the medium. Let us define the linear microscopic polarizability β via:

$$p = \varepsilon_0 \beta E \tag{3.17}$$

Then the polarizability turns out to be complex and frequency-dependent according to:

$$\beta = \frac{q^2}{\varepsilon_0 m}\,\frac{1}{\omega_0^2 - \omega^2 - 2\mathrm{i}\omega\gamma} \tag{3.18}$$

Equation (3.18) describes a resonant behaviour of the microscopic dipole, when the angular frequency ω of the field approaches the angular eigenfrequency of the dipole. In this resonance condition, the interaction between radiation and matter is expected to be most effective.

Note that the linear polarizability has the dimension of the volume. The model that was described here is sometimes called the Lorentzian oscillator model.

3.2.2 Microscopic Fields

From (3.17) and (3.18), it seems straightforward to calculate the macroscopic polarization vector $\boldsymbol{P}$ from the induced dipole moments. After that, we may

find the susceptibilities. But there is a further problem in condensed matter optics, which concerns the electric field fixed in (3.17)

The problem is as follows: (2.5) describes the macroscopic response of the medium, and the electric field fixed in (2.5) is the average field in the medium. It is formed from the external field and the field of the dipoles in the medium. On the contrary, (3.17) describes a microscopic dipole moment, and the field is the microscopic (or local) field acting on the selected dipole. The question is, whether or not these fields are identical.

In the general case, these fields are different, and the aim of this section is to derive an equation that allows us to calculate the microscopic field for the special case of optically isotropic materials.

Let us regard a single induced dipole in the medium. The field acting on the dipole is built from two constituents: the external field and the field caused by all other dipoles, except the considered one. Of course, nobody would start the calculation from the external field, subsequently adding the response of 10^{23} dipoles. Instead, we make use of the superposition principle. We will subtract the field of our regarded dipole from the average field in the medium, and this way we can find the field that is acting on the dipole itself.

In continuum electrodynamics, this calculation is easy to perform, regarding the dipole as a sphere (in accordance with the assumed isotropy) with a diameter much smaller than the wavelength, so that the average field $\boldsymbol{E}$ may be assumed to be spatially homogeneous. This is the so-called quasistatic case, where the field is oscillating with time, but homogeneous with respect to the dimensions of the discussed dipole. The latter could be, for example, an elementary cell in a cubic crystal, a molecule with a rather spherical shape, or simply an atom.

In Fig. 3.2, this situation is presented. On the left, we have the continuous medium with a small microscopic spherical hole inside. The field inside this hole corresponds to the mentioned microscopic field, because it may be con-

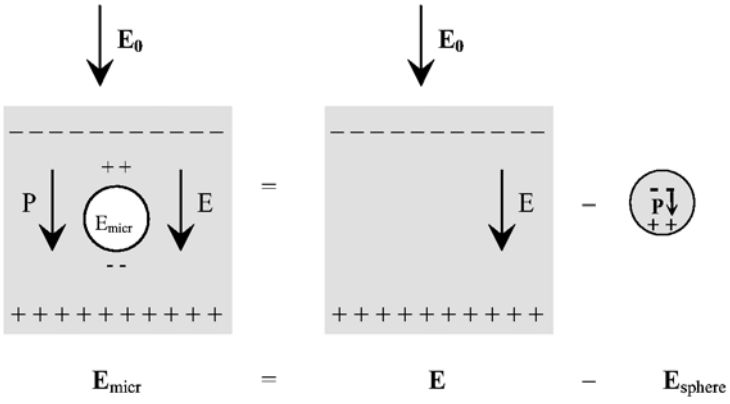

Fig. 3.2. Calculation of the microscopic field

sidered as the field in the compact medium less the field in a homogeneous polarized sphere.

We therefore have:

$$E_{\text{micr}} = E - E_{\text{sphere}} \tag{3.19}$$

As the field in the polarized sphere is equal to $-P/3\varepsilon_0$, we obtain from (3.19)

$$E_{\text{micr}} = E + \frac{P}{3\varepsilon_0} \tag{3.20}$$

Finally, we find for the macroscopic polarization $\boldsymbol{P}$:

$$P = Np = N\varepsilon_0\beta E_{\text{micr}} = N\varepsilon_0\beta E + \frac{N\varepsilon_0\beta}{3\varepsilon_0} * P$$

$$P = \frac{N\beta\varepsilon_0 E}{1 - \dfrac{N\beta}{3}}$$

where N is again the concentration of the dipoles. The susceptibility is:

$$\begin{aligned} \chi &= \frac{N\beta}{1 - \dfrac{N\beta}{3}} \\ \beta &= \frac{q^2}{\varepsilon_0 m} \; \frac{1}{\omega_0^2 - \omega^2 - 2\mathrm{i}\omega\gamma} \end{aligned} \tag{3.21}$$

where β is given by (3.18).

For small concentrations ($N \to 0$) the susceptibility equals $N\beta$, which is valid for diluted gases.

Before coming to the discussion of (3.21), let us make one remark. Because of $\boldsymbol{P} = \varepsilon_0\chi\boldsymbol{E} = \varepsilon_0(\varepsilon - 1)\boldsymbol{E}$, from (3.20) it follows immediately that

$$E_{\text{micr}} = \frac{\varepsilon + 2}{3} E \tag{3.22a}$$

This is valid for the assumed spherical cavity in the continuum. For $\varepsilon > 1$, the microscopic field exceeds the average field due to the surface charges at the cavity borders, as indicated in Fig. 3.2. In fact, our treatment also allows to account for simple cases of optical anisotropy. In this case, the spherical cavity shape must be replaced by another suitable cavity shape, which leads to modifications in (3.22a). Thus, for a thin needle-like cavity parallel to $\boldsymbol{E}$, the surface charges at the bottom and the top of the cavity are of negligible influence, so that one has

$$E_{\text{micr}} = E \tag{3.22b}$$

On the contrary, in a pancake-shaped cavity perpendicular to $\boldsymbol{E}$, the surface charges in the cavity would completely compensate those at the outer boundary of the dielectric, so that the microscopic field inside the cavity equals

Table 3.1. Depolarisation factors L. For completeness, the general expression for calculating L for an ellipsoid along the three main axes l_a, l_b, and l_c is included (without derivation)

Type of cavity	$\boldsymbol{E}$ parallel to the cavity axis	$\boldsymbol{E}$ perpendicular to the cavity axis
Ellipsoid with main axes l_a, l_b, l_c	$L_\xi = \frac{l_a l_b l_c}{2} \int_0^\infty \frac{\mathrm{d}s}{\left(s + l_\xi^2\right) \sqrt{\left(s + l_a^2\right)\left(s + l_b^2\right)\left(s + l_c^2\right)}}\,; \quad \xi = a, b, c$	
Sphere	$^1/_3$	
Needle	0	$^1/_3$
Pancake	1	0

the external electric field that would be measured outside the dielectric. We therefore have for a pancake cavity:

$$E_{\mathrm{micr}} = \varepsilon E \tag{3.22c}$$

These equations may be written in a generalized form according to:

$$E_{\mathrm{micr}} = [1 + (\varepsilon - 1)L]\, E \tag{3.22d}$$

where L is the so-called depolarisation factor. For important cases, the depolarisation factors are summarized in Table 3.1.

3.2.3 The Clausius–Mossotti and Lorentz–Lorenz-Equations

From (3.21), we find the dielectric function of a medium with respect to local field effects according to:

$$\varepsilon = 1 + \frac{N\beta}{1 - \dfrac{N\beta}{3}} \tag{3.23}$$

This gives us the Clausius–Mossotti-Equation:

$$\frac{\varepsilon - 1}{\varepsilon + 2} = \frac{N\beta}{3} \tag{3.24}$$

or the Lorentz–Lorenz-Equation

$$\begin{aligned} \frac{\hat{n}^2 - 1}{\hat{n}^2 + 2} &= \frac{N\beta}{3} \\ \hat{n} &= n + \mathrm{i}K \end{aligned} \tag{3.25}$$

The significance of these rather simple equations is in that they relate microscopic optical parameters (the polarizability β) to macroscopically measurable parameters (optical constants). In other words, measurements on the macroscopic scale, which yield the optical constants of a material, give further access to microscopic parameters such as molecular or atomic polarizabilities. In fact, this is the point from where analytical optical spectroscopy starts.

Let us have a look at the consequences. We have a microscopic polarizability according to (3.18) and a dielectric function from (3.23). In combination, that yields:

$$\varepsilon(\omega) = 1 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - 2\mathrm{i}\gamma\omega - \dfrac{\omega_p^2}{3}} \equiv 1 + \frac{\omega_p^2}{\tilde{\omega}_0^2 - \omega^2 - 2\mathrm{i}\gamma\omega} \tag{3.26}$$

where

$$\tilde{\omega}_0^2 \equiv \omega_0^2 - \frac{\omega_p^2}{3} \tag{3.27}$$

is the resonance frequency valid for the dielectric function. The dielectric function has exactly the same spectral shape as the polarizability, but the resonance position in ε is red-shifted with respect to that of the polarizability. The larger the density, the larger is the red-shift. For arbitrary depolarisation factors L, (3.27) generalizes to:

$$\tilde{\omega}_0^2 \equiv \omega_0^2 - L\omega_p^2$$

Figure 3.3 shows the real and imaginary parts of a dielectric function according to (3.26), and Fig. 3.4 the optical constants. We see, that in the vicinity of the resonance frequency, the imaginary parts of both the dielectric function and the index of refraction show a local maximum. That means, that at this frequency the light wave is effectively damped. The imaginary part of the dielectric function therefore describes an absorption line with a characteristic shape, which is called a Lorentzian line. In the region of strong damping, the refractive index n decreases with increasing frequency (anomalous dispersion). On the contrary, in the transparency regions, where damping is negligible, n increases with increasing frequency (normal dispersion). For high frequencies, the relationships (3.6) are again valid.

We will now have to make a few remarks. First of all, we derived two classical models for the dielectric function that were based on classical equations

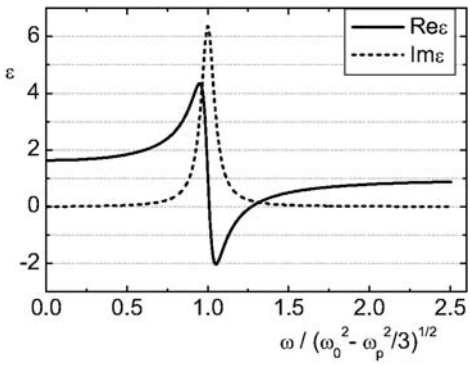

Fig. 3.3. Dielectric function according to (3.26)

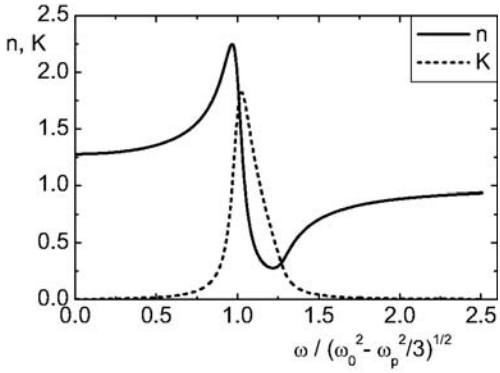

Fig. 3.4. Optical constants, according to Fig. 3.3

of motion for the charge carriers. In both models, we introduced energy dissipation by the damping factor γ. As seen from both equations (3.5) and (3.26), the introduction of energy dissipation leads to a non-zero imaginary part of the dielectric function. On the contrary, a non-vanishing Imε is only possible with a non-zero damping constant. Therefore, it is the imaginary part of the dielectric function that indicates the presence of light absorption in the sense that energy is transferred from the electromagnetic field to specific degrees of freedom in the medium. The large extinction coefficient in Fig. 3.1 (metal optics) rather leads to high reflection than to light absorption, because the imaginary part of the dielectric function is negligible in that spectral region.

The next remark concerns the local field effects discussed in Sect. 3.2.2. In our treatment, we assumed that the properties of the microscopic dipole itself do not change with increasing particle density. This is clearly a classical approach, because in reality chemical reactions start to occur when the molecules come close enough together for their electronic shells to overlap. However, in materials without covalent bonding, this simple theory (which is

Table 3.2. Recommendations on when to apply the local field correction, as collected from relevant textbooks. Here, E denotes the energy

Source		System	Applicability of the local field correction
R Feynman (Noble Price Winner) *The Feynman lectures of physics*		Metal	no
		Dielectric	yes
N. Bloembergen (Noble Price Winner) *Nonlinear Optics*		Ionic crystals (for example CuCl)	rather yes
		Valence electrons in crystalline semiconductors such as GaAs	no
		Liquids	yes
		Solids with a complicated elementary cell	rather yes
A.S. Davydov	Quantum mechanics	Systems with discrete energy levels	yes
	Solid state theory	Systems with energy bands $E = E(\boldsymbol{k})$	no

in fact 150 years old) may pretty work up to packing densities characteristic for a solid.

A further remark concerns the Drude-function as derived in the previous section. In this case, we did not distinguish between microscopic and average macroscopic fields. Why?

At least at low frequencies, a classical free electron may travel a considerably long way until the changing field direction forces it to return to its starting position. Therefore, during an oscillation, the electron rather feels the average field than a local one. The space "probed" by an oscillating free electron resembles the needle-like cavity shape with a vanishing depolarisation factor, so that (3.22b) rather holds than (3.22a). But this is a formal argument, and it remains unclear how to deal with high frequencies or a very low field strength.

In this situation we must remember that we deal with *models*. In every special case, one must accurately check whether or not the application of a given model makes sense. Concerning the validity of (3.22a), there is in fact no general recipe when it should be applied and when not. As a thumb rule, in optically isotropic media with well bound electrons, the application of (3.22a) makes sense. Presently, there is no general theory valid for media with any symmetry or more freely moving charge carriers. Table 3.2 summarizes some recommendations from the literature, concerning the application of the mentioned local field theory.

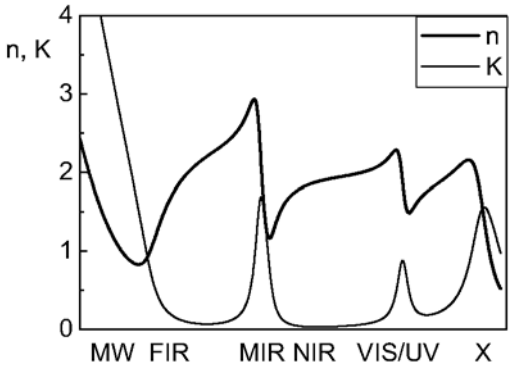

Fig. 3.5. Principal shape of the optical constants in different spectral regions

Table 3.3. Overview on spectral regions. The wavelength (and related) data may slightly differ in different sources

spectral region	vacuum wavelength λ nm	wavenumber ν $\nu = 1/\lambda$ cm^{-1}	angular frequency ω $\omega = 2\pi\nu c$ s^{-1}	origin of absorption (examples)
Far Infrared **FIR**	10^6–5×10^4	10–200	1.9×10^{12}–3.8×10^{13}	free carriers; orientation
Middle Infrared **MIR**	5×10^4–2.5×10^3	200–4000	3.8×10^{13}–7.5×10^{14}	free carriers; vibrations
Near Infrared **NIR**	2.5×10^3–8×10^2	4000–12500	7.5×10^{14}–2.4×10^{15}	free carriers; vibrational overtones
Visible **VIS**	8×10^2–4×10^2	12500–25000	2.4×10^{15}–4.7×10^{15}	excitation of valence
Ultra-violet **UV**	4×10^2–10	25000–10^6	4.7×10^{15}–1.9×10^{17}	electrons
X-ray **X**	10– 0.005	10^6–2×10^9 (unusual)	1.9×10^{17}–3.8×10^{20}	excitation of core electrons

3.3 Probing Matter in Different Spectral Regions

In summary, we are now familiar with three classical models that are useful to describe the optical properties of condensed matter. The orientation and reorientation of permanent molecular dipoles is very inert, and it will cause a remarkable optical response only in the microwave (MW) or far infrared (FIR) spectral regions. In liquids and also in some solids (for example ice) it may be tackled by means of Debye's equations. Drude's function describes the optical properties of free charge carriers, and depending on their concentration, it may be of use from the microwave up to the visible (VIS) spectral regions. The Lorentzian oscillator model is suitable for the description of absorption and dispersion in the presence of distinct spectral lines. In the middle infrared (MIR), it may be used to describe the response of core vibrations in molecules and solids. The excitation of valence electrons in atoms or molecules causes absorption lines in the visible or ultraviolet (UV) spectral regions, while core electron excitation dominates the x-ray region. An overview on the possible optical spectrum of condensed matter is given in Fig. 3.5. More quantitative information may be obtained from Table 3.3.

4 Derivations from the Oscillator Model

4.1 Natural Linewidth

From the dispersion models derived so far, the oscillator model is the most important one. Formally, it even contains the Drude model as the particular case $\tilde{\omega}_0 = 0$. We will therefore use this chapter to discuss in more detail the physics that are hidden in the simple equation (3.26).

Let us start from (3.18) for the polarizability β obtained for a single microscopic oscillator. If damping is weak, we have $\gamma^2 \ll \omega_0^2$. In the immediate vicinity of the resonance frequency, we assume $\omega_0 \approx \omega$, and then the imaginary part of the polarizability becomes:

$$\mathrm{Im}\beta \approx \frac{q^2\gamma}{2\omega_0\varepsilon_0 m} \frac{1}{(\omega_0 - \omega)^2 + \gamma^2}$$

This is a symmetric lineshape called a Lorentzian line. It describes the shape of an absorption line in terms of the classical oscillator model. As already mentioned in the previous section, the imaginary part of the dielectric function has the same shape as the polarizability, and such a lineshape is presented in Fig. 3.3 as the dashed line. Apart from resonance, $\mathrm{Im}\beta$ and $\mathrm{Im}\varepsilon$ decrease and achieve 50% of the maximum value at the frequencies:

$$\omega - \omega_0 = \pm\gamma$$

The value 2γ therefore represents the so-called Full Width at Half Maximum (FWHM) and is an important characteristic of a spectral line. In the present version of our classical theory, the width of a spectral line is exclusively determined by damping.

The FWHM as defined before is closely related to the decay time of the damped harmonic oscillator. Indeed, let us assume that the oscillator has been excited in the past and is now performing damped oscillations. Clearly, the energy will dissipate with time from the oscillator. The equation of free motion of a damped oscillator will be:

$$m\ddot{x} + 2\gamma m\dot{x} + m\omega_0^2 x = 0$$

and may be solved by means of the approach:

$$x = x_0 \, \mathrm{e}^{\xi t}$$

That leads to:

$$\xi = -\gamma \pm \sqrt{\gamma^2 - \omega_0^2} = -\gamma \pm \mathrm{i}\sqrt{\omega_0^2 - \gamma^2}$$

We further assume weak damping:

$$\omega_0 \gg \gamma^2$$

and obtain

$$x \approx x_0 \, \mathrm{e}^{-\gamma t} \, \mathrm{e}^{\pm \mathrm{i}\omega_0 t} \tag{4.1}$$

Equation (4.1) described the expected damping of the oscillation amplitude, with a decay time of $\tau_{\mathrm{amplitude}} = \gamma^{-1}$. As the energy is proportional to the square of the amplitude, it will dissipate with half the decay time, so that we get:

$$\tau_E = (2\gamma)^{-1} \tag{4.2}$$

Hence, the decay time for the energy τ_E equals the reciprocal value of the FWHM, when the latter is given in angular frequency units. The longer the energy remains in the system, the narrower is the corresponding absorption line. The linewidth defined by (4.2) is called the natural linewidth of the oscillator. This is a further example for the strong interconnections between the time response of a system and its spectral behaviour. If one is able to measure the natural linewidth experimentally, the decay time may be calculated.

But what use do we have from the decay time? We defined the decay time as the time necessary for the energy to dissipate from a microscopic oscillator. More accurately, it is the time when the energy has been decreased for e times. If we regard atoms or molecules as the microscopic oscillators, this classically defined decay time corresponds to the quantum mechanical lifetime of an excited atomic or molecular level. As it will be shown in Chap. 10, that lifetime in turn is connected to functions that describe the dynamics of molecules or atoms on a quantum mechanical level, and therefore yields information on their fundamental physics.

For intensive spectral lines, the lifetime is about 10^{-8}s. For so-called metastable levels, it may be of the order of 10^{-1} to 10^{-5}s.

4.2 Extended Detail: Homogeneous and Inhomogeneous Line Broadening Mechanisms

In practice, it is not so easy to measure the natural linewidth determined by (4.2), because in real matter it is not only the energy dissipation that causes

the finite linewidth. Other mechanisms may be of importance as well. First of all, we will have to consider that in most measurements we have not only one single oscillator, but rather an assembly. This makes the situation more complicated. We have already postulated that the single oscillator may loose its energy, but if there are several excited oscillators, they may come into collisions that destroy the phase of the vibration, but not the amplitude. If several such phase interruptions occur during τ_E, a line broadening will occur, because the phase interruptions distort the periodicity, so that the Fourier spectrum of the oscillation of the discussed molecule becomes broader. If all oscillators are in identical physical conditions, then all of them will suffer this broadening mechanism in an equal manner. In this case we speak on *homogeneous* line broadening.

There is another situation when the oscillators are in different physical conditions. For example, in disordered condensed matter, the particles (molecules, atoms) may "feel" different local fields, which leads to different spectral shifts according to (3.27). As a consequence, there are groups of oscillators absorbing at slightly different frequencies. Such an assembly may have a very broad absorption line, although every single oscillator shows a narrow line. The macroscopically detected broad line appears to be a superposition of a tremendous number of narrow lines resulting from oscillators that are placed in different physical conditions. Such a line broadening is called an *inhomogeneous* one. Clearly, in the case of inhomogeneous broadening, the line shape may significantly differ from the Lorentzian line shape.

As standard examples of a homogeneous and an inhomogeneous line broadening mechanism, the collision and Doppler broadening mechanisms of spectral lines in gases will be briefly addressed.

Collision Broadening

Stochastic elastic collisions between particles destroy the phase of their vibrations. Let us assume that the average time between two collisions $\tau_{\text{collision}}$ is much smaller than the energy decay time τ_E. In the case that the macroscopic polarization decays according to $\mathrm{e}^{-t/\tau_{\text{collision}}}$, according to (4.1), we get a linewidth

$$\Gamma = \tau_{\text{collision}}^{-1} \Rightarrow \text{FWHM} \equiv 2\Gamma = \frac{2}{\tau_{\text{collision}}} \tag{4.3}$$

The thus defined FWHM is called the homogeneous width of a spectral line. As our treatment of dielectric functions or the optical constants always concerns a large number of oscillators, starting from now we will use Γ in the dielectric function rather than γ that appeared in the microscopic oscillator. In collision broadening, the Lorentzian spectral shape is preserved.

In the general case, when both energy dissipation and collision broadening contribute to the observed FHWM, a more general equation is obtained as a merger between equations (4.2) and (4.3):

$$\text{FWHM} \equiv 2\Gamma = \frac{2}{\tau_{\text{collision}}} + \frac{1}{\tau_E} \tag{4.4}$$

According to (4.4), the previously discussed natural linewidth appears as a particular case of the homogeneous linewidth with negligible phase distortion.

Doppler Broadening

In contrast to collision broadening, the Doppler broadening in gases is an inhomogeneous broadening mechanism. As in equilibrium conditions the distribution of gas particles with respect to their velocities is symmetric, some of them fly in direction to the line source, and some of them away from the source. Due to the Doppler effect, the molecules moving in direction to the source may absorb at a slightly lower light frequency than those moving away. Hence, the molecules differ with respect to the physical condition essential for the process of light absorption. The full absorption line will be composed from a large number of narrow lines shifted with respect to each other due to the Doppler effect. This is a typical situation for an inhomogeneous line broadening.

This particular case may be mathematically treated in an exact manner. Let us assume that the light wave moves along the z-axis. Due to Maxwell's distribution, the number of molecules with a given z-component of their velocity is:

$$N(\nu_z)\,\mathrm{d}\nu_z \propto \mathrm{e}^{-\frac{m\nu_z^2}{2k_\mathrm{B}T}}\,\mathrm{d}\nu_z$$

m is the mass of a molecule, k_B Boltzmann's constant, and T the absolute temperature. Let ω_0 be the resonance frequency of the molecule in rest. Due to the movement along z, the molecule absorbs no longer at ω_0, but at a shifted frequency ω_D:

$$\omega_D = \omega_0\left(1 + \frac{\nu_Z}{c}\right)$$

The number of molecules absorbing at ω_D is then:

$$N(\nu_Z)\,\mathrm{d}(\nu_Z) = N(\nu_Z)\,\frac{\mathrm{d}\nu_Z}{\mathrm{d}\omega_D}\,\mathrm{d}\omega_D \equiv N(\omega_D)\;\mathrm{d}\omega_D$$

so that we finally get:

$$N(\omega_D) = \left[N(\nu_Z)\,\frac{\mathrm{d}\nu_Z}{\mathrm{d}\omega_D}\right]_{\nu_Z = f(\omega_D)} \tag{4.5}$$

Expression (4.5) reveals the probability density distribution for the Doppler-shifted absorption frequency ω_D in an assembly of gas molecules or atoms. In the case that this distribution is considerably broader than the

homogeneous linewidth, the absorption line shape of the assembly will be dominated by (4.5). We then find a Gaussian spectral shape with an FWHM given as:

$$\Delta\omega_D = \frac{2\omega_0}{c}\sqrt{\frac{2\ln 2 k_\mathrm{B} T}{m}}$$

which is, of course, dependent on the temperature.

These two examples were to show, that the important information on the natural linewidth is not easily accessible by the experiment, because line broadening mechanisms occur that broaden the line and may even change the spectral shape. For Doppler broadening, we got a Gaussian lineshape, but other lineshapes are possible as well in inhomogeneously broadend assemblies.

4.3 Oscillators with More Than One Degree of Freedom

We will now generalize the oscillator model to the so-called multi-oscillator model. Instead of one resonance frequency, we have now a set of M resonance frequencies $\{\omega_{0j}\}$ for each oscillator, and the natural generalization of (3.25) becomes:

$$\beta = \frac{q^2}{\varepsilon_0 m}\sum_{j=1}^{M}\frac{f_j}{\omega_{0j}^2 - \omega^2 - 2\mathrm{i}\omega\Gamma_j} = \frac{3}{N}\,\frac{\hat{n}^2-1}{\hat{n}^2+2} = \frac{3}{N}\,\frac{\varepsilon-1}{\varepsilon+2} \tag{4.6}$$

The factor f_j describes the relative strength of the absorption lines caused by the different degrees of freedom. If we deal with a molecule, for example, different normal vibrations of the cores or various electronic oscillations may thus be taken into account. Figure 4.1 compares the dispersion of the dielectric functions for single- and multioscillator models.

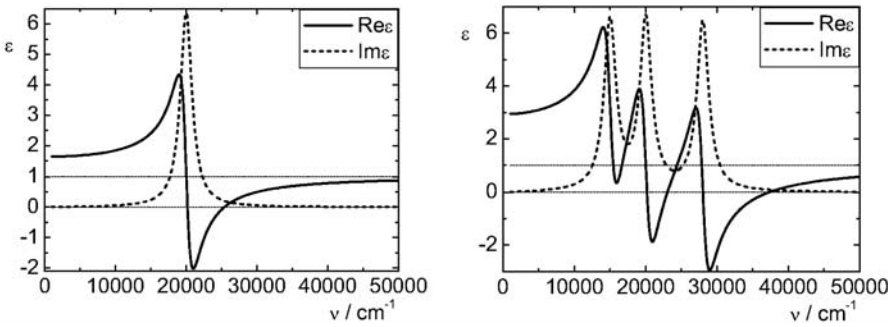

Fig. 4.1. Comparison between the dielectric functions for a single- (**left**) and multioscillator (**right**) models

The static value of the dielectric function is influenced by all resonances and may be calculated assuming $\omega = 0$ in (4.6):

$$\beta_{\text{stat}} = \frac{q^2}{\varepsilon_0 m} \sum_{j=1}^{M} \frac{f_j}{\omega_{0j}^2} = \frac{3}{N} \frac{\varepsilon_{\text{stat}} - 1}{\varepsilon_{\text{stat}} + 2}$$

By the way, this expression describes the previously excluded case of ferroelectrics as the limiting case $\beta_{\text{stat}} \to 3N^{-1}$. In this limit, $\varepsilon_{\text{stat}} \to \infty$.

4.4 Sellmeier's and Cauchy's Formulae

This section will not yield any new physics. There exists a variety of common dispersion formulae that may be regarded as special cases of (4.6). They are often cited in the literature, so that it is worth to mention some of them.

As seen from Fig. 4.1, in the multioscillator model there may be spectral regions where the dielectric function is almost real. This is the case when the condition:

$$(\omega_{0j} - \omega)^2 \gg \Gamma_j^2 \quad \forall j$$

is fulfilled. These are the transparency regions of real materials, which are utmost important for their use as optical materials. In that non-resonant case, the dielectric function following from (4.6) may be simplified.

Let us start from the dielectric function of the multioscillator model. Generally, from (4.6) it follows that the dielectric function may be written as:

$$\varepsilon = 1 + \sum_{j=1}^{M} \frac{\tilde{f}_j}{\tilde{\omega}_{0j}^2 - \omega^2 - 2\mathrm{i}\omega\Gamma_j}. \tag{4.7}$$

Again, $\tilde{f}_j$ is responsible for the intensity of the spectral lines. It does not make sense to accurately write out the full classical intensity expressions that would follow from (4.6) during the following derivations, because relevant expressions have to be obtained later in terms of the semiclassical mechanical treatment. We only mention here that expression (4.7) may always be obtained from (4.6) by an expansion into partial fractions. That will also give the final expressions for $\tilde{f}_j$ and $\tilde{\omega}_{0j}$, respectively. Note that expression (3.27) is not valid if more than one resonances are involved.

Far from any resonances, (4.7) may be written as:

$$\varepsilon = \mathrm{Re}\varepsilon \approx 1 + \sum_{j=1}^{M} \frac{\tilde{f}_i}{\tilde{\omega}_{0j}^2 - \omega^2}$$
$$\mathrm{Im}\varepsilon = 0$$

Replacing ω by λ via

$$\omega = 2\pi \frac{c}{\lambda}$$

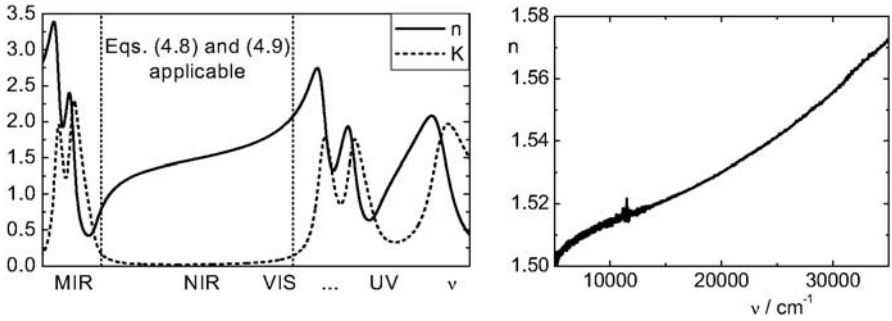

Fig. 4.2. Applicability region of equations (4.8) and (4.9) (**left**) and measured refractive index of an optical glass (**right**)

and utilizing

$$\frac{\lambda^2}{\lambda^2 - \tilde{\lambda}_{0j}^2} \equiv 1 + \frac{\tilde{\lambda}_{0j}^2}{\lambda^2 - \tilde{\lambda}_{0j}^2} \,.$$

Equation (4.7) may be written as:

$$\varepsilon - 1 = n^2 - 1 = a + \sum \frac{b_j}{\lambda^2 - \tilde{\lambda}_{0j}^2} \tag{4.8}$$

where a and b_j are constant coefficients. They are interconnected with each other due to the requirement that the refractive index must approach one when the wavelength approaches zero. Equation (4.8) is known as Sellmeier's dispersion formula. In Fig. 4.2, an example for the applicability of (4.8) is sketched.

Another common dispersion formula is obtained expanding (4.8) into a power series. We rewrite (4.8) according to:

$$n^2(\lambda) = 1 + a - \sum_{\tilde{\lambda}_{0j} > \lambda} \frac{b_j}{\tilde{\lambda}_{0j}^2 - \lambda^2} + \sum_{\tilde{\lambda}_{0j} < \lambda} \frac{b_j}{\lambda^2 - \tilde{\lambda}_{0j}^2}$$

where the first sum contains the long-wavelength resonances, and the second one the short-wavelength ones. Expanding the first sum into a power series of $\left(\lambda/\tilde{\lambda}_{0j}\right)^2$ and the second sum into a series of $\left(\lambda/\tilde{\lambda}_{0j}\right)^{-2}$, by means of

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots$$

we find

$$n^2 = A + B\nu^2 + C\nu^4 + \ldots - B'\nu^{-2} - C'\nu^{-4} - \ldots \tag{4.9}$$

Here, the A-, B- and C-values are new constants. Equation (4.9) is again applicable in the range indicated in Fig. 4.2. For illustration, the experimentally determined refractive index of an optical glass is also presented. More details on this subjects will be given in Chap. 6.

Sometimes, (4.9) is called Cauchy's dispersion formula. In other sources, the terminus 'Cauchy's formula' is only applied to gases, where the refractive index is close to one because of the low particle concentration. In this case, we have:

$$n^2 - 1 = (n+1)(n-1) \approx 2(n-1)$$

so that instead of (4.9) one obtains:

$$n = A + B\nu^2 + C\nu^4 + \ldots - B'\nu^{-2} - C'\nu^{-4} - \ldots \tag{4.10}$$

where the A-, B- and C-values generally differ from those in (4.9).

Another version of such simplified dispersion equations is often applied in the infrared spectral region. In resonance spectroscopy, it makes sense to separate the resonant contributions in (4.7) from the rest of the full dielectric function. These are the terms for which $\tilde{\omega}_{0j} \approx \omega$. We thus define the resonant contribution to the susceptibility as:

$$\chi_{\mathrm{res}} \equiv \sum_{\omega \approx \tilde{\omega}_{0j}} \frac{\tilde{f}_j}{\tilde{\omega}_{0j}^2 - \omega^2 - 2\mathrm{i}\omega\Gamma_j}$$

The other terms in the sum in (4.7) form the nonresonant contribution χ_{nr}. We find:

$$\varepsilon(\omega) = 1 + \chi_{\mathrm{res}}(\omega) + \chi_{\mathrm{nr}}(\omega)$$

In the infrared, the non-resonant contribution mainly originates from high-frequency electronic resonances. It is therefore a common practice to neglect the dispersion of the non-resonant terms and to define a purely real 'background' dielectric function ε_∞ as:

$$\varepsilon_\infty = 1 + \chi_{\mathrm{nr}}$$

In this language, we may rewrite the dispersion formulae derived so far in the following manner:

$$\text{Debye:} \quad \varepsilon = \varepsilon_\infty + \frac{\varepsilon_{\mathrm{stat}} - \varepsilon_\infty}{1 - \mathrm{i}\omega\tau}$$

$$\text{Drude:} \quad \varepsilon = \varepsilon_\infty - \frac{\frac{\sigma_{\mathrm{stat}}}{\varepsilon_0}}{\omega^2\tau + \mathrm{i}\omega}$$

$$\text{Single-Lorentz-Oszillator:} \quad \varepsilon = \varepsilon_\infty + \frac{\tilde{f}}{\tilde{\omega}_0^2 - \omega^2 - 2\mathrm{i}\omega\Gamma}$$

All these equations represent special cases of the previously discussed equation (3.1). In the particular form as written here, they may be approximately valid in the frequency region where the separated polarization mechanism is close to resonance. They are not appropriate for describing the asymptotic behaviour far from resonance.

4.5 Optical Properties of Mixtures

4.5.1 Motivation and Example from Practice

In practice, one often has to deal with situations where the optical properties of material mixtures are of interest. It is naturally to assume, that the optical constants of the mixture represent some kind of superposition of the optical constants of their constituents. The question is, how to superimpose the optical constants of the constituents?

First of all, let us assume that each of the constituents numbered by j occupies a certain volume fraction V_j of the material, and this volume fraction determines the filling factor p_j of the material via:

$$p_j \equiv \frac{V_j}{V}$$

where V is the full volume occupied by the mixture. Obviously,

$$\sum_j p_j = 1$$

One could now assume that it makes sense to superimpose the dielectric functions of the constituents linearly to obtain the so-called effective dielectric function of the mixture via:

$$\varepsilon_{eff} = \sum_j p_j \varepsilon_j \tag{4.11}$$

Let us look at an example how such a simple approach would work.

We regard a material that is composed from aluminum oxide Al_2O_3 with embedded small silver particles. The terminus 'small' means that the diameter of the particles and their average distances are small compared to the wavelength, so that the material appears to be optically homogeneous, although it might be heterogeneous on a nanometer scale. Such composite materials are quite easy to produce by evaporation in vacuum conditions and surprise by their beautifully coloured appearance. Of course, for practical applications (for example in absorber designs) one must accurately know their optical constants, so that this example will accompany us throughout this section.

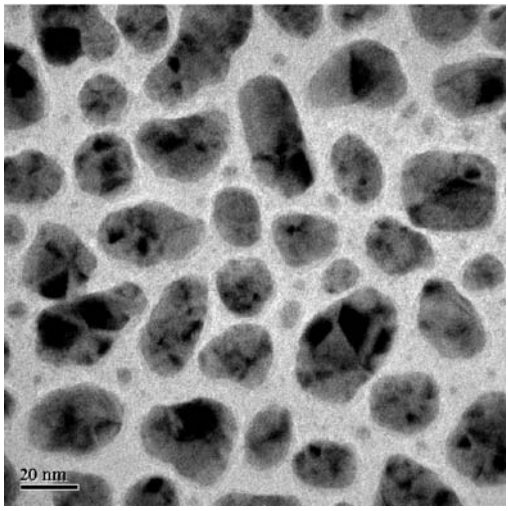

Fig. 4.3. TEM-image of a composite thin film built from Al_2O_3 with embedded isolated silver particles

But before starting with optics, let us have a look at the real structure of such a composite material. Figure 4.3 shows a transmission electron microscopy (TEM) image of this material. The bar in the left corner on the bottom of the image indicates a length of 20 nm. In this image, the silver particles are seen as dark spots. Obviously, they differ from each other in size, shape and relative orientation.

Nevertheless, the cluster size is well below the wavelength in the visible spectral region, so that we will treat the material as optically homogeneous. In particular that means, that we may make use of the quasistatic approximation. In the present sample, the filling factor of the silver fraction is approximately 0.3. Accordingly, that of Al_2O_3 is 0.7.

Let us now see how (4.11) will work. Figure 4.4 presents the dielectric functions from the individual constituents of the mixture (Ag and Al_2O_3). The imaginary part of the dielectric function of Al_2O_3 is negligible in the discussed spectral range when compared to the real part, so that it is not shown in the figure. The dielectric function of silver is obviously dominated by a Drude term (compare figure 3.1) while the dispersion of Al_2O_3 could be close to the Sellmeier-type dispersion sketched in the central region of Fig. 4.2.

It is now straightforward to apply (4.11) with the mentioned filling factors. The result is shown in Fig. 4.5, together with the experimentally determined data.

As seen from Fig. 4.5, the agreement between experiment and theory needs some improvements. Soft phrases of this kind in any scientific work always have the meaning: there is no agreement at all. Consequently, at least in the present example, the simple linear superposition of dielectric functions to model the optical behaviour of a mixture does not work at all.

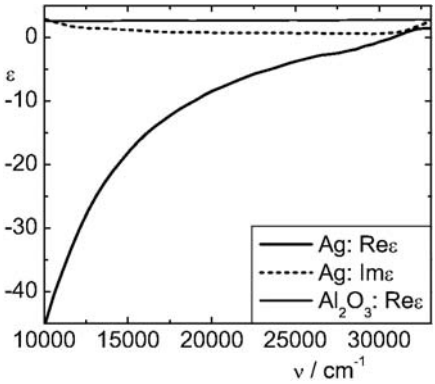

Fig. 4.4. Dielectric functions of Ag and Al_2O_3

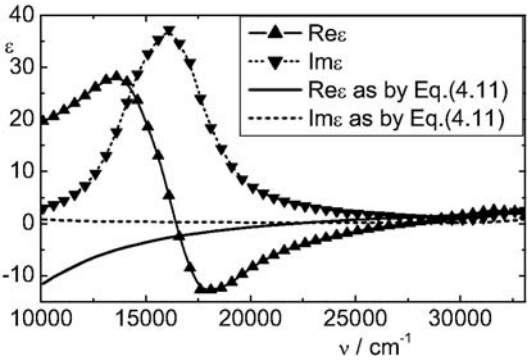

Fig. 4.5. Dielectric function of the composite material

But what is wrong? Is this the moment, where we have to turn to the quantum mechanical description? Or is there any explanation in the framework of the classical theory?

First of all, let us remember that (4.11) has not been derived, but purely guessed. And the guess was wrong, at least in application to the system shown in Fig. 4.3.

Secondly, let us remark that (fortunately) there is still no need to apply the apparatus of quantum mechanics. The behaviour of the dielectric function from Fig. 4.5 may be reproduced in terms of classical electrodynamics, but only after serious modifications in our theoretical description of the optical behaviour of the composite. By the way, let us state in advance, that (4.11) may work well in particular cases, which will be specified later. Our task at the moment is rather to understand what was wrong, and to derive a more general equation for the dielectric function of mixtures as (4.11).

A glance at the dispersion shown in Fig. 4.5 already reveals a crucial point: The dielectric function resembles the shape like shown in Fig. 3.3, so that it is the oscillator model that could be suitable to the system. But

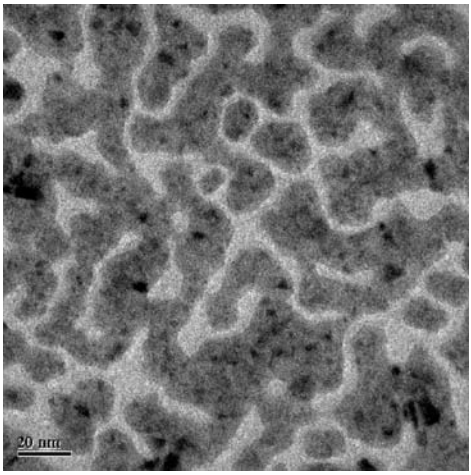

Fig. 4.6. TEM-image of a composite thin film built from Al_2O_3 with embedded percolated silver particles

that has been derived for *bound* electrons. On the contrary, the spectral behaviour of bulk silver is close to that of a Drude metal and therefore rather determined by free electrons. The key point is, that due to the confinement in small particles (clusters) as shown in Fig. 4.3, the 'free' electrons are not really free, but rather 'bound' in the clusters. When applying a static electric field, no remarkable electrical current would flow through such a system. Consequently, the straightforward implementation of the dielectric response of silver into (4.11) cannot be correct.

It is easy to give an experimental cross-check on the validity of this assumption. The only thing we must do is to prepare a system from the same material in such a manner that the silver particles are not isolated from each other, but form a closed network so that a direct current could flow (percolation of the clusters). Such a system is demonstrated in Fig. 4.6.

If our assumption was true, then the optical behaviour of this system should be completely different from that in Fig. 4.3 (and perhaps closer to that predicted by (4.11)). Figure 4.7 answers this question.

Quite obviously, the dielectric function of the composite material is sensitive to the *morphology* of the composite. Although the constituent materials are essentially the same, the optical behaviour of the systems from Fig. 4.3 and 4.6 are completely different from each other. Clearly, in (4.11), the morphology doesn't play any role, and that might be the mistake we have made so far. By the way, the output from (4.11) is at least qualitatively comparable to the behaviour of the percolated system. Hence, it will now be our task to derive a more general sophisticated mathematical apparatus suitable for the description of the optical properties of material mixtures.

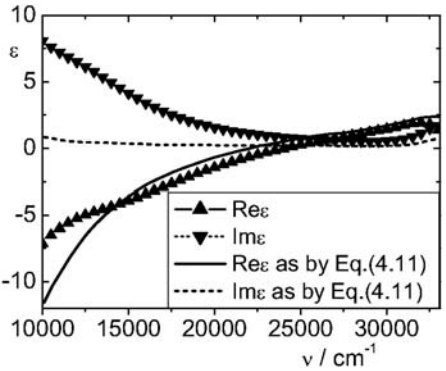

Fig. 4.7. Dielectric function of the material shown in Fig. 4.6

4.5.2 Extended Detail: The Maxwell Garnett, Bruggeman, and Lorentz–Lorenz Mixing Models

Let us have one further glance at Fig. 4.3. Obviously, there are particles of one material (the inclusions) that are embedded in another material (the host). In complete analogy with the philosophy from Sect. 3.2.2, we will start our treatment with the discussion of *spherical* inclusions. And we will again assume, that the inclusions may be regarded as polarized spheres in a homogeneous electric field (quasistatic approximation). The difference to Sect. 3.2.2. is, that the sphere is now embedded in a host with another dielectric function.

That leads to a modification in the expression of the microscopic field. Instead of (3.22a) we now have:

$$E_{\text{micr}} = \tfrac{\varepsilon+2\varepsilon_h}{3\varepsilon_h} E \tag{4.12a}$$

Here, E is the average field in the sphere, ε the dielectric function of the inclusion material, and ε_h that of the host. If the host dielectric function is equal to 1, then (4.12a) and (3.22a) are identical. The derivation of (4.12a) may be found in textbooks on electrodynamics.

In order to make the discussion complete, we present the analogue to (3.22d) for other inclusion shapes:

$$E_{\text{micr}} = \tfrac{\varepsilon_h+(\varepsilon-\varepsilon_h)L}{\varepsilon_h} E \tag{4.12b}$$

where L has the same meaning as in equation(3.22d) and Table 3.1. In particular, for $L = 0$ (needle) we have $E_{\text{micr}} = E$, and for $L = 1$ (pancake) $\varepsilon_h E_{\text{micr}} = \varepsilon E$. These are the familiar boundary conditions for the electric field tangential and normal to a surface.

The last step is to calculate the polarization. Every inclusion may be characterized by its linear polarizability β. Because the polarizabilities of the inclusion and the host β_h are different, an excess dipole moment is formed at the boundary of the inclusion. It may be calculated via

$$p = \varepsilon_0(\beta - \beta_h)E_{\text{micr}} = \varepsilon_0(\beta - \beta_h)\frac{\varepsilon_h + (\varepsilon - \varepsilon_h)L}{\varepsilon_h}E \tag{4.13}$$

Let N be the number of inclusions per occupied volume. Their dipoles contribute to the full polarization according to:

$$Np = \varepsilon_0(\chi - \chi_h)E = \varepsilon_0(\varepsilon - \varepsilon_h)E = N\varepsilon_0(\beta - \beta_h)\frac{\varepsilon_h + (\varepsilon - \varepsilon_h)L}{\varepsilon_h}E$$

From here we find immediately:

$$(\beta - \beta_h) = \varepsilon_h V\frac{(\varepsilon - \varepsilon_h)}{\varepsilon_h + (\varepsilon - \varepsilon_h)L}$$

Here, V is the average volume occupied by a single inclusion. Let us now assume that we have different kinds of inclusions numbered by j, each of them polarizing in the same host material. They cause the full polarizability

$$\sum_j(\beta_j - \beta_h) = \varepsilon_h\sum_j V_j\frac{(\varepsilon_j - \varepsilon_h)}{\varepsilon_h + (\varepsilon_j - \varepsilon_h)L} \tag{4.14}$$

At the same time, the medium may be thought to be built from identical structural units, embedded into the host medium occupying a volume V and having some average 'effective' dielectric function ε. Of course, their polarizability β must be equal to what is supplied by the real dipoles. Hence, we assume for the 'effective' medium:

$$\beta - \beta_h = \varepsilon_h V\frac{(\varepsilon - \varepsilon_h)}{\varepsilon_h + (\varepsilon - \varepsilon_h)L}; \quad V = \sum_j V_j \tag{4.15}$$

and demand:

$$\sum_j(\beta_j - \beta_h) = \beta - \beta_h \tag{4.16}$$

Then, from (4.14)–(4.16) we finally obtain the general mixing formula:

$$\frac{(\varepsilon - \varepsilon_h)}{\varepsilon_h + (\varepsilon - \varepsilon_h)L} = \sum_j p_j\frac{(\varepsilon_j - \varepsilon_h)}{\varepsilon_h + (\varepsilon_j - \varepsilon_h)L} \tag{4.17a}$$

For spherical inclusions, $L = 1/3$, and (4.17a) becomes:

$$\frac{\varepsilon - \varepsilon_h}{\varepsilon + 2\varepsilon_h} = \sum_j p_j\frac{\varepsilon_j - \varepsilon_h}{\varepsilon_j + 2\varepsilon_h} \tag{4.17b}$$

Equations (4.17a) or (4.17b) represent general optical mixing formulae. Of course, all dielectric functions here may be complex and frequency dependent. The effective dielectric function of the mixture appears to depend on

the dielectric functions of the constituents, their filling factors, and the morphology (via L). However, it is still a function of the somewhat dubious value ε_h, that was necessary to introduce at the beginning of the derivation, but, unfortunately, didn't disappear at the end. The following approaches exist to deal with ε_h.

Maxwell Garnett (MG) Approach

It might be the most natural choice to regard one of the constituents (say, the l-th one) as the host material, and the others as the inclusions. In the case of Fig. 4.3, it clearly makes sense to regard silver as inclusion and the dielectric as the host. That is the philosophy of the Maxwell Garnett approach. In this case, we have:

$$\frac{(\varepsilon - \varepsilon_l)}{\varepsilon_l + (\varepsilon - \varepsilon_l)L} = \sum_{j \neq l} p_j \frac{(\varepsilon_j - \varepsilon_l)}{\varepsilon_l + (\varepsilon_j - \varepsilon_l)L} \tag{4.17c}$$

Note that the sum of the filling factors on the right hand is now less than 1. In application, one must keep in mind that (4.17c) depends on the choice of the host function: It makes a great difference whether material 1 is embedded in material 2 or vice versa.

Lorentz-Lorenz (LL) Approach

As in Sect. 3.2.2, the Lorentz–Lorenz approach assumes that all inclusions polarize in vacuum ($\varepsilon_h = 1$). We therefore obtain:

$$\frac{(\varepsilon - 1)}{1 + (\varepsilon - 1)L} = \sum_{j} p_j \frac{(\varepsilon_j - 1)}{1 + (\varepsilon_j - 1)L} \tag{4.17d}$$

Effective Medium Approximation (EMA) or Bruggeman Approach

Another possibility is to assume that the effective dielectric function itself acts as the host medium for the inclusions. This leads to the following mixing formula:

$$0 = \sum_{j} p_j \frac{(\varepsilon_j - \varepsilon)}{\varepsilon + (\varepsilon_j - \varepsilon)L} \tag{4.17e}$$

There is no general recipe which of these approaches works best. As a rule, the MG theory works best when the constituents clearly may be subdivided into inclusions and one matrix material. On the contrary, in the presence of percolation or in molecular mixtures, the application of the EMA may lead to the best results. Finally, highly porous materials might be well fitted within the LL approach.

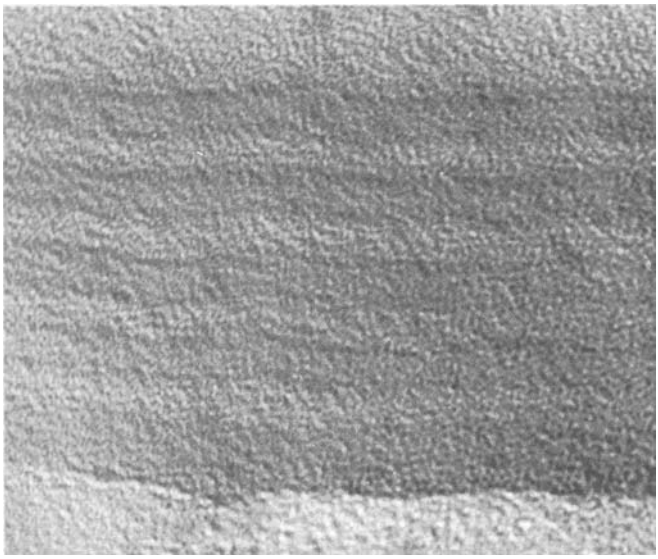

Fig. 4.8. Ultrathin multilayer structure, composed from amorphous silicon (a-Si, dark) and an organic material (copperphthalocyanine CuPc). In contrast to the previous images, this is a cross-sectional image. In average, each a-Si layer has a thickness of approximately 3.5 nm, and each CuPc layer a thickness of about 2 nm. In the VIS, the thickness values are much smaller than the wavelength, so that for in-plane polarization of the electromagnetic wave, (4.11) may find application

Before comparing our newly derived equations with the experimental composite data, let us make a final comment on (4.11). In which cases it makes sense to apply (4.11)?

Let us assume a pancake structure with an electrical field perpendicular to the cavity axis. From Table 3.1 we find, that $L = 0$. Equation (4.17a) immediately becomes:

$$\varepsilon = \sum_j p_j \varepsilon_j$$

which is identical to (4.11). Due to $L = 0$, the system may be regarded as a layered structure, while the electric field vector is parallel to the planes. On the other hand, the quasistatic approximation still holds, so that the layer thickness must be much smaller than the wavelength (see Fig. 4.8). In this case, the system behaves like a couple of capacitors in parallel combination. Those capacitances add up to the full capacitance, and (4.11) is a natural conclusion from there.

On the other hand, when the electric field vector is normal to the planes, the same argumentation would lead to the statement that the capacitors are now combined in series. In this case, one would expect:

$$\varepsilon^{-1} = \sum_j p_j \varepsilon_j^{-1}$$

In fact, in order to deal with this situation, we have now to assume $L = 1$ and to use (4.17e). From that, the given equation will be easily obtained.

4.5.3 Extended Detail: Remarks on Surface Plasmons

Let us now return to our practical problem. We have derived a couple of equations that could be helpful to reproduce the experimentally observed dispersion. Figure 4.9 shows the effective dielectric functions obtained for our case by (4.17c)–(4.17e) assuming $L = 1/3$ (spherical inclusions).

It turns out, that the different models give quite different results concerning the effective dielectric function. In any case, we succeeded to qualitatively reproduce the experimentally observed behaviour of the dielectric function from Fig. 4.5: At least in the MG and LL models, we get a distinct absorption line that has some similarity with the behaviour we wanted to reproduce. So that these models seem to be applicable in real life.

But what is the reason for the absorption line? As already mentioned, in small metal islands, the motion of the previously "free" electrons is confined inside the particle, so that the electrons behave optically in a similar manner like bound electrons. These electrons may perform collective oscillations (plasma oscillations). Clearly, the electrons are inert and suffer a restoring force, hence their oscillation causes a resonant absorption behaviour. This is the *physical* reason of the observed absorption line.

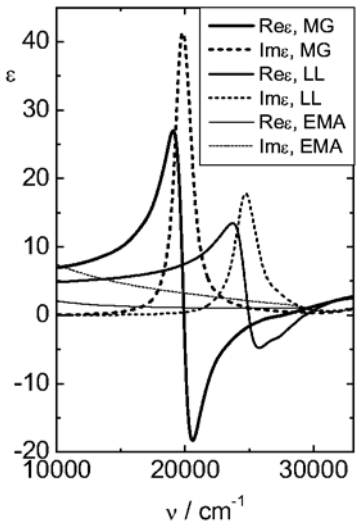

Fig. 4.9. Dielectric functions as obtained from different mixing models; $L = 1/3$

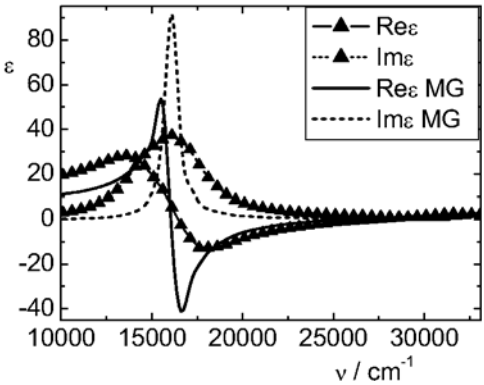

Fig. 4.10. Experimental data and calculation by the Maxwell Garnett approach with an assumed depolarisation factor of 0.21 (*elongated particles*)

In the language of quantum physics, the collective motion of electrons is identical to a superposition of elementary oscillations called plasmons. The term "surface plasmon" originates from the fact, that in a small metal particle a net charge appears only at its surface – inside the particle the charges of the electrons are compensated by the positive charges of the cores in the same manner as it would happen in the bulk metal. The surface charges may form a dipole moment of the particle, that may effectively couple to the impinging light, so that such surface plasmons are easily excited through absorption of electromagnetic radiation.

It is not the purpose of this chapter to go into full detail here. We only mention, that the resonant behaviour *mathematically* originates from vanishingly small denominators such like those in (4.17c). For a small metal particle in a dielectric host, a resonance therefore occurs, when

$$|\varepsilon_h(\omega) + [\varepsilon_{metal}(\omega) - \varepsilon_h(\omega)]\, L| \to \min \tag{4.18}$$

is fulfilled. As the real part of the metal dielectric function is usually negative, there will exist one or more frequencies where condition (4.18) is fulfilled. A further analysis of this condition shows, that the resonance frequency depends on the particle shape (via L) and the value of the dielectric function of the embedding medium. As a thumb rule, one may assume that an increase in the host dielectric function decreases the surface plasmon resonance frequency, while an increase in L increases the resonance frequency.

Let us finally directly compare the result obtained from the Maxwell Garnett model with the experimental data. As expected, the MG data lead to the best results here, because we really deal with isolated inclusions in the matrix. Nevertheless, the agreement between the theory from Fig. 4.9 and experiment is still only qualitative, due to the mismatch between the resonance frequencies. This is partially caused by the assumed spherical geometry ($L = 1/3$). The problem may be overcome by choosing another depolarisa-

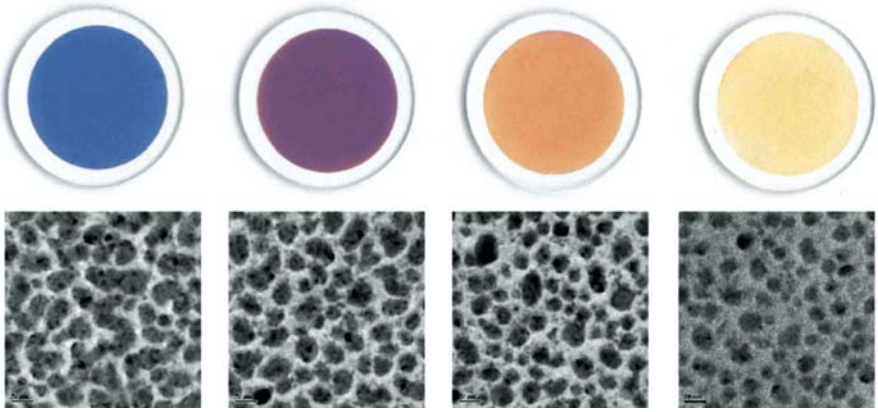

Fig. 4.11. Optical appearance of silver island films in lanthanum fluoride (**upper part**) and transmission electron micrographs of the samples

tion factor of $L = 0.21$, corresponding to moderately elongated particles. The result is shown in Fig. 4.10.

The agreement between experiment and theory has now improved. Clearly, as seen from Fig. 4.3, the assumption about spherical inclusions was very rough. Most of the silver material is concentrated in elongated clusters, so that resonances with $L < 1/3$ significantly contribute to the response of the system. In practice, they cause a red-shift of the absorption line. Moreover, the particles are statistically distributed with respect to shape and orientation. That leads to an inhomogeneous broadening of the line. For that reason, the experimentally obtained resonance is broader than that predicted by the simple MG-calculation. Of course, there exist more sophisticated approaches that generalize (4.17a) to a statistical superposition of particles with different depolarisation factors. That will clearly result in a better agreement between theory and experiment. We will not go into these details, but will state here that the general features of the dielectric function of mixtures may be reproduced basing on the equations derived so far. More powerful algorithms to calculate the response of such composite systems are provided by the Bergman theory, the Mie-theory and its modern generalizations to spherical cluster assemblies, and the Rigorous Coupled Wave Approach (RCWA). The interested reader is here referred to the special literature.

Finally, let us demonstrate the optical appearance of silver island films at a concrete example. Figure 4.11 (upper part) shows the colour of silver islands embedded in lanthanum fluoride. Below, for each sample, a transmission electron micrograph is presented, corresponding to a sample area of $170\,\mathrm{nm} \times 170\,\mathrm{nm}$. The silver islands are seen as dark spots. It is clearly seen, that the different colour of the samples correspond to quite different silver island geometries, as they are characteristic for each of the individual samples.

4.5.4 Extended Detail: The Effect of Pores

As the last example in this chapter, we turn to another problem that is of practical significance for thin film optical materials. Although we did not yet deal with optical films, it should be intuitively clear that the optical response of optical films is determined by both material properties and their geometry. As most optical films are grown today 'from bottom to top' in vacuum conditions with a considerably high growth rate, it is clear that they may contain plenty of defects, among them pores. These pores may be empty or filled with water, and in any case they will affect the optical constants of the film material.

To consider the effects of pores in our theory, the simplest possibility is to regard the film material as a mixture of the 'pure material' (with a refractive index n_0) and the pores. The problem then appears as a particular case of the theory developed in Sect. 4.5.2, and is traditionally solved by means of (4.17c). For convenience, we will assume that the indices of refraction are purely real, but this does not affect the generality of the model.

That still gives a lot of possibilities to consider different types of morphology as well as different optical constants of the 'pore material' (empty pores or water). The remaining task is to derive an explicit expression for the optical constants of the film material. That may easily be done by the reader himself, and we restrict ourselves to one practically important particular case, namely the case of free standing cylindrical rods. This is a model system that is often applied to films with a columnar structure, as they may be grown by evaporation techniques. For the sake of illustration, Fig. 4.12 shows an electron micrograph of a magnesium fluoride film that exhibits such a columnar structure.

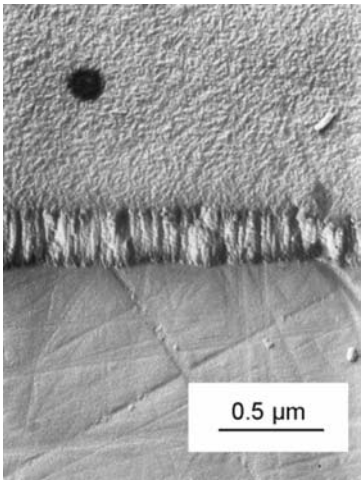

Fig. 4.12. Columnar structure of a magnesium fluoride film. Courtesy of Norbert Kaiser

The filling factor of the solid fraction p (filling factor of the rods) is often called the *packing density* of the film. For normal incidence, the electric field vector is perpendicular to the rods. According to our classification from Chap. 3, these rods should correspond to the previously discussed *needles*. Therefore, as seen from Table 3.1, the corresponding depolarisation factor must be close to 0.5. Assuming that the rods are embedded in air or vacuum with a refractive index of 1, (4.17c) (which is in this particular case identical with (4.17d)) will yield:

$$n^2 = \frac{2 + \left(n_0^2 - 1\right)(1+p)}{2 + \left(n_0^2 - 1\right)(1-p)} \tag{4.19a}$$

In the more general case of filled pores, (4.19a) becomes

$$\begin{aligned} n^2 &= n_v^2 \left. \frac{n_v^2 + [L(1-p)+p]\left(n_0^2 - n_v^2\right)}{n_v^2 + L(1-p)\left(n_0^2 - n_v^2\right)} \right|_{L=\frac{1}{2}} \\ &= n_v^2 \frac{2n_v^2 + (1+p)\left(n_0^2 - n_v^2\right)}{2n_v^2 + (1-p)\left(n_0^2 - n_v^2\right)} = \frac{(1-p)n_v^4 + (1+p)n_v^2 n_0^2}{(1+p)n_v^2 + (1-p)n_0^2} \end{aligned} \tag{4.19b}$$

where n_v is the refractive index of the void material (usually water). Quite fortunately, from Sect. 2.5 we already have some idea on the optical constants of water. With a refractive index of 1.33 in the VIS, it still has a lower refractive index than the optical materials that are typically in use today. Equation (4.19b) is known as the mixing formula of Bragg and Pippard.

Let us look at these equation in some more detail. Obviously, for a vanishing packing density, the refractive index approaches that of the void fraction. If the packing density is equal to 1, the film's refractive index becomes equal to that of the pure film material. For intermediate packing densities we will find indices that are in between n_v and n_0. On the other hand, the mass density ρ of the film is:

$$\rho = p\rho_0 + (1-p)\rho_v \tag{4.20}$$

Therefore, (4.17c–e) and (4.20) define characteristic relationships between the mass density and the refractive index, which are experimentally accessible and may, in principle, be used to decide which of the possible mixing models best describes the given film material. Particularly, for empty pores and negligible absorption, one will obtain that the refractive index generally increases with increasing mass density. This seems to be quite understandable, because the response of the system is expected to become stronger when more oscillators are available, which is consistent with a higher mass density.

In fact, one could finish the chapter with this fine result. The only problem is, that our approach does often work, but unfortunately not always.

We have already mentioned, that thin solid films are often produced by evaporation in vacuum conditions. This is a common method, but there exist

other techniques such as for example the sputtering techniques, which generally yield films with other properties than the evaporated ones. In these films, the pores may be much smaller. For example, in sputtered amorphous silicon and germanium films, pores have been found with diameters of about 0.7 nm, and the film's refractive index has been found to be essentially *higher* than that of completely dense (crystalline) films. How can we explain this behaviour?

In this connection, we come to the generally important question of the *size* of the inclusions assumed in the mixing models. In our models described so far, the size does not play any role, as long as the inclusions are much smaller than the wavelength of the light. If they become too large, the medium can no longer be regarded as optically homogeneous, which will result in light scattering that causes a turbid appearance of the material. In this case, our theory will clearly be at stake. However, we are now confronted with the new fact that the inclusions should not be too small. For very small inclusions, our theory may give misleading results as well.

The new thing we find here is called a *size effect*. When the size of the inclusions comes into play, equations (4.17) are no more valid. We speak on *extrinsic* size effects when the size of the inclusions becomes too large compared to the wavelength of the light. These extrinsic size effects are of purely classical nature and may be calculated in terms of Mie's famous theory. On the contrary, when the inclusions are too small, it is no longer correct to describe them in terms of the conventional dielectric function. The dielectric function itself is a macroscopic measure accurately defined for a thermodynamically relevant number of bulk oscillators. If the inclusion consists only of a few atoms, its dielectric behaviour will deviate from the bulk one – for classical reasons as well as for quantum mechanical ones. The size effects that arise on this basis are called *intrinsic* size effects.

In terms of classical physics, intrinsic size effects occur when the number of surface atoms of the inclusion cannot be neglected with regard to the bulk atoms. Clearly, the smaller the inclusion (or the pore), the higher is the ratio between surface and bulk oscillators. As the behaviour of surface atoms is usually different from that of the bulk, we obtain an optical behaviour that depends on the size of the inclusions.

In application to very small void, the simple Lorentz–Lorenz equation (3.25) is sufficient to show that the void fraction may increase the refractive index of the mixture. Let us see how this may happen:

We start from a dense packed solid and create pores by removing a number of N_b atoms per unit volume. At the same time, we automatically create N_s surface states (for example dangling bonds), each with a polarizability β_s. The dense solid had a refractive index according to:

$$n_0^2 = 1 + \frac{N_0\beta}{1 - \dfrac{N_0\beta}{3}}$$

After removing the atoms, we have two types of oscillators, and the new refractive index may be written as:

$$n^2 = 1 + \frac{(N_0 - N_b)\,\beta + N_s\beta_s}{1 - \dfrac{(N_0 - N_b)\,\beta + N_s\beta_s}{3}}$$

It is obvious that the removal of atoms leads to a decrease of the refractive index, when

$$N_b\beta > N_s\beta_s \Leftrightarrow \frac{N_b}{N_s} > \frac{\beta_s}{\beta}$$

is fulfilled. On the other hand, N_b scales with the third power of the characteristic pore size l^3 (the volume), while N_s scales with l^2 (the surface). Therefore, for a refractive index increase with increasing density, we get the condition

$$l \propto \frac{N_b}{N_s} > \frac{\beta_s}{\beta}$$

So that the pores must not be too small. On the contrary, when the pores are smaller than the limit defined by the upper condition, the appearance of pores is expected to lead to an increase of the refractive index. Let us estimate the characteristic 'critical' pore radius where the derivative of the refractive index with respect to density changes its sign.

In order to get such an estimation, we have to make a few model assumptions. Let us assume, that the pores are spherical with the same radius R. That leads to:

$$N_b = \frac{4\pi}{3} N_p \frac{R^3}{a^3}\,; \quad N_s = 4\pi N_p \frac{R^2}{a^2}$$

With N_p – pore concentration and a – the interatomic spacing. When the polarizabilities are nearly equal ($\beta \approx \beta_s$), we get immediately:

$$R > 3a$$

When the interatomic spacing is approximately 0.2 nm, we should have at least a pore diameter of 1.2 nm to get a decrease in the refractive index due to the pores. This is consistent with the experimental findings mentioned above.

These considerations show, that the refractive index is not necessary an unambiguous function of the mass density. It really depends on the specifics of the pores whether or not the simple mixing formulae like (4.17) may find application. The general conclusion is, that for applying equations (4.17), the pore size l should be of an order so that the following condition is fulfilled:

$$1nm < l \ll \frac{\lambda}{n}$$

Of course, one may deal with intrinsic size effects by replacing the conventional dielectric function of the inclusion by a size-dependent *ad-hoc* dielectric function, but this does not change the principal conclusion of this chapter.

One last remark concerning the discussed subnanometer voids. Their effect is not to cause a small correction to the bulk refractive index. Instead, they may change the picture completely. Thus, crystalline silicon has a refractive index of nearly 3.45 in the NIR. Amorphous sputtered silicon films have been found to show refractive indices up to 4.2 in the same spectral region.

5 The Kramers–Kronig Relations

5.1 Derivation of the Kramers–Kronig Relations

This is the last chapter of the first part of this book, which deals with the classical theory of linear optical constants. The purpose of this chapter is to highlight some general analytical properties of the dielectric function, as they follow from the fundamental physical principle of causality.

Let us start with the case of dielectrics. From Sect. 2.5 we now, that their linear dielectric susceptibility may be written as:

$$\chi(\omega) = \int_0^\infty \kappa(\xi)\,\mathrm{e}^{\mathrm{i}\omega\xi}\,\mathrm{d}\xi = \int_{-\infty}^\infty \tilde{\kappa}(\xi)\,\mathrm{e}^{\mathrm{i}\omega\xi}\,\mathrm{d}\xi = \int_{-\infty}^\infty \tilde{\kappa}(\xi)\theta(\xi)\,\mathrm{e}^{\mathrm{i}\omega\xi}\,\mathrm{d}\xi$$

These identities directly follow from the principle of causality, which makes the response function invariant with respect to the multiplication with the step function $\theta(t)$. Let us now execute a Fourier transform according to:

$$\tilde{\kappa}(\xi) = \frac{1}{2\pi}\int_{-\infty}^\infty \chi(\omega)\,\mathrm{e}^{-\mathrm{i}\omega\xi}\,\mathrm{d}\omega$$

$$\theta(\xi) = \frac{1}{2\pi}\int_{-\infty}^\infty \Theta(\omega)\,\mathrm{e}^{-\mathrm{i}\omega\xi}\,\mathrm{d}\omega$$

That leads us to:

$$\begin{aligned}\chi(\omega) &= \frac{1}{(2\pi)^2}\int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}\omega\xi}\,\mathrm{d}\xi \int_{-\infty}^\infty \Theta(\omega_1)\,\mathrm{e}^{\mathrm{i}\omega_1\xi}\,\mathrm{d}\omega_1 \int_{-\infty}^\infty \chi(\omega_2)\,\mathrm{e}^{\mathrm{i}\omega_2\xi}\,\mathrm{d}\omega_2 \\ &= \frac{1}{(2\pi)^2}\int_{-\infty}^\infty\int_{-\infty}^\infty \Theta(\omega_1)\,\chi(\omega_2)\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2 \int_{-\infty}^\infty \mathrm{e}^{\mathrm{i}(\omega-\omega_1-\omega_2)\xi}\,\mathrm{d}\xi \\ &= \frac{1}{2\pi}\int_{-\infty}^\infty\int_{-\infty}^\infty \Theta(\omega_1)\,\chi(\omega_2)\,\delta(\omega-\omega_1-\omega_2)\,\mathrm{d}\omega_1\,\mathrm{d}\omega_2\end{aligned}$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \Theta(\omega - \omega_2)\, \chi(\omega_2)\, \mathrm{d}\omega_2 = \chi(\omega)$$

where the identity:

$$\int_{-\infty}^{+\infty} \mathrm{e}^{\mathrm{i}(\omega-\omega_1-\omega_2)\xi}\, \mathrm{d}\xi = 2\pi\delta(\omega - \omega_1 - \omega_2)$$

has been used with $\delta(x)$ – Dirac's delta-function. The Fourier-spectrum of the step function may be calculated according to:

$$\begin{aligned}
\Theta(\omega) &= \int_{-\infty}^{\infty} \theta(\xi)\, \mathrm{e}^{\mathrm{i}\omega\xi}\, \mathrm{d}\xi = \int_{0}^{\infty} \mathrm{e}^{\mathrm{i}\omega\xi}\, \mathrm{d}\xi = \lim_{T\to\infty} \int_{0}^{\infty} \mathrm{e}^{-\frac{\xi}{T}} \mathrm{e}^{\mathrm{i}\omega\xi}\, \mathrm{d}\xi \\
&= -\lim_{T\to\infty} \frac{1}{(-\frac{1}{T} + \mathrm{i}\omega)} = \lim_{T\to\infty} \frac{1}{-\mathrm{i}\omega + \frac{1}{T}} = \lim_{T\to\infty} \frac{T^{-1}}{T^{-2} + \omega^2} + \lim_{T\to\infty} \frac{\mathrm{i}\omega}{T^{-2} + \omega^2} \\
&= \pi\delta(\omega) + \frac{\mathrm{i}}{\omega}
\end{aligned}$$

so that one obtains:

$$\Theta(\omega - \omega_2) = \pi\delta(\omega - \omega_2) + \frac{\mathrm{i}}{\omega - \omega_2} \rightarrow$$

$$\chi(\omega) = \frac{1}{2\pi} VP \int_{-\infty}^{\infty} \left[\pi\delta(\omega - \omega_2) + \frac{\mathrm{i}}{\omega - \omega_2} \right] \chi(\omega_2)\, \mathrm{d}\omega_2$$

We thus come to the relationship:

$$\chi(\omega) = \frac{\mathrm{i}}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi(\omega_2)}{\omega - \omega_2}\, \mathrm{d}\omega_2$$

where 'VP' denotes Cauchy's principal value of the integral. Separating the real (χ') and imaginary (χ'') parts, we obtain the result:

$$\chi'(\omega) = -\frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi''(\omega_2)\, \mathrm{d}\omega_2}{\omega - \omega_2} = \frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi''(\omega_2)\, \mathrm{d}\omega_2}{\omega_2 - \omega}$$

$$\chi''(\omega) = \frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi'(\omega_2)\, \mathrm{d}\omega_2}{\omega - \omega_2} = -\frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi'(\omega_2)\, \mathrm{d}\omega_2}{\omega_2 - \omega}$$

In application to the real and imaginary parts of the dielectric function (ε' and ε'') we finally get the Kramers–Kronig Relations:

$$\varepsilon' = 1 + \frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\varepsilon''(\omega_2)\,\mathrm{d}\omega_2}{\omega_2 - \omega} \tag{5.1}$$

$$\varepsilon'' = -\frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\left[\varepsilon'(\omega_2) - 1\right]\mathrm{d}\omega_2}{\omega_2 - \omega} \tag{5.2}$$

The important conclusion is, that as a consequence of causality, the dispersion of the real and imaginary parts of the dielectric function are interconnected to each other via the integral transformations (5.1) and (5.2).

In the present form, these equations are valid only for dielectrics. The reason is clear: The integration interval involves the argument $\omega_2 = 0$, but as we have mentioned in Sect. 3.1.2, (2.25) cannot be used to describe the low-frequency behaviour of conductors. Instead, the approach (2.25) with respect to (3.13) has to be replaced by an approach like (3.15). That gives us the possibility to generalize equations (5.1) and (5.2) to the case of conductors. Indeed, the series (3.15) may be rewritten as:

$$\chi^{\mathrm{conductor}}(\omega) \equiv \mathrm{i}\,\frac{\sigma_{\mathrm{stat}}}{\varepsilon_0 \omega} + \chi^{\mathrm{opt}}(\omega)$$

Of course, χ^{opt} behaves 'regular' with respect to an expansion into a power series like (3.13). Therefore, for χ^{opt}, the Kramers–Kronig relations hold:

$$\chi'^{\mathrm{opt}}(\omega) = \frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi''^{\mathrm{opt}}(\omega_2)\,\mathrm{d}\omega_2}{\omega_2 - \omega}$$

$$\chi''^{\mathrm{opt}}(\omega) = -\frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi'^{\mathrm{opt}}(\omega_2)\,\mathrm{d}\omega_2}{\omega_2 - \omega}$$

Because the first term in (3.15) is purely imaginary, we have

$$\chi'(\omega) = \chi'^{\mathrm{opt}}(\omega)$$

So that, for a conductor, we find:

$$\begin{aligned} \chi''(\omega) &= \chi''^{\mathrm{opt}}(\omega) + \frac{\sigma_{\mathrm{stat}}}{\varepsilon_0 \omega} \rightarrow \\ \chi''(\omega) &= -\frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi'(\omega_2)\,\mathrm{d}\omega_2}{\omega_2 - \omega} + \frac{\sigma_{\mathrm{stat}}}{\varepsilon_0 \omega} \end{aligned} \tag{5.3}$$

and

$$\chi'(\omega) = \frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi''^{\text{opt}}(\omega_2)\,\mathrm{d}\omega_2}{\omega_2 - \omega} =$$

$$= \frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi''(\omega_2)\,\mathrm{d}\omega_2}{\omega_2 - \omega} - \frac{\sigma_{\text{stat}}}{\pi\varepsilon_0} \underbrace{VP \int_{-\infty}^{\infty} \frac{\mathrm{d}\omega_2}{\omega_2(\omega_2 - \omega)}}_{=0} \tag{5.4}$$

$$\rightarrow \chi'(\omega) = \frac{1}{\pi} VP \int_{-\infty}^{\infty} \frac{\chi''(\omega_2)\,\mathrm{d}\omega_2}{\omega_2 - \omega}$$

In (5.1)–(5.4), the susceptibility or the dielectric functions have to be defined for positive and negative frequencies. That does not cause any problems. According to (3.13) and (3.15), the imaginary part of the dielectric function has to be regarded as an odd function of the frequency, while the real part is an even one (For that deeper reason, the Sellmeier- and Cauchy – formulae in Sect. 4.4 contain only even powers of the wavelength or the wavenumber). Consequently, (5.3) and (5.4) may be rewritten in the more familiar form:

$$\varepsilon'(\omega) = 1 + \frac{2}{\pi} VP \int_{0}^{\infty} \frac{\varepsilon''(\omega_2)\omega_2\,\mathrm{d}\omega_2}{\omega_2^2 - \omega^2} \tag{5.5}$$

$$\varepsilon''(\omega) = -\frac{2\omega}{\pi} VP \int_{0}^{\infty} \frac{\left[\varepsilon'(\omega_2) - 1\right]}{\omega_2^2 - \omega^2}\,\mathrm{d}\omega_2 + \frac{\sigma_{\text{stat}}}{\varepsilon_0\omega} \tag{5.6}$$

5.2 Some Conclusions

Let us use this short section to present some useful relationships that immediately follow from equations (5.5) and (5.6). We start with the derivation of a simple dispersion formula (Wemple's dispersion formula), which is obtained for the refractive index dispersion in a frequency region well below the region of absorption. We postulate, that absorption (a non-zero imaginary part of the dielectric function) is restricted to a frequency range $[\omega_A, \omega_B]$. According to the mean-value-theorem, we have

$$\varepsilon'(\omega) = n^2(\omega) = 1 + \frac{2}{\pi} VP \int_{0}^{\infty} \frac{\varepsilon''(\omega_2)\,\omega_2\,\mathrm{d}\omega_2}{\omega_2^2 - \omega^2} = 1 + \frac{2}{\pi} \int_{\omega_A}^{\omega_B} \frac{\varepsilon''(\omega_2)\,\omega_2\,\mathrm{d}\omega_2}{\omega_2^2 - \omega^2} =$$

$$= 1 + (\omega_B - \omega_A)\frac{\varepsilon''(\bar{\omega})\,\bar{\omega}}{\bar{\omega}^2 - \omega^2} \equiv 1 + \frac{\text{const.} \times \bar{\omega}}{\bar{\omega}^2 - \omega^2} = n^2(\omega)$$

with $\quad \bar{\omega} \in [\omega_A, \omega_B];\ \omega \ll \omega_A < \omega_B;\ \varepsilon''(\omega) = 0.$

By structure, this dispersion formula is similar to the Sellmeier formula, obtained for one single oscillator. In fact, in our derivation, the full absorption structure has been replaced by a single oscillator centred at ω.

Another conclusion concerns the static dielectric constant for non-conductive materials. From (5.5), we get for $\omega = 0$:

$$\varepsilon_{\text{stat}} = 1 + \frac{2}{\pi} \int_0^\infty \frac{\varepsilon''(\omega_2)}{\omega_2} \, \mathrm{d}\omega_2 \tag{5.7}$$

So we see that the value of the static dielectric constant is directly connected to the high-frequency behaviour of Imε. In dielectrics, it will be always larger than one. On the other hand, for very high frequencies, we find from (5.5):

$$\varepsilon'(\omega) \,|_{\omega \to \infty} = 1 - \frac{2}{\pi \omega^2} \int_0^\infty \varepsilon''(\omega_2) \, \omega_2 \, \mathrm{d}\omega_2 \tag{5.8}$$

This approach is valid when the frequency is far above the frequency regions where absorption occurs (it does not work for Debye's equations, because of the slow descent of the imaginary part with increasing frequency). We see, that for very high frequencies, we still have to expect normal dispersion, but the refractive index is below the value one! This is a typical situation in the X-ray region.

From here, we may conclude that the real part of the dielectric function *must* show anomalous dispersion in the vicinity of absorption structures, no matter whether or not the oscillator model is applicable. Indeed, in the static case it is larger than the value one (5.7). As far as we have no absorption, it further increases with frequency according to Wemple's formula. When the frequency is well above the absorption structures, we find again normal dispersion, but the refractive index is below 1. Consequently, in the neighbourhood of absorption structures, the refractive index must decrease with frequency (as far as it is regarded to be a continuous function of the frequency).

Let us finally come to an important sum rule. In terms of a multioscillator model with M degrees of freedom, at sufficiently high frequencies the dielectric function of a material may be approximated by:

$$\varepsilon(\omega) = 1 - \frac{N_{free}}{\varepsilon_0} \, \frac{q^2}{m} \, \frac{1}{\omega^2 + 2\mathrm{i}\gamma\omega} + \sum_{j=1}^{M} \frac{\dfrac{N_j q^2}{\varepsilon_0 m}}{\tilde{\omega}_{0j}^2 - \omega^2 - 2\mathrm{i}\omega\gamma_j} \tag{5.9}$$

where N_j is the concentration of bound electrons that belong to the j-th degree of freedom. In the asymptotic case, we get

$$\varepsilon(\omega) \,|_{\omega \to \infty} \approx 1 - \frac{q^2}{\varepsilon_0 m} \left(N_{free} + \sum_{j=1}^{M} N_j \right) \times \frac{1}{\omega^2} = 1 - \frac{q^2}{\varepsilon_0 m \omega^2} N \tag{5.10}$$

where N is the full electron concentration. Comparing (5.8) and (5.10) leads us to the sum rule:

$$N = \frac{2\varepsilon_0 m}{\pi q^2} \int_0^\infty \varepsilon''(\omega_2)\, \omega_2 \, \mathrm{d}\omega_2 \tag{5.11}$$

Hence, the integral absorption is connected to the concentration of dipoles that cause the absorption. Rewriting (5.11) in terms of the optical constants, one immediately obtains:

$$N = \frac{2\varepsilon_0 mc}{\pi q^2} \int_0^\infty n(\omega)\, \alpha(\omega) \, \mathrm{d}\omega \tag{5.12}$$

Equation (5.12) is in the fundament of any quantitative spectroscopic analysis, where the integral absorption is measured in order to determine the concentration of any kind of absorption centres (molecules, impurities, and so on). Of course, in any practical application, one will always use a finite frequency interval where the integration in (5.12) is performed.

5.3 Resume from Chapters 2–5

5.3.1 Overview on Main Results

As mentioned in the introduction, the Chap. 2–5 together form the first part of this book and deal with the classical theory of linear optical constants. Let us shortly recall the main results we have obtained so far:

- For homogeneous, isotropic, and nonmagnetic materials, the linear optical constants (refractive index, absorption coefficient) are determined by the complex dielectric function of the material.
- As a result of causality, the dielectric function and the optical constants depend on the frequency (dispersion). The dispersion of the real and imaginary parts of the dielectric function are related to each other by integral transformations called the Kramers–Kronig relations.
- The dielectric functions of systems of permanent dipoles as well as of induced dipoles created by the oscillation of free and bound charge carriers have been explicitely derived. The results are expressed in terms of Debye's equations as well as the Drude- and Lorentz-formulae.
- In an assembly of oscillators, the width of the absorption line usually differs from that predicted by the model of a microscopic oscillator. This is caused by homogeneous and inhomogeneous line broadening effects.
- The optical constants of material mixtures depend on the mixing ratio as well as on the microstructure. Different optical mixing models have been derived, including the classical Maxwell Garnett theory, the Lorentz–Lorenz theory, and the effective medium approximation. On this basis, the optical behaviour of metal-dielectric-composites and of porous solids could be discussed. For materials with a columnar microstructure, the mixing model of Bragg and Pippard has been derived.

– The relation between microstructure and optical constants could be exemplified. Thus, the optical constants depend on the shape of possible inclusions. In simple cases, unambiguous relations between the refractive index and the mass density of a material may exist.

5.3.2 Problems

1. In (2.23), find out the angular frequency where $\mathrm{Im}\varepsilon$ reaches its maximum value. What is the value of $\mathrm{Re}\varepsilon$ at this frequency?
   Answer: $\omega = \tau^{-1}$; $\mathrm{Re}\varepsilon = 1 + \chi_{\mathrm{stat}}/2$.
2. From $\hat{n} = \sqrt{\varepsilon}$, find explicit expressions for the real and imaginary parts of the complex index of refraction as a function of $\mathrm{Re}\varepsilon$ and $\mathrm{Im}\varepsilon$.
   Answer:
   $n = \frac{1}{\sqrt{2}}\sqrt{\sqrt{(\mathrm{Re}\varepsilon)^2 + (\mathrm{Im}\varepsilon)^2} + \mathrm{Re}\varepsilon}$; $K = \frac{1}{\sqrt{2}}\sqrt{\sqrt{(\mathrm{Re}\varepsilon)^2 + (\mathrm{Im}\varepsilon)^2} - \mathrm{Re}\varepsilon}$
3. Find a general expression for the polarizability of a small sphere with the radius R embedded in vacuum in the quasistatic limit.
   Answer: $\beta(\omega) = 4\pi R^3 \dfrac{\varepsilon(\omega) - 1}{\varepsilon(\omega) + 2}$
4. Basing on the result of the third problem, assume a metallic sphere with a dielectric function given by the Drude function. Derive an explicit expression for $\beta(\omega)$.
   Answer: $\beta(\omega) = \dfrac{4\pi}{3} R^3 \dfrac{\omega_p^2}{\frac{\omega_p^2}{3} - \omega^2 - 2\mathrm{i}\omega\gamma}$; a resonance occurs at $\omega = \dfrac{\omega_p}{\sqrt{3}}$
5. Repeat problems 3 and 4 assuming an ellipsoid with the volume V, L being the relevant depolarisation factor
   Answer: $\beta(\omega) = V \dfrac{\varepsilon(\omega) - 1}{1 + [\varepsilon(\omega) - 1]\, L}$
   Drude-metal: $\beta(\omega) = V \dfrac{\omega_p^2}{L\omega_p^2 - \omega^2 - 2\mathrm{i}\omega\gamma}$; resonance at $\omega = \omega_p \sqrt{L}$
6. Calculate the so-called dielectric loss function for a Drude metal and for the oscillator model: The loss function is defined as $-\mathrm{Im}(1/\varepsilon)$.
   Answer: $-\mathrm{Im}\dfrac{1}{\varepsilon} = \dfrac{2\omega\gamma\omega_p^2}{\left(\tilde{\omega}_0^2 + \omega_p^2 - \omega^2\right)^2 + 4\omega^2\gamma^2}$

   A resonance appears at $\omega \approx \sqrt{\tilde{\omega}_0^2 + \omega_p^2}$. The resonance of the loss function is thus always blue-shifted with respect to that of the dielectric function. In metals, $\tilde{\omega}_0^2 = 0$, and therefore the resonance of the loss function is expected at the plasma frequency. The loss function will become important later in Chap. 6 when we discuss the reflectivity of metal surfaces.

7. Estimate the classical plasma frequency for different noble metals. From that, estimate the different resonance frequencies (and wavelength) derived in the problems 4–6.
8. In order to account for the effects of bound electrons in a metal, in classical physics, a merger of the Drude and the oscillator models may be applied, according to (3.1). Try to find out the effect of the bound electron resonances on the dielectric loss function, assuming that the bound oscillator resonance frequency is below the plasma frequency of the free electrons.
   Answers to 7 and 8 will be given in Chap. 6.
9. Check the Kramers–Kronig consistency of the real and imaginary parts of the dielectric functions given by Debye's model, the Drude formula, and the oscillator model.
10. Assume a porous columnar material with a packing density of 97%. At room temperature, the pores are assumed to be filled with water ($n_v = 1.33$). The room temperature refractive index of the porous material has been measured to be 2.10. Estimate the refractive index of the material at 100°C, when the pores are expected to be empty.
    Answer: $n = 2.07$
11. Make sure that the expression (5.12) yields the correct dimensionality of a particle concentration (m^{-3})!

Part II

Interface Reflection and Interference Phenomena in Thin Film Systems

6 Planar Interfaces

6.1 Transmission, Reflection, Absorption, and Scattering

6.1.1 Definitions

The previous Chaps. 2–5 discussed the classical treatment of linear optical constants, necessary for the calculation of any kind of optical spectrum. The Chaps. 6–9 form the second part of this book, and their main purpose is to deal with the calculation of the optical spectra of thin films and film stacks.

In terms of the philosophy claimed in Chap. 2, the calculation of any spectrum consists of two subtasks. The first one is to elaborate a theory for the wavelength behaviour of the optical material constants. This is what we have done in the first part of this book while restricting ourselves to classical models and linear optics. The second subtask is to calculate the propagation of electromagnetic waves in the given materials while considering the specific geometry of the system under investigation. We will not deal here with a general theory, but again restrict our attention to a particular case, namely on thin film spectra. Again, this will be a purely classical treatment. So that the material provided in the Chaps. 2–9 in sum shall enable the reader to calculate the linear optical properties of arbitrary thin film systems in terms of classical electrodynamics.

Let us start with some useful definitions. We regard a system as exemplified in Fig. 6.1.

Figure 6.1 shows an object (the sample) that is irradiated with light under a given angle of incidence. First of all, the incoming light has to penetrate the surface of the object to come into interaction with the material contained in the bulk of the sample. It is therefore clear that the optical properties of surfaces and interfaces will be of utmost importance for the optical behaviour of the whole system. It will be the subject of Sect. 6.2 to discuss this essential point. After interaction with the sample, light may leave the sample in several directions. From the phenomenological point of view, the light may either be

- *transmitted* through the sample (in a well-defined direction), or
- *specularly reflected* from the sample, or
- *diffusely scattered* at the sample surfaces or in its volume, or
- *absorbed* at the sample surfaces or in its volume.

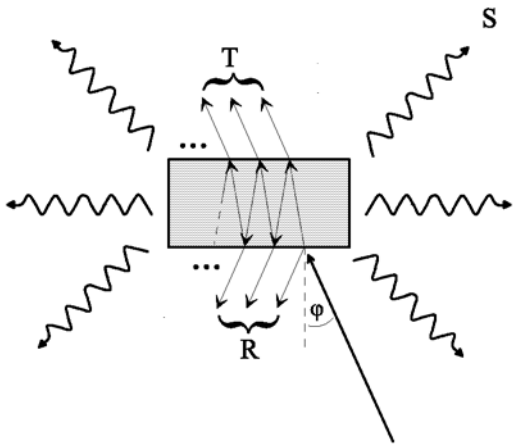

Fig. 6.1. To the definitions of T, R, and S. φ is the incidence angle

Let us for simplicity focus on the intensities of the signals. It is a common practice to define the *transmittance* T of the sample as the ratio of the intensity of the transmitted light I_T and that of the incoming light I_E:

$$T \equiv \frac{I_T}{I_E}$$

Accordingly, we define the *specular reflectance* R as the ratio of the specularly reflected intensity I_R and the incoming one:

$$R \equiv \frac{I_R}{I_E}$$

If we deal with a sample that does neither diffusely scatter nor absorb the irradiation, then the thus defined transmittance and reflectance must sum up to the value one – simply as a result of the energy conservation law. In practice, a certain fraction of the light intensity is diffusely scattered. That leads us to the definition of the optical *scatter* S as the ratio of the intensities of the light participating in scattering processes I_S and the incoming intensity:

$$S \equiv \frac{I_S}{I_E}$$

Analogously, we define the *absorptance* A as the ratio of the absorbed intensity I_A and the incoming one:

$$A \equiv \frac{I_A}{I_E}$$

In the presence of absorption and scatter, the energy conservation law may be written as:

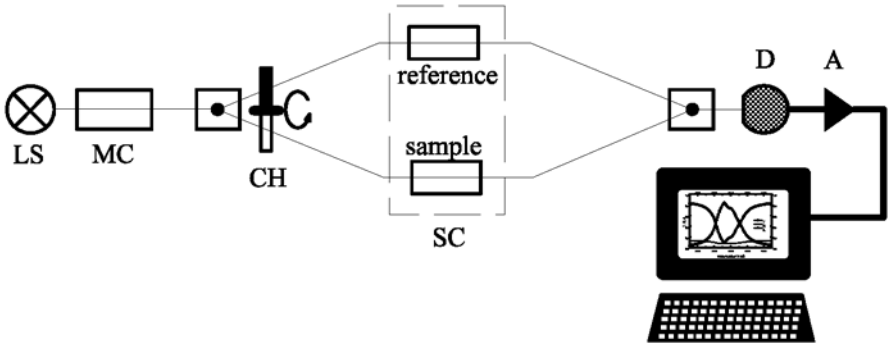

Fig. 6.2. Principal scheme of a double beam dispersive spectrophotometer; (LS) light source, (MC) monochromator, (CH) chopper, (SC) sample compartment, (D) detector, (A) amplifier

$$T + R + A + S = 1 \tag{6.1}$$

So that these four quantities are not independent from each other, and accurate knowledge of three of them allows the fourth to be immediately calculated. Nevertheless all four quantities T, R, S, and A may, in principal, be measured independently from each other. The algebraic sum of absorption and scatter is often called *optical loss* L:

$$L \equiv S + A = 1 - T - R \tag{6.2}$$

The values T, R, S, and A are characteristic for a sample in specific experimental conditions. That means, that both sample material and its geometry (including the experiment geometry) are responsible for the signal. Of course, all these values are additionally dependent on the wavelength of the light. But their wavelength dependence does *not* necessarily resemble that of the optical constants in a simple manner (especially in thin film samples).

6.1.2 Experimental Aspects

Concerning the measurements of T, R, A, and S, the transmittance is most simply to be measured. Today, transmission spectrophotometers belong to the commercially available standard equipment in many university and industrial labs. Typical spectrophotometers are either designed for the UV/VIS region ($\lambda > 185\,\text{nm}$; so-called UV/VIS-spectrometers, which often work with spectrally dispersive monochromators as shown schematically in Fig. 6.2) or for the MIR (so-called IR-spectrometers). IR spectrometers are produced today to an increasing extent as Fourier-Transform-spectrometers. This has led to the abbreviation FTIR (Fourier-Transform-Infrared). Compared to dispersive spectrophotometers, they allow a much faster spectra registration.

The NIR region is usually accessible in so-called UV/VIS/NIR - dispersive spectrometers, or as an optional upgrading of FTIR spectrometers. The latter type of spectrometer also allows to be upgraded for FIR-measurements.

In its standard version, a transmission spectrometer performs measurements of the transmittance with an accuracy of approximately 0.002 to 0.01, depending on the quality of the spectrometer and the wavelength range. Usually, a suitable specular reflectance attachment is optionally available, so that T and R may be measured.

Clearly, in any real situation, from the knowledge of two data (T and R) only the full optical loss may be determined utilizing (6.1). A discrimination between absorption and scatter losses is then impossible without additional model assumptions on the nature of the sample and their realization in refined mathematical spectra fitting procedures. An indication of surface scatter at the first sample surface may be drawn from the specular reflectance: If the first surface is rough, the specular reflectance gradually decreases down to zero with increasing frequency.

Another principal problem occurs in connection with the measurement of small loss values. As T and R are measured with a finite accuracy, the measurement of small losses (typically below 0.01) becomes impossible by this method. This is simply a consequence of the high underground signal, provided by the transmittance and the reflectance spectra. In such cases, one should rather directly measure the optical loss, and not conclude on it from T and R measurements.

Occasionally, corresponding attachments may also be combined with the above-mentioned spectrophotometers. Backscattering losses (back into medium 1) and forward scatter can be measured in so-called integrating sphere attachments, where the diffusely scattered light is collected and brought to the detector. These spheres are commercially provided for the NIR/VIS/UV spectral regions (coated with $BaSO_4$ or Spectralon) or for the MIR (coated with infragold). From the viewpoint of their size, these spheres reach from minispheres (a few centimeters in diameter) up to devices with more than one meter in diameter.

The accurate measurement of *absorption losses* bases on the idea, that the energy absorbed in the sample must either leave the sample (with a certain time delay) or enhance its temperature. In other words: The absorbed energy portion will participate in relaxation processes, and this is our chance to detect it. In order to detect very small absorption losses, absorption measurements are often accomplished with high incident light intensities, reliably supplied from laser sources.

It will depend on the nature of the sample and its environmental conditions (for example temperature), which of the relaxation channels works most rapidly. If radiative relaxation is fast enough, the fluorescence intensity allows us to conclude on the previously absorbed energy, and thus to determine the absorptance. This is what is done by the *fluorescence method*. If nonradiative relaxation is faster, then the absorbed energy will finally lead to

sample heating. As the temperature increase may be conveniently measured, the absorptance of the sample may be determined. Thus we have *calorimetric methods* of absorptance measurements.

Other absorption measurement techniques make use of the sample heating without direct temperature measurements. Thus, the *optoacustical measurements* detect the sound wave generated in a medium as a result of the absorption of pulsed light due to thermal expansion. Further methods detect the deformation of the sample surface, caused by thermal expansion due to light absorption. This deformation may be optically detected by the angular deflection of a weak probe beam. The corresponding method is called *Photothermal Deflection Spectroscopy* (PDS). Alternatively, the thermal expansion of the embedding medium surrounding the sample surface may be detected through its refractive index change. If the probe beam is of grazing incidence, the refractive index gradient in the vicinity of the heated surface leads to an angular deflection of the probe beam, which may be detected (*Mirage-effect*).

6.1.3 Remarks on the Absorbance Concept

There is a further concept describing optical sample properties that is especially popular in chemical physics – the so-called *absorbance* concept. The absorbance may be directly obtained by means of a spectrophotometer like shown in Fig. 6.2, placing the sample of interest into the sample beam, and a suitable reference sample into the reference beam. Then, the spectrophotometer measures the ratio between the sample transmittance T and that of the reference T_{ref}. From that, the absorbance may be calculated according to the definition:

$$\text{absorbance} \equiv -\lg \frac{T}{T_{\mathrm{ref}}}$$

It is immediately clear, that the thus defined absorbance is not an absolute measure of the sample properties, but depends on the optical properties of the (arbitrarily chosen) reference. That ambiguity is in clear contrast to the properties of the absorptance as defined before, and makes it difficult to apply the absorbance concept to quantitative analysis of the optical sample properties. One may overcome the problem leaving the reference beam blank, so that $T_{\mathrm{ref}} = 1$ for any wavelength. That leads to:

$$\text{absorbance} = -\lg T = -\lg(1 - A - S - R)$$

The thus defined absorbance is no further dependent on the properties of a reference sample. On the other hand, any non-vanishing reflectance, scatter, or absorption signal will give rise to a finite "absorbance" signal, so that, in fact, the measurement of a finite absorbance needs not to be connected with a physical absorption process. Hence, we will not use this absorbance concept

here, because it is only a mathematical transformation of the transmittance to a logarithmic scale and does not contain any new information. By the way, for the same reason it is misleading to refer to a transmission spectrophotometer as shown in Fig. 6.2 as an 'absorption spectrometer'.

Before concluding this section, let us make final remark concerning a specific version of the absorbance concept. In application to the spectroscopy of thin films, the supporter of the absorbance concept make use of the fact, that a thin film system is usually deposited on a thick substrate, which supplies the necessary mechanical support. It seems promising to define the thin film absorbance through the ratio of the sample (film on substrate) transmittance T and that of the uncoated substrate T_{sub}. This leads to:

$$\text{absorbance} \equiv -\lg \frac{T}{T_{\mathrm{sub}}}$$

But this re-definition is not helpful, on the contrary, it causes further confusion. Imagine an anti-reflection coating *without* absorption on a non-absorbing substrate (we will describe such systems in the following chapters in more detail). Such a sample has a *higher* transmittance than the bare substrate. Consequently, we get a negative absorbance (whatever that should mean), although there is no absorption at all – neither in the film nor in the substrate!

So that the absorbance concept may be convenient in the spectroscopy of liquids and gases in cells, but it should not be applied in solid state spectroscopy, and it lacks any use in thin film spectroscopy.

6.2 The Effect of Planar Interfaces: Fresnel's Formulae

The calculation of the transmittance and the reflectance of a thin film system belongs to the standard tasks in thin film spectroscopy. Impinging on a thin film system, the electromagnetic wave first comes into contact with the thin film surface. Therefore, the first step in understanding thin film spectra is to understand what happens with the electromagnetic wave at surfaces and interfaces.

That will lead us to a theoretical apparatus based on Fresnel's formulae. They are utmost important in thin film spectroscopy, and we will discuss them in full detail. But before deriving these equations, it is worth to make a remark concerning the history of these equations:

As it will be seen in the following, we will derive Fresnel's equations from Maxwell's electromagnetic theory of light. Fresnel couldn't make use of this theory for the simple reason, that he lived before Maxwell was born (Augustin Fresnel: 1788–1827; James Clerk Maxwell: 1831–1879). At his time, Fresnel obtained these equations from the elastic theory of ether, assuming transversal elastic waves impinging onto an interface. Of course, in our treatment we will use Maxwell's theoretical apparatus and imagine a plane electromagnetic

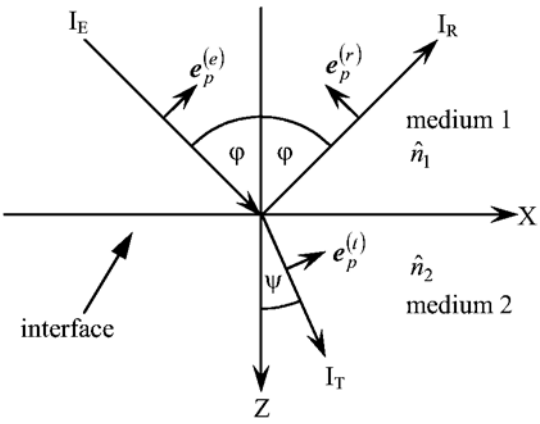

Fig. 6.3. Derivation of Fresnel's formulae

wave impinging on an interface with an angle of incidence φ. This situation is scetched in Fig. 6.3.

As seen from the figure, we assume an absolutely flat and sharp interface. Again, the media above and below the interface are regarded to be optically homogeneous, isotropic, and nonmagnetic. The (possibly complex) refractive index of the first medium is $\hat{n}_1$, and that of the second $\hat{n}_2$.

In the given geometry, it is natural to assume that there is one transmitted wave in the second medium, while we expect two waves in the first medium: the impinging wave and a reflected one. Let ψ be the angle between the interface normal and the propagation direction of the transmitted beam (angle of refraction). The incidence angle and the refractive angle are mutually connected by Snell's law of refraction, which is supposed to be known to the reader. In electromagnetic theory, it may be derived as a direct consequence of the requirement that the horizontal components of the wave vectors above and below the interface are identical. The latter requirement is a straightforward conclusion from Maxwell's boundary conditions for the parallel (to the interface) components of the electric field. Because the wave vectors may be complex, Snell's law may be formally written in the following way:

$$\frac{\sin\varphi}{\sin\psi} = \frac{\hat{n}_2}{\hat{n}_1} \tag{6.3}$$

For the same continuity reasons, the reflected wave forms the same angle with the interface normal as the impinging one.

It may happen that (6.3) yields a complex angle of refraction. That does not cause any problems. Indeed, let us assume the case of a thin film system. In this case, all interfaces are parallel to each other. Clearly, some of the interfaces may separate materials from each other that have complex indices of refraction, so that Snell's law forces us to work with complex angles of refraction. Their physical sense is quite simple. Let us assume, that the very

first (incident) medium is free of damping. Then, the refractive index is purely real, and of course the incidence angle is real as well. So the product $n_1 \sin\varphi$ is real. Due to Snell's law, $n_2 \sin\psi$ is then also real, even when the second medium has a complex index of refraction. Applying (6.3) subsequently to all interfaces, we will find that this conclusion is true for all media. On the other hand, the product of the refractive index and the sinus of the propagation angle is proportional to the tangential component of the wavevector. That component must remain real, because it was real in the incident medium and must be continuous. Consequently, when the refractive index becomes complex, the refractive angle must become complex as well.

On the other hand, the product $n_2 \cos\psi$ becomes complex when n_2 is complex. That means, that the normal (to the surface) component of the wavevector is complex. This is again a familiar result, because such a complex normal component of the wavevector is necessary to describe damping of the light intensity inside the film.

In summary, the complex refractive angles as introduced by (6.3) are a very convenient construction. In terms of the geometry as described in Fig. 6.3, they describe damping of the light intensity along the z-axis, while along the x-axis, the intensity doesn't change.

For oblique incidence of the impinging wave, we now define the plane of incidence as the plane containing the surface normal and the wave vector of the impinging light. In the geometry of Fig. 6.3, the plane of incidence is identical to the x–z-plane. The y-axis is directed in a manner that the x-, y-, and z-axes form a right handed Cartesian coordinate system. The interface between the media is thus identical with the x–y-plane.

The purpose of the following treatment is to derive equations that allow to calculate the transmittance and reflectance of a plane wave impinging on the given interface, supposed that the angle of incidence is known as well as the refractive indices of the materials and the polarization state of the wave. We start from Maxwells boundary conditions for the E- and H-fields at the plane interface. In the first medium, the full field strength appears as the sum of the fields of the impinging and reflected waves, while in the second medium, we only have one transmitted wave. As the tangential (to the interface) components of the E- and H-fields must be continuous, we may write:

$$\begin{aligned} E_x^{(e)} + E_x^{(r)} &= E_x^{(t)} \\ E_y^{(e)} + E_y^{(r)} &= E_y^{(t)} \end{aligned} \tag{6.4}$$

$$\begin{aligned} H_x^{(e)} + H_x^{(r)} &= H_x^{(t)} \\ H_y^{(e)} + H_y^{(r)} &= H_y^{(t)} \end{aligned} \tag{6.5}$$

where (e), (r), and (t) mark the incident, reflected, and transmitted waves.

Up to now, we have not yet discussed the polarization state of the waves. Let us focus on the electric field vector. As the electric field vector is normal to the propagation direction, it may be represented as the sum of two compo-

nents: one of them being parallel to the incidence plane (p-component), and one normal to the incidence plane (parallel to the y-axis – s-component). In the following, both of these special cases will be treated separately.

For the unambiguous characterization of the polarization state, we introduce unit vectors $\boldsymbol{e}_s$ and $\boldsymbol{e}_p$ in the following manner:

$$\begin{aligned}\boldsymbol{E}^{(e)} &= \boldsymbol{E}_s^{(e)} + \boldsymbol{E}_p^{(e)} \equiv E_s^{(e)}\boldsymbol{e}_s + E_p^{(e)}\boldsymbol{e}_p^{(e)}\\ \boldsymbol{E}^{(r)} &= \boldsymbol{E}_s^{(r)} + \boldsymbol{E}_p^{(r)} \equiv E_s^{(r)}\boldsymbol{e}_s + E_p^{(r)}\boldsymbol{e}_p^{(r)}\\ \boldsymbol{E}^{(t)} &= \boldsymbol{E}_s^{(t)} + \boldsymbol{E}_p^{(t)} \equiv E_s^{(t)}\boldsymbol{e}_s + E_p^{(t)}\boldsymbol{e}_p^{(t)}\end{aligned}$$

where the unit vectors for the s-component are all directed along the y-axis, while those for p-polarization are defined as indicated in Fig. 6.3. The electric field components fixed in (6.4) may now be expressed through the s- and p-components of the electric fields according to:

$$\begin{array}{lll} E_x^{(e)} = E_p^{(e)}\cos\varphi & E_x^{(r)} = -E_p^{(r)}\cos\varphi & E_x^{(t)} = E_p^{(t)}\cos\psi \\ E_y^{(e)} = E_s^{(e)} & E_y^{(r)} = E_s^{(r)} & E_y^{(t)} = E_s^{(t)} \\ E_z^{(e)} = -E_p^{(e)}\sin\varphi & E_z^{(r)} = -E_p^{(r)}\sin\varphi & E_z^{(t)} = -E_p^{(t)}\sin\psi \end{array}$$

Then, (6.4) may be rewritten as:

$$\cos\varphi\left(E_p^{(e)} - E_p^{(r)}\right) = E_p^{(t)}\cos\psi \tag{6.6}$$

$$E_s^{(e)} + E_s^{(r)} = E_s^{(t)} \tag{6.7}$$

Thus, the interface has a different effect on the s- and p-components of the impinging field. Only for normal incidence equations (6.6) and (6.7) are physically identical (one must take into consideration, that the vectors

$$\boldsymbol{e}_p^{(e)} \quad \text{and} \quad \boldsymbol{e}_p^{(r)}$$

are mutually antiparallel for normal incidence, while those for the s-polarization are always mutually parallel).

Let us describe the effect of the interface in terms of interface transmission and reflection coefficients for the electric fields as defined by:

$$\begin{aligned} r_p &= \frac{E_p^{(r)}}{E_p^{(e)}}, \qquad r_s = \frac{E_s^{(r)}}{E_s^{(e)}} \\ t_p &= \frac{E_p^{(t)}}{E_p^{(e)}}, \qquad t_s = \frac{E_s^{(t)}}{E_s^{(e)}} \end{aligned}$$

Of course, the two equations (6.6) and (6.7) are insufficient to calculate four unknown values t_s, t_p, r_s, r_p. We need two further equations, and these may

be obtained from (6.5). The only task is to rewrite the magnetic fields through electric fields. This will be done now.

Let us express $curl\boldsymbol{E}$ in Cartesian coordinates. That yields:

$$curl\boldsymbol{E} = \begin{vmatrix} \boldsymbol{e}_x & \boldsymbol{e}_y & \boldsymbol{e}_z \\ \dfrac{\partial}{\partial_x} & \dfrac{\partial}{\partial_y} & \dfrac{\partial}{\partial_z} \\ E_x & E_y & E_z \end{vmatrix}$$

where $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, and $\boldsymbol{e}_z$ are unit vectors directed along the coordinate axes. In a plane electromagnetic wave, we already know that $\boldsymbol{E}$ may be written as:

$$\boldsymbol{E} = \boldsymbol{E}_0 \, \mathrm{e}^{-\mathrm{i}(\omega t - \boldsymbol{kr})} \tag{6.8}$$

Combining the last two equations, we obtain:

$$curl\boldsymbol{E} = \mathrm{i} \begin{vmatrix} \boldsymbol{e}_x & \boldsymbol{e}_y & \boldsymbol{e}_z \\ k_x & k_y & k_z \\ E_x & E_y & E_z \end{vmatrix} = \mathrm{i}\boldsymbol{k} \times \boldsymbol{E}$$

In full analogy to (6.8), we have for the magnetic field:

$$\boldsymbol{H} = \boldsymbol{H}_0 \, \mathrm{e}^{-\mathrm{i}(\omega t - \boldsymbol{kr})}$$

From Maxwells equations we have:

$$curl\boldsymbol{E} = -\frac{\partial \boldsymbol{B}}{\partial \boldsymbol{t}}$$
$$\boldsymbol{B} = \mu_0 \boldsymbol{H}$$

The wavevector may be written as:

$$\boldsymbol{k} = \boldsymbol{e}\frac{\omega}{c}\hat{n}$$

so that we finally obtain:

$$\boldsymbol{k} \times \boldsymbol{E} = \mu_0 \omega \boldsymbol{H} \rightarrow \frac{\hat{n}}{\mu_0 c}\boldsymbol{e} \times \boldsymbol{E} = \boldsymbol{H} \tag{6.9}$$

The vector $\boldsymbol{e}$ without any sub- and superscripts denotes the unit vector along the propagation direction of the wave. For the impinging wave, it may be written as:

$$\boldsymbol{e} = \boldsymbol{e}_x \sin\varphi + \boldsymbol{e}_z \cos\varphi$$

From (6.9), the magnetic field of the impinging wave may be written as:

$$\boldsymbol{H} = \frac{\hat{n}}{\mu_0 c}\begin{vmatrix} \boldsymbol{e}_x & \boldsymbol{e}_y & \boldsymbol{e}_z \\ \sin\varphi & 0 & \cos\varphi \\ E_x^{(e)} & E_y^{(e)} & E_z^{(e)} \end{vmatrix}$$

$$= \frac{\hat{n}}{\mu_0 c}\left\{\boldsymbol{e}_x\left(-E_y^{(e)}\cos\varphi\right) + \boldsymbol{e}_y\left(E_x^{(e)}\cos\varphi - E_z^{(e)}\sin\varphi\right) + \boldsymbol{e}_z\left(E_y^{(e)}\sin\varphi\right)\right\}$$

That gives us the tangential components:

$$H_x^{(e)} = -\frac{\hat{n}_1}{\mu_0 c}E_s^{(e)}\cos\varphi \tag{6.10}$$

$$H_y^{(e)} = \frac{\hat{n}_1}{\mu_0 c}E_p^{(e)} \tag{6.11}$$

where we considered that the impinging wave moves in the first medium with the refractive index n_1. In the reflected wave, we have:

$$\boldsymbol{e} = \boldsymbol{e}_x\sin\varphi - \boldsymbol{e}_z\cos\varphi$$

It is not necessary to repeat the full calculation. Instead, we only have to replace $\cos\varphi$ by $-\cos\varphi$ to get the relevant expressions for the reflected fields. From (6.10) and (6.11) it is then obtained:

$$H_x^{(r)} = \frac{\hat{n}_1}{\mu_0 c}E_s^{(r)}\cos\varphi$$

$$H_y^{(r)} = \frac{\hat{n}_1}{\mu_0 c}E_p^{(r)}$$

In the transmitted wave, we have

$$\boldsymbol{e} = \boldsymbol{e}_x\sin\psi + \boldsymbol{e}_z\cos\psi$$

so that $\cos\varphi$ has to be replaced by $\cos\psi$ in (6.10) and (6.11). Moreover, as the transmitted wave propagates in the second medium, n_1 has to be replaced by n_2. That leads us to:

$$H_x^{(t)} = -\frac{\hat{n}_2}{\mu_0 c}E_s^{(t)}\cos\psi$$

$$H_y^{(t)} = \frac{\hat{n}_2}{\mu_0 c}E_p^{(t)}$$

The equations (6.5) may now be expressed in terms of the electric field as follows:

$$\hat{n}_1\cos\varphi\left(E_s^{(e)} - E_s^{(r)}\right) = \hat{n}_2\cos\psi\,E_s^{(t)} \tag{6.12}$$

$$\hat{n}_1\left(E_p^{(e)} + E_p^{(r)}\right) = \hat{n}_2 E_p^{(t)} \tag{6.13}$$

Equations (6.6), (6.7), (6.12) and (6.13) form a system of four equations that allow to calculate four unknown values r_s, r_p, t_s, and t_p. For p-polarization, we find from (6.6) and (6.13)

$$r_p = \frac{\hat{n}_2 \cos\varphi - \hat{n}_1 \cos\psi}{\hat{n}_2 \cos\varphi + \hat{n}_1 \cos\psi} \tag{6.14}$$

$$t_p = \frac{2\hat{n}_1 \cos\varphi}{\hat{n}_2 \cos\varphi + \hat{n}_1 \cos\psi} \tag{6.15}$$

For s-polarization, we have to use (6.7) and (6.12). That gives

$$r_s = \frac{\hat{n}_1 \cos\varphi - \hat{n}_2 \cos\psi}{\hat{n}_1 \cos\varphi + \hat{n}_2 \cos\psi} \tag{6.16}$$

$$t_s = \frac{2\hat{n}_1 \cos\varphi}{\hat{n}_1 \cos\varphi + \hat{n}_2 \cos\psi} \tag{6.17}$$

Equations (6.14)–(6.17) form the couple of Fresnel's equations. In the present form, they are valid only for isotropic and nonmagnetic materials.

The reflectance of a an interface may now be calculated when remembering that the intensity is proportional to the square of the modulus of the electric field amplitude, the dielectric function and the z-component of the velocity of light. That leads to:

$$R = |r|^2 \tag{6.18}$$

Accordingly, the transmittance through the interface is:

$$T = 1 - R = \frac{\mathrm{Re}\,(\hat{n}_2 \cos\psi)}{\mathrm{Re}\,(\hat{n}_1 \cos\varphi)} |t|^2 \tag{6.19}$$

In the simplest case of normal incidence and purely real refractive indices, we get the well-known formula:

$$R = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2 \tag{6.20}$$

Figure 6.4 demonstrates the principal angle dependence of the reflectivities of s- and p-polarized light.

As seen from Fig. 6.4, in the case of non-absorbing materials there appears a particular angle of incidence where the reflectivity of p-polarized light becomes zero. This is the so-called Brewsters angle. When a sample is illuminated at Brewsters angle of incidence, the reflected light will be linearly polarized along the s-direction, because the p-component is absent in the reflected light.

Brewsters angle may be simply calculated requiring $r_p = 0$. From that, one obtains:

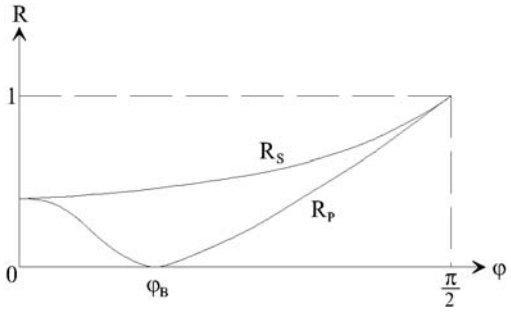

Fig. 6.4. Dependence of the reflectivity on the angle of incidence. The refractive indices are real

$$\tan \varphi_B = \frac{n_2}{n_1} \tag{6.21}$$

In this case,

$$\psi = \frac{\pi}{2} - \varphi \tag{6.22}$$

is also fulfilled. Consequently, when light impinges at Brewsters angle, the transmitted and reflected wavevectors are perpendicular to each other.

This allows to give a simple geometrical interpretation of the effect. Indeed, as seen from Fig. 6.5, at Brewsters angle the dipoles in the medium ($\boldsymbol{P}$) oscillate *parallel* to the propagation direction of the reflected light. However, an oscillating dipole does never irradiate into the direction of its vibration. Hence, when a surface is illuminated by p-polarized light at Brewster's angle, there are no dipoles that could contribute to the reflected wave. Consequently, there is no reflectance. Of course, such a situation is impossible for *s*-polarization.

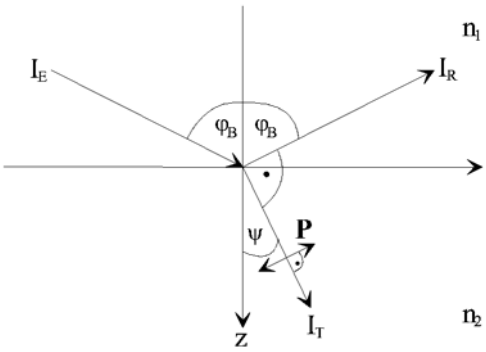

Fig. 6.5. Interpretation of Brewsters angle

6.3 Total Reflection of Light

6.3.1 Conditions of Total Reflection

As seen from (6.1), the maximum possible value of the reflectance is one. In this case all of the impinging light intensity is reflected by the sample (in this chapter, we do not regard the case of active laser media, where light is amplified). We therefore speak on *total* reflection of light. Let us see what kind of conditions must be fulfilled in order to obtain total light reflection at a single interface.

The mathematical formulation of the problem is simple. We have to demand, that

$$|r| = 1 \quad \text{or} \quad |r|^2 = 1 \tag{6.23}$$

is satisfied. We restrict our attention to the practically important situation, where the first medium has a purely real index of refraction, and analyse the case of s-polarization. For p-polarization, the same results will be obtained.

From (6.23), we have

$$1 = \left| \frac{n_1 \cos\varphi - \hat{n}_2 \cos\psi}{n_1 \cos\varphi + \hat{n}_2 \cos\psi} \right|$$

That is identical to:

$$\begin{aligned} &\left(n_1 \cos\varphi - \mathrm{Re}\left(\hat{n}_2 \cos\psi\right)\right)^2 + \left(\mathrm{Im}\left(\hat{n}_2 \cos\psi\right)\right)^2 \\ &= \left(n_1 \cos\varphi + \mathrm{Re}\left(\hat{n}_2 \cos\psi\right)\right)^2 + \left(\mathrm{Im}\left(\hat{n}_2 \cos\psi\right)\right)^2 \end{aligned}$$

These equations are automatically fulfilled when one of the participating refractive or extinction indices becomes infinitively large (*first* type of solution). At present, we will only fix this solution, its physical relevance shall be discussed later. Evaluating the upper equation, we come to the condition.

$$n_1 \cos\varphi \mathrm{Re}\left(\hat{n}_2 \cos\psi\right) = 0$$

Obviously, the next solutions are $n_1 = 0$ (*second* solution) and $\varphi = \pi/2$ (*third* solution). These are rather trivial solutions. In terms of (6.19), they correspond to a vanishing transmittance because of $t = 0$. In other words, the electric field is zero in the second medium.

The more interesting case is mathematically hidden in the remaining condition:

$$\mathrm{Re}\left(\hat{n}_2 \cos\psi\right) = 0$$

According to (6.19), there will be no transmittance through the interface in this case, no matter how large the Fresnel coefficient t (and the electric field) might be. Here, the reason is that there is no energy flux along the z-axis.

From Snell's law (6.3) we have

$$\cos\psi = \sqrt{1 - \frac{n_1^2}{\hat{n}_2^2}\sin^2\varphi}$$

so that the remaining solutions must suffice the condition:

$$\mathrm{Re}\sqrt{\hat{n}_2^2 - n_1^2\sin^2\varphi} = 0$$

In other words, the square root must be imaginary. For that, we must demand that the radicand is purely real but non-positive. On the other hand, the radicand may be written in the form:

$$n_2^2 - K_2^2 - n_1^2\sin^2\varphi + 2in_2K_2$$

This expression becomes real, when n_2 or K_2 are zero. Let us fix the case of $n_2 = 0$ as the *fourth* solution. In this case, the radicand is always zero or negative. If, on the contrary, $n_2 \neq 0$, we must require that $K_2 = 0$. In this case, we have a real radicand, which becomes non-positive when the additional condition

$$\sin\varphi \geq \frac{n_2}{n_1} \tag{6.24}$$

is fulfilled (*fifth* solution). As the light is incident from a medium with a real index of refraction, the incidence angle should also be regarded as real, and his sinus value cannot exceed 1. Therefore, the refractive index of the incident medium must be higher than that of the second one. This is the 'classical' condition of total internal reflection.

6.3.2 Discussion

Let us start with the first solution from Sect. 6.3.1. Total reflection is consequently obtained when the refractive index of one of the participating media becomes infinitively large by modulus. We are familiar with such a situation – it is obtained from Drude's formula when the frequency of the light approaches zero. Accordingly, Fresnel's equations predict a high reflection at metal surfaces in the long wavelength range, which is a quite reasonable result.

The second and fourth solutions deal with the situation, that the real part of one of the refractive indices is zero. The refractive index is then purely imaginary; accordingly, the dielectric function is real but negative. Again, such a situation makes sense in metals. When damping is negligible ($\omega \gg \gamma$), the dielectric function according to Drude's formula may be written as

$$\varepsilon = 1 - \frac{\omega_p^2}{\omega^2} \tag{6.25}$$

It is clearly negative for frequencies below the plasma frequency. In fact, Fresnel's equations yield the result that a plasma without damping reflects all light that has a frequency lower than the plasma frequency.

The third solution predicts high reflectance at grazing incidence – this is an intuitively clear result from everybody's daily experience.

Finally, the fifth solution represents the familiar case of 'total internal reflection'. It appears at the interface between a high-refractive and a low-refractive index material *without* damping, when the angle of incidence exceeds a critical angle defined by condition (6.24). At the critical angle, the refractive angle is 90°. Above the critical angle, the sinus of the refractive angle becomes larger than one, which is impossible in terms of a real angle of refraction. In fact, the angle of refraction becomes complex, while its real part is still 90°. It is correct to assume, that in this case there is no light transmitted into the second medium. When absorption is absent, all the light must consequently be reflected (therefore *total* internal reflection).

But there is completely another story when absorption is present. From Fresnel's equations it turns out that total reflection will only appear when $K = 0$ is fulfilled. Otherwise, one has to expect that the total reflection is attenuated, although there is still no real angle of refraction, as may be easily checked from Snell's law of refraction. So that, in this case, the light must be partly reflected, and partly absorbed, when penetrating into the second medium.

6.3.3 Attenuated Total Reflection ATR

It is worth mentioning that there is a spectroscopy modification that is entirely based on the specific conditions necessary for total internal reflection of light. We are speaking about attenuated total reflection spectroscopy (ATR). The idea is simple: when two absorption-free materials are in optical contact, one of them having a high refractive index, and the other one a low, so above some critical angle of incidence all light should be reflected. That means, a spectral scan of the reflectivity should give a straight 100% line, as long as the dispersion of the refractive indices does not violate condition (6.24). On the contrary, as soon as damping comes into play, total reflection will be destroyed, and the reflectivity will drop in a manner so that it resembles the spectral behaviour of the absorption coefficient of the medium. Consequently, these regions of 'attenuated' total reflection may give you an idea about the spectral behaviour of the absorption coefficient. For that reason attenuated total reflection spectroscopy has become an important tool mainly in physical chemistry where it may be used to identify substances by their thus identified 'absorption spectrum'. In practice the method is applied in a way that the light is incident from a high refractive index material that is assumed to be free of damping, and bounces onto the interface of a low-index material that has to be investigated.

Let us try to understand where the mentioned absorption of light may occur. In the absence of absorption, the electric field strength in the second medium may be written as:

$$\begin{aligned} E_2 &= E_{20}\,\mathrm{e}^{-\mathrm{i}(\omega t - k_x x - k_z z)}\,; \\ k_x &= k\sin\psi \\ k_z &= k\cos\psi \\ k &= \frac{\omega}{c} n_2 \end{aligned} \tag{6.26}$$

From Snell's law, we find:

$$\sin\psi = \frac{n_1}{n_2}\sin\varphi$$

$$\cos\psi = \sqrt{1 - \frac{n_1^2}{n_2^2}\sin^2\varphi} = \mathrm{i}\sqrt{\frac{n_1^2}{n_2^2}\sin^2\varphi - 1}$$

So that the wavefunction from (6.26) becomes:

$$\mathrm{e}^{-\mathrm{i}\left(\omega t - x n_1 \frac{\omega}{c}\sin\varphi\right)}\,\mathrm{e}^{-z\frac{\omega}{c}\sqrt{n_1^2\sin^2\varphi - n_2^2}} \tag{6.27}$$

Equation (6.27) describes a wave that travels along the interface in the x-direction, while its amplitude quickly damps in the direction into the film. Such waves are called *evanescent*. Due to this wave, the electric field extends into the second medium with a certain penetration depth. When the second medium is absorbing, it becomes clear that the wave is no longer evanescent and suffers absorption in the second medium.

In principal, ATR may work in broad spectral regions, but its main application field is the middle infrared spectral range. Then, the penetration depth is of the order of some microns. As the high refractive index material (the so-called ATR-crystals), germanium may be utilized ($n \approx 4.0$) or KRS5 ($n \approx 2.37$). In order to achieve a better sensitivity in practice, multiple bouncing of the light beam onto the interface is allowed, so that even weak absorption lines that appear in the attenuated total reflection spectrum are enhanced after a few bounces.

6.4 Metal Surfaces

6.4.1 Metallic Reflection

We are now in the position to discuss the particular reflection behaviour of metallic surfaces. From daily experience, everybody knows that metals have a high reflectance. This statement is valid in different spectral regions, and we will treat them separately.

Let us assume, that we deal with normal incidence of the light, and that the incidence medium is air. We will therefore assume, that the refractive index $n_1 = 1$, while the metal (second medium) has a refractive index n and an extinction coefficient K. The normal incidence reflectance of the metal surface according to Fresnel's equations is:

$$R = \frac{(n-1)^2 + K^2}{(n+1)^2 + K^2} \tag{6.28}$$

To determine the reflectance of the metal surface, we have to remember that the classical response of the free electrons is described by Drude's function (3.5) or (3.10). Let us start with the low-frequency limit, when $\omega \ll \gamma$ is fulfilled. From the asymptotic behaviour of Drude's function (3.15) we have:

$$\varepsilon \approx \mathrm{i}\,\frac{\sigma_{\mathrm{stat}}}{\varepsilon_0 \omega}$$

and therefore

$$n \approx K \approx \sqrt{\frac{\sigma_{\mathrm{stat}}}{2\varepsilon_0 \omega}}$$

These expressions may be used to evaluate the reflectance (6.28) up to the lowest order of ω. As the result, we obtain the Hagen-Rubens-Equation:

$$R|_{\omega \to 0} = 1 - \sqrt{\frac{8\varepsilon_0 \omega}{\sigma_{\mathrm{stat}}}} \tag{6.29}$$

The lower the frequency, and the higher the conductivity, the closer the reflectance will approach the 100%-value. That high reflection is caused by the large values of n and K and may be regarded as an example of the first solution of the total reflection condition as obtained in Sect. 6.3.1.

At higher frequencies, one could assume that Drude's function should be able to explain the experimentally established metallic reflectance. As an example, Fig. 6.6 shows the normal incidence reflectance of several noble metal surfaces. As we deal with bulk samples here, the transmittance is definitely equal to zero. The silver surface has a high reflectance over the whole VIS, and correspondingly, it does not appear in any colour when being illuminated with white light. On the contrary, gold absorbs the blue and violet, so it has an orange-yellow appearance when illuminated with white light. In copper, even green light is absorbed, thus causing the typical reddish appearance of this metal.

Let us now see what will be predicted from Drude's theory. In all of the mentioned metals, the plasma frequency of the free charge carriers corresponds to a wavenumber between $70000\,\mathrm{cm}^{-1}$ and $75000\,\mathrm{cm}^{-1}$, which is in the ultraviolet spectral region. The collision times between the electrons are also similar to each other – they range between 1×10^{-14}s and 4×10^{-14}s.

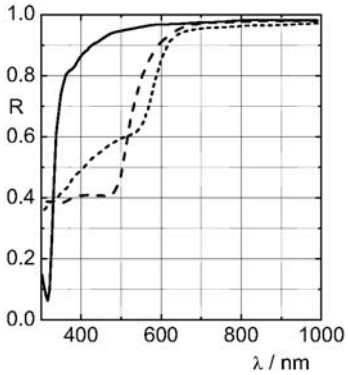

Fig. 6.6. Normal incidence reflectance of silver (*solid line*), gold (*dash*), and copper (*short dash*)

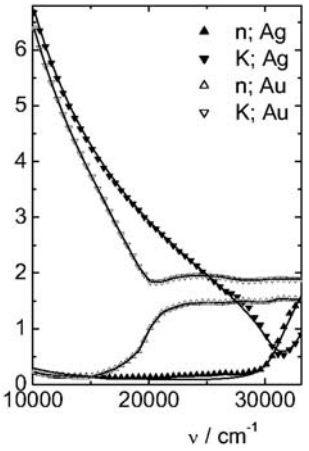

Fig. 6.7. Optical constants of silver and gold (*symbols*). The *solid lines* correspond to the fit by (5.9)

With these parameters, Drude's function predicts a reflectivity of approximately 99% over the whole VIS for all of these metals, which is not consistent with the experimental findings.

On the other hand, when the wavelength exceeds 650 nm, the reflectivities of the different metals are indeed close to each other, which suggests that Drude's function may at least be used to describe the optical response of metal surfaces at higher wavelength values.

In order to understand what is going on here, it makes sense to look at the optical constants of some of these metals. Figure 6.7 depicts the optical constants of gold and silver, which are consistent with the measured reflectance (symbols). In order to compare with Fig. 3.1, we have now chosen an abscissa that is proportional to the frequency, namely the wavenumber.

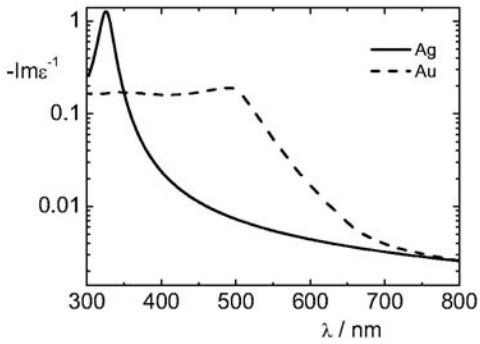

Fig. 6.8. Calculated dielectric loss function of silver and gold

As it is seen from Fig. 6.7 and 3.1, at higher wavelength, the behaviour of the optical constants of real metals indeed resembles that of Drude's function. In this region, we have optical constants satisfying $n \ll K$, which again results in high reflection, but now in terms of the fourth solution of the total reflection conditions. However, in the short wavelength region, serious deviations from Drude's theory occur, so that the response of free electrons alone seems insufficient to explain the observed behaviour.

The key for understanding the optical constants as shown in Fig. 6.7 is to include the response of the bound electrons into the description ((3.1) or (5.9)). This is still a completely classical approach, but it leads to an astonishingly good reproduction of the metal's optical constants.

The solid lines in Fig. 6.7 demonstrate the dispersion of the metal optical constants as they may be calculated by a merger of Drude's formula with a multioscillator model according to (5.9). To fit the optical constants of silver in the given spectral region, it is sufficient to introduce one oscillator that accounts for the bound electrons. In the case of gold, five oscillators have been used.

So that we can state, that Drude's function is suitable for describing the optical constants of metals at a sufficiently long wavelength. When the wavelength becomes too short, the response of bound electrons must be considered, which is done in classics by means of Lorentzian oscillators.

To conclude this subchapter, let us return to the reflectance curves from Fig. 6.7. As already mentioned, at a high wavelength there is generally a high reflection. It starts to drop when the wavelength becomes shorter than a certain threshold value, which is material dependent. When comparing with the optical constants from Fig. 6.7, we see that the drop in the reflectivity is accompanied by characteristic features in the optical constants. The question is: Is there any simple function of the optical constants that may predict the mentioned "threshold" wavelength?

Fortunately, there is. It is the so-called dielectric loss function, as introduced in the Sect. 5.3 (Problem 6). Figure 6.8 shows the loss function corresponding to the optical constants depicted in Fig. 6.7.

A comparison between Figs. 6.6 and 6.8 demonstrates, that the loss function resembles the spectral behaviour of the reflectance, while the peaks in the loss function are spectrally close to valleys in the reflection curve.

6.4.2 Extended Detail: Propagating Surface Plasmons

We will conclude Chap. 6 with two more complicated examples, which deal with surprising optical effects that may be observed at interfaces and may at least partially be described within the theoretical framework of Fresnels equations. In this subsection 6.4.2, we will shortly derive the dispersion relation of propagating surface plasmons at metal surfaces, which are utmost important in applied optical surface spectroscopy. The second example (Sect. 6.5) is dedicated to the effects of giant birefringent optics occurring at the interfaces between optically anisotropic materials.

We have already dealt with surface plasmons in small spheres (Sect. 4.5.3). In that case, the plasmon has been excited at a spherical surface. We will now consider the case of plane surfaces.

Let us start again from a thought experiment. Imagine, that we are looking for a spectroscopic method that is extremely sensitive to any effects located near the interface. One would naturally choose a geometry where the electric field at the interface would be large. Correspondingly, one should require that the energy of the impinging wave is neither transmitted through the interface, nor reflected from the interface, but accumulated 'in' the interface region. Of course, as seen from (6.19), that may never happen as long as we deal with one single ideal interface. However, let us now assume at the first time in this chapter, that we do not deal with a single surface, but with a *thin film*. This situation is sketched in Fig. 6.9.

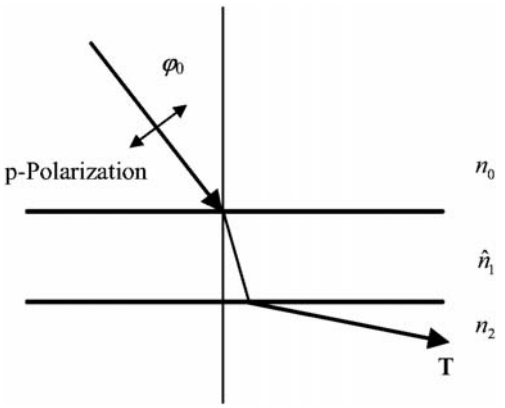

Fig. 6.9. Assumed thin film system

Let us further assume, that the light impinges from the 'zeros' medium with an angle of incidence φ_0. The refractive index n_0 shall be purely real as well as the refractive index n_2 of the last medium. We want to find the physical conditions that lead to a high electric field at the interface between the media 1 and 2. From Snell's law, we have

$$n_0 \sin\varphi_0 = \hat{n}_1 \sin\varphi_1 = n_2 \sin\varphi_2$$

According to our general idea, we require that the transmittance of the whole system vanishes. Additionally, we try to achieve a vanishing reflectance at the interface between the media 1 and 2. To do so, we assume p-polarization and tune the angle φ_1 to the Brewster's angle. Both conditions together may be written as:

$$\left\{ \begin{array}{l} T = 0 \Rightarrow 1 > \sin\varphi_0 \geq \dfrac{n_2}{n_0} \\ \\ R_{p,12} = 0 \Rightarrow \tan\varphi_1 = \dfrac{n_2}{\hat{n}_1} \end{array} \right\}$$

Because the angles φ_1 and φ_0 are mutually connected by Snell's law, one may show that the upper conditions are fulfilled when (6.30) is satisfied:

$$1 > \sin^2\varphi_0 = \frac{n_2^2}{n_0^2}\,\frac{\hat{n}_1^2}{\hat{n}_1^2 + n_2^2} \geq \frac{n_2^2}{n_0^2} \tag{6.30}$$

Let us look at condition (6.30) in more detail. It has been obtained assuming the geometry of Fig. 6.9 and requiring, that the transmittance of the whole system vanishes as well as the reflectance at the second interface. Condition (6.30) states that this is principally possible, when several additional requirements are fulfilled. First of all, it is obvious that the incidence medium must have a higher index of refraction than the last one. On the contrary, medium 1 must have a purely imaginary index of refraction. This is clearly impossible, but we know, that several metals have refractive indices with an imaginary part that is much higher than the real one, thus coming close to what would be necessary. So, we may assume that the film material 1 is a metal, for example silver. In order to fulfil the right-hand inequality in (6.30), we must further demand:

$$\hat{n}_1^2 < -n_2^2 \Rightarrow \varepsilon_1 < -\varepsilon_2 \tag{6.31a}$$

The left-hand inequality in (6.30) restricts the index of refraction of the incidence medium. Indeed, we must require:

$$n_0^2 > \frac{\hat{n}_1^2 n_2^2}{\hat{n}_1^2 + n_2^2} \tag{6.31b}$$

Then, (6.30) defines an incidence angle (the resonance angle) where the reflectance of p-polarized light of the whole system should be minimized, while the transmittance is definitely zero.

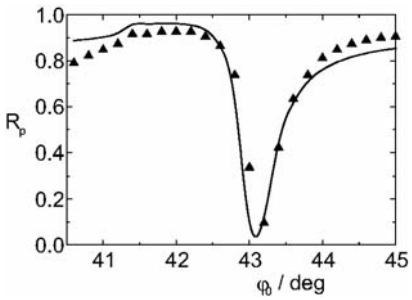

Fig. 6.10. Reflectance of p-polarized light ($\lambda = 632.8\,\mathrm{nm}$) of the system glass/silver/air as a function of the incidence angle

Let us look at an example. Figure 6.10 shows you an experimental angular reflectance scan in a thin film system, where a 50 nm thick silver film has been deposited on quartz glass. The light ($\lambda = 632.8\,\mathrm{nm}$) was incident from the quartz side. We see, that at an angle of approximately 43 deg, the reflectivity has a sharp gap. As the transmittance is zero, the light intensity is indeed absorbed in the system. But we do not know, where.

Fortunately, there is a couple of strong indications that the effect is really located at the surface between silver (medium 1) and air (medium 2). Let us utilize (6.30) to estimate the resonance angle φ_0 predicted by our theory. Equation (6.30) is now only an approximation, because, in fact, the refractive index of silver is not purely imaginary. But the real part is small, as seen from Fig. 6.7, so that we assume $\hat{n}_1^2 \approx -16$ at the given wavelength. The refractive index of air is 1, and that of glass close to 1.5. That leads to a theoretical resonance angle of 43.5 deg, quite close to the value obtained from the experimental data given in Fig. 6.10.

Additionally, the solid line in Fig. 6.10 shows the result of a theoretical fit of the experimental data. The theoretical apparatus for the fit is derived in Chap. 7, but it is worth mentioning that the given fit corresponds to a silver thickness of 46.2 nm, which is in good agreement to the intentional value of 50 nm.

Let us now understand what happens at the interface between silver and air. The horizontal component of the wavevector k_x may be written as:

$$k_x = \frac{\omega}{c} n_0 \sin\varphi_0 = \frac{\omega}{c}\sqrt{\frac{n_2^2 \hat{n}_1^2}{\hat{n}_1^2 + n_2^2}} = \frac{\omega}{c}\sqrt{\frac{\varepsilon_1 \varepsilon_2}{\varepsilon_1 + \varepsilon_2}} \tag{6.32}$$

When (6.31a) is satisfied, expression (6.32) yields a real value for the horizontal component of the wavevector. Let us look at the normal components. In difference to the horizontal ones, they are different in each material. We have:

$$
\begin{aligned}
k_{z,0} &= \frac{\omega}{c} n_0 \cos\varphi_0 \rightarrow \text{real} \\
k_{z,1} &= \frac{\omega}{c} \hat{n}_1 \cos\varphi_1 = \frac{\omega}{c} \hat{n}_1 \sqrt{1 - \sin^2\varphi_1} = \frac{\omega}{c} \sqrt{\varepsilon_1} \sqrt{\frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2}} \\
&= \pm \mathrm{i} \frac{\omega}{c} \sqrt{\left| \frac{\varepsilon_1^2}{\varepsilon_1 + \varepsilon_2} \right|} \rightarrow \text{imaginary} \\
k_{z,2} &= \frac{\omega}{c} n_2 \cos\varphi_2 = \frac{\omega}{c} n_2 \sqrt{1 - \sin^2\varphi_2} \\
&= \pm \mathrm{i} \frac{\omega}{c} \sqrt{\left| \frac{\varepsilon_2^2}{\varepsilon_1 + \varepsilon_2} \right|} \rightarrow \text{imaginary}
\end{aligned}
\tag{6.33}
$$

As expected, in the incident medium we have a propagating wave, while in media 1 and 2, no wave propagation is possible (in fact, we have again total reflection conditions here).

Accordingly, the electric fields in the media 1 and 2 may be written as:

$$
E_1 = E_{10}\, \mathrm{e}^{-\mathrm{i}\left(\omega t - \frac{\omega}{c}\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}x\right)}\, \mathrm{e}^{\mp\left(\frac{\omega}{c}\right)\sqrt{\left|\frac{\varepsilon_1^2}{\varepsilon_1+\varepsilon_2}\right|}z}
$$

$$
E_2 = E_{20}\, \mathrm{e}^{-\mathrm{i}\left(\omega t - \frac{\omega}{c}\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}x\right)}\, \mathrm{e}^{\mp\left(\frac{\omega}{c}\right)\sqrt{\left|\frac{\varepsilon_2^2}{\varepsilon_1+\varepsilon_2}\right|}z}
$$

It makes no sense to assume, that we have an exponentially increasing electric field at infinity in medium 2, so that we choose the *descending* electric field here. For continuity reasons at the interface, in medium 1 we choose the *ascending* solution. So that we obtain finally:

$$
\begin{aligned}
E_1 &= E_{10}\, \mathrm{e}^{-\mathrm{i}\left(\omega t - \frac{\omega}{c}\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}x\right)}\, \mathrm{e}^{\left(\frac{\omega}{c}\right)\sqrt{\left|\frac{\varepsilon_1^2}{\varepsilon_1+\varepsilon_2}\right|}z} \\
E_2 &= E_{20}\, \mathrm{e}^{-\mathrm{i}\left(\omega t - \frac{\omega}{c}\sqrt{\frac{\varepsilon_1\varepsilon_2}{\varepsilon_1+\varepsilon_2}}x\right)}\, \mathrm{e}^{-\left(\frac{\omega}{c}\right)\sqrt{\left|\frac{\varepsilon_2^2}{\varepsilon_1+\varepsilon_2}\right|}z}
\end{aligned}
\tag{6.34}
$$

Expressions (6.34) describe an evanescent wave travelling along the surface, while descending by amplitude into the media 1 and 2. Consequently, the electric field is mainly concentrated at the interface, and this is indeed what we wanted to achieve. The travelling evanescent wave excites the movement of free electrons at the metal (medium 1) surface, and their collective oscillation is called a propagating surface plasmon. The couple of equations (6.32) and (6.33) forms the dispersion law of the surface plasmons (any $k(\omega)$-dependence is called a dispersion law). The penetration depth of the evanescent field becomes infinitively small, when $\varepsilon_1 \rightarrow -\varepsilon_2$, and in this case the field is most effectively confined to the interface region.

The excitation of propagating surface plasmons is an outstandingly effective experimental tool in surface spectroscopy. Indeed, when one has to

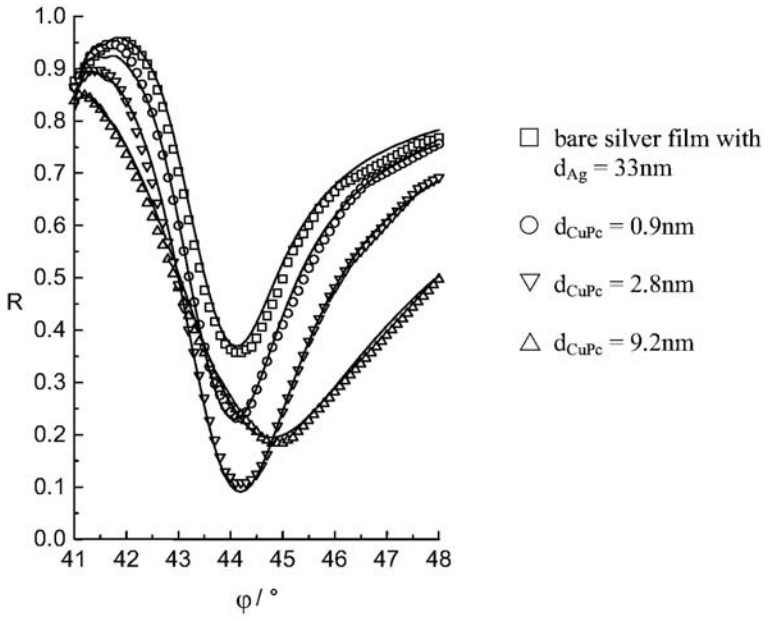

Fig. 6.11. Fit of reflectance curves of a system glass/silver/adsorbate CuPc/air, obtained at 560 nm wavelength. The silver thickness d_{Ag} is 33 nm for all samples

detect weak absorption centres at a metal surface, any background signal resulting from the bulk will be highly disturbing. On the contrary, when a strong electric field is confined in the surface region, the bulk background signal may be strongly reduced in significance.

In practice, an ultrathin but absorbing adsorbate layer at the silver-air-interface will significantly change the spectral features occurring in the reflectance of the system. To give an idea on the sensitivity of the method, Fig. 6.11 presents reflectance spectra obtained from a 33 nm silver film on glass with ultrathin organic adsorbate layers. As the adsorbate, a blue-coloured organic dye (copperphthalocyanine CuPc) has been choosen. The figure demonstrates, that even ultrathin adsorbate layers with a thickness down to 0.9 nm are easily detectable.

Again, the full line in Fig. 6.11 correspond to theoretical fits obtained from equations that will be derived in Chap. 8.

But what is the reason for the high sensitivity? The point is, that the electric field strength at the interface may be extremely high, even when no light intensity is transmitted through the interface. In the present geometry, the local electric field strength amplitude at the interface may exceed that of the impinging wave for several orders of magnitude. Again, some patience will be of use, because for an accurate calculation we would need the theoretical apparatus, which will be derived in Chap. 7.

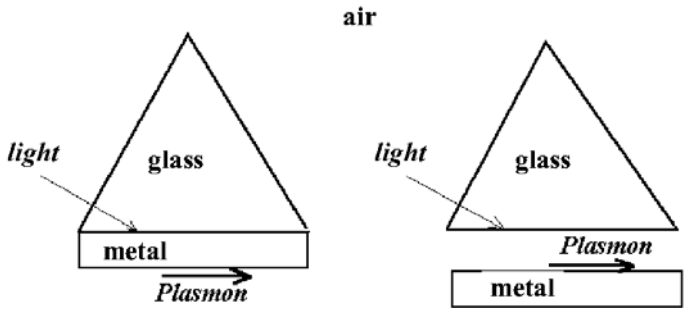

Fig. 6.12. Prism coupler geometries: **left**: Kretschmar–Raether geometry, **right**: Otto geometry

In practice, propagating plasmons at the metal-air interface are most easily excited by means of so-called prism couplers. In Fig. 6.12, two experimental geometries are presented, which are commonly in use for the optical excitation of propagating surface plasmons. The arrangement shown in Fig. 6.9 corresponds to the Kretschmar–Raether geometry.

6.5 Extended Detail: Anisotropic Materials

6.5.1 Interface Reflection Between an Isotropic and an Anisotropic Material

Concluding the treatment of interfaces, let us briefly discuss the generalization of Fresnel's equations to the interface between an optically isotropic and an optically anisotropic material. In principle, this subject is beyond the frames of this book as they have been defined in the introduction. Nevertheless, optically anisotropic films may become important for specific future applications, and therefore it is worth to give a brief introduction to this field.

This section will not deal with derivations of formulae. Instead, some knowledge on light propagation in anisotropic media will be presumed, and the corresponding modifications that occur in Fresnel's reflection coefficients will be rather guessed than derived. The material given in Sect. 6.5 will not be necessary to understand the following chapters, so that the section may be skipped by the reader as well.

We restrict our discussion to the special case of uniaxial and nonabsorbing anisotropic materials. In this case, instead of (2.9), the electrical displacement vector will be given by:

$$\boldsymbol{D} = \varepsilon_0 \boldsymbol{\varepsilon} \boldsymbol{E}$$

Where $\boldsymbol{\varepsilon}$ is now a symmetric tensor. In a suitable coordinate system (which is not necessarily identical to the coordinate system used in this chapter so far!!), it may be written in the diagonalized form:

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & 0 & 0 \\ 0 & \varepsilon_{yy} & 0 \\ 0 & 0 & \varepsilon_{zz} \end{pmatrix}$$

Let us assume an uniaxial material with an optical axis parallel to the z-axis. In this case,

$$\varepsilon_{xx} = \varepsilon_{yy} \neq \varepsilon_{zz}$$

This leads to a completely different behaviour of an electromagnetic wave. Indeed, let us regard the wave (2.2). When the wave travels along the z-axis, the displacement vector $\boldsymbol{D}$ may contain an x- and a y-component, both of them 'feeling' the same dielectric function ε_{xx}. This wave travels with a given phase velocity determined by ε_{xx}, no matter how the light is polarized.

Let us now discuss the case of a wave propagating into a direction, that forms an angle ϑ to the z-axis. Let us for simplicity assume, that the $\boldsymbol{k}$-vector lies in the x–z-plane. For vertical polarization ($\boldsymbol{D}$ parallel to the y-axis) only ε_{yy} is relevant, independent of the angle ϑ. Therefore, such a wave travels with a phase velocity that does not depend on ϑ. It is called an *ordinary* wave. On the contrary, for the other polarization ($\boldsymbol{D}$ is in the x–z-plane), it 'feels' a superposition of the different dielectric functions ε_{xx} and ε_{zz}, while their relative weights depend on the angle ϑ. The phase velocity of this wave will depend on ϑ. Due to this quite unusual behaviour, such a wave is called an *extraordinary* one. Any arbitrarily polarized wave that impinges onto the surface of an uniaxially anisotropic material will split off into the ordinary and extraordinary waves, a behaviour that is called optical *birefringence*.

Let us regard some useful equations. We rewrite the condition of uniaxiality in the following manner:

$$\varepsilon_{xx} = \varepsilon_{yy} \equiv \varepsilon_{\perp} \neq \varepsilon_{zz} \equiv \varepsilon_{\parallel}$$

The ordinary wave travels with the so-called ordinary refractive index defined as:

$$n_o \equiv \sqrt{\varepsilon_{\perp}}$$

Without derivation, we present the expression for the angle-dependent refractive index valid for the extraordinary wave n_a:

$$n_a = n_a(\vartheta) = \frac{n_o n_e}{\sqrt{n_e^2 \cos^2\vartheta + n_o^2 \sin^2\vartheta}}$$

$$n_e \equiv \sqrt{\varepsilon_{\parallel}}$$

Where n_e is the so-called extraordinary refractive index. Obviously, for a given ϑ, the relative weights of the ordinary and extraordinary refractive

indices are given by $\sin\vartheta$ and $\cos\vartheta$, respectively. n_o and n_e form the pair of principal refractive indices of an uniaxial material.

This is the key to our simplified treatment of Fresnel's reflection coefficients in the anisotropic case. It will enable us to guess the correct expressions. Let us rewrite Fresnel's equations valid for the isotropic case in the following manner:

$$r_s = \frac{n_1\cos\varphi - \sqrt{n_2^2 - n_1^2\sin^2\varphi}}{n_1\cos\varphi + \sqrt{n_2^2 - n_1^2\sin^2\varphi}}$$
$$r_p = \frac{n_2\cos\varphi - n_1\sqrt{1 - \frac{n_1^2}{n_2^2}\sin^2\varphi}}{n_2\cos\varphi + n_1\sqrt{1 - \frac{n_1^2}{n_2^2}\sin^2\varphi}} \tag{6.35}$$

Again, we will now assume, that medium 1 is isotropic, while medium 2 is anisotropic. The problem is, that different orientations of the optical axis with respect to the incidence plane are possible. Let us regard three special cases:

a) optical axis perpendicular to the surface

In this case, s-polarized light always senses the ordinary refractive index. Therefore, in r_s, one only has to replace n_2 by the ordinary index $n_{2\mathrm{o}}$. For p-polarization, a superposition of $n_{2\mathrm{o}}$ and the extraordinary index $n_{2\mathrm{e}}$ is sensed. The higher the incidence angle, the higher the contribution of $n_{2\mathrm{e}}$. Therefore, in the pre-factor of $\cos\varphi$, we replace n_2 by $n_{2\mathrm{o}}$. On the contrary, in the prefactor of $\sin\varphi$, n_2 will be replaced by $n_{2\mathrm{e}}$. We obtain:

$$r_s = \frac{n_1\cos\varphi - \sqrt{n_{2o}^2 - n_1^2\sin^2\varphi}}{n_1\cos\varphi + \sqrt{n_{2o}^2 - n_1^2\sin^2\varphi}}$$
$$r_p = \frac{n_{2o}\cos\varphi - n_1\sqrt{1 - \frac{n_1^2}{n_{2e}^2}\sin^2\varphi}}{n_{2o}\cos\varphi + n_1\sqrt{1 - \frac{n_1^2}{n_{2e}^2}\sin^2\varphi}} \tag{6.35a}$$

b) optical axis parallel to both the incidence plane and the surface

Again, s-polarized light always senses the ordinary refractive index. Therefore, for r_s, (6.35a) remains valid. For p-polarization, again a superposition of $n_{2\mathrm{o}}$ and the extraordinary index $n_{2\mathrm{e}}$ is sensed. In difference to the previous case, $n_{2\mathrm{e}}$ and $n_{2\mathrm{o}}$ interchange their roles. We find:

$$r_s = \frac{n_1 \cos\varphi - \sqrt{n_{2o}^2 - n_1^2 \sin^2\varphi}}{n_1 \cos\varphi + \sqrt{n_{2o}^2 - n_1^2 \sin^2\varphi}}$$
$$r_p = \frac{n_{2e} \cos\varphi - n_1 \sqrt{1 - \frac{n_1^2}{n_{2o}^2} \sin^2\varphi}}{n_{2e} \cos\varphi + n_1 \sqrt{1 - \frac{n_1^2}{n_{2o}^2} \sin^2\varphi}} \tag{6.35b}$$

c) optical axis perpendicular to the incidence plane

This is the simplest case. No matter what incidence angle is applied, the s-polarization always senses the extraordinary index, while the p-polarization always senses the ordinary one. From (6.35) it is therefore obtained:

$$r_s = \frac{n_1 \cos\varphi - \sqrt{n_{2e}^2 - n_1^2 \sin^2\varphi}}{n_1 \cos\varphi + \sqrt{n_{2e}^2 - n_1^2 \sin^2\varphi}}$$
$$r_p = \frac{n_{2o} \cos\varphi - n_1 \sqrt{1 - \frac{n_1^2}{n_{2o}^2} \sin^2\varphi}}{n_{2o} \cos\varphi + n_1 \sqrt{1 - \frac{n_1^2}{n_{2o}^2} \sin^2\varphi}} \tag{6.35c}$$

Equations (6.35a–c) represent important special cases of Fresnel's reflection coefficients for the interfaces between optically isotropic and anisotropic media.

6.5.2 Giant Birefringent Optics

Quite interesting optical effects may arise at the interfaces between isotropic and anisotropic materials. They form the field of the so-called Giant Birefringent Optics GBO.

The general idea is to match one of the principal indices of the anisotropic material to the index of the incident medium. For example, let us regard case c) from Sect. 6.5.1. When $n_1 = n_{20}$ is fulfilled, from (6.35c) it follows:

$$r_p = \frac{n_{2o} \cos\varphi - n_1 \sqrt{1 - \frac{n_1^2}{n_{2o}^2} \sin^2\varphi}}{n_{2o} \cos\varphi + n_1 \sqrt{1 - \frac{n_1^2}{n_{2o}^2} \sin^2\varphi}} = 0 \quad \forall\, \varphi$$

Instead of one well-defined Brewster's angle, the p-polarized light will not be reflected at *any* angle of incidence, although s-polarized light is still reflected.

Of course, such effects facilitate the design of effective polarizers. The effect is simple to understand, because in the given geometry, the p-polarization senses only the ordinary refractive index. In the case that the latter is matched to the incidence medium index, of course, light reflection cannot occur because there is no difference in the refractive indices.

As another example, let us regard case a) in Sect. 6.5.1. We require, that $n_1 = n_{2\mathrm{e}}$. From (6.35a) it follows that

$$r_p = \frac{n_{2o}\cos\varphi - n_1\sqrt{1 - \dfrac{n_1^2}{n_{2e}^2}\sin^2\varphi}}{n_{2o}\cos\varphi + n_1\sqrt{1 - \dfrac{n_1^2}{n_{2e}^2}\sin^2\varphi}} = \frac{n_{2o} - n_1}{n_{2o} + n_1} \neq r_p\,(\varphi)$$

Therefore, Brewster's angle is completely absent in this case, instead, the reflectance of p-polarized light is completely independent on the angle of incidence. Such a behaviour may be of use when one wants to design omnidirectional mirrors.

Table 6.1 gives an overview on important GBO-effects. Today, GBO-effects are in practical use in polymer coatings, because in polymers a well-defined optical anisotropy may easily be induced by mechanical stretching of the films.

Table 6.1. Examples for GBO-Effects: "e" denotes the extraordinary, and "o" denotes the ordinary principal refractive index. φ is the angle of incidence, and the subscripts "s" and "p" denote s- or p-polarization. z is the direction perpendicular to the film surface

Optical axis in material 1	Optical axis in material 2	Matching condition	GBO-effect
isotropic	$\parallel$ to surface and $\parallel$ to incidence plane	$n_1 = n_{2\mathrm{o}}$	$R_s = 0\ \forall\ \varphi$
isotropic	$\perp$ to incidence plane	$n_1 = n_{2\mathrm{o}}$	$R_p = 0\ \forall\ \varphi$
$\parallel$z	$\parallel$z	$n_{1\mathrm{e}} = n_{2\mathrm{o}}$ and $n_{1\mathrm{o}} = n_{2\mathrm{e}}$	$R_s = R_p\ \forall\ \varphi$
Isotropic or $\parallel$z	$\parallel$z	$n_{1\mathrm{e}} = n_{2\mathrm{e}}$	$R_p \neq R_p(\varphi)$

7 Thick Slabs and Thin Films

7.1 Transmittance and Reflectance of a Thick Slab

In the previous chapter, we discussed Fresnel's formulae that were necessary to understand the effects of plane interfaces on a plane electromagnetic wave. In fact, up to now we regarded the effect of a *single* interface. Because a practically available sample such as shown in Fig. 6.1 always contains more than one interfaces, it will now be our purpose to understand the interplay of the effects caused by different interfaces contributing to the overall spectrum of the sample.

Again, we will restrict our attention to parallel interfaces, the typical situation in thin film optics. We start from the simplest case – a thick slab of a transparent material. In the following, we will use the terminus 'transparent' in the sense that the optical losses are negligible. For example, a typical window pane has two parallel surfaces, so that the light transmitted through the window is at least determined by the transmission coefficients of the two surfaces.

In fact, the situation in somewhat more involved. This becomes clear from Fig. 7.1. The figure shows the sample geometry relevant for the regarded thick slab. So we have a first medium, from where the light is incident. The impinging light passes, of course, the first interface with a transmittance determined by (6.19). To avoid confusion, we return to the symbols used in Sect. 6.2 and regard the incidence medium as the first one, while the slab material forms the second one. The transmittance T_{12} denotes the transmittance through the interface, when the light comes from medium 1 and passes into medium 2. Impinging on the second interface, the primarily transmitted light my again be transmitted into the 'third' medium, which is regarded to be identical to the first one. Therefore, the relevant transmittance is T_{21}. So that, first of all, the transmittance of the whole system including both interfaces depends on the product $T_{12}T_{21}$.

But this is not the whole story. At both interfaces, a certain fraction of the light may be reflected. The light reflected at the first surface (reflectance R_{12}) clearly contributes to the reflectance of the whole slab. But there is another situation in the case of light reflected at the second surface (R_{21}). It travels back to the first surface, and contributes to the slab's reflectance when it is transmitted through the first surface (now with a transmittance T_{21}). How-

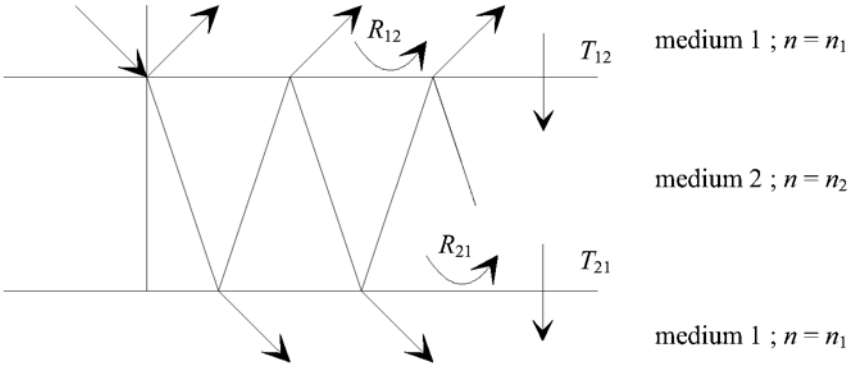

Fig. 7.1. Geometry of the sample

ever, it may be again reflected (R_{21}) and impinges onto the second surface for the second time. Again, it has a chance to be transmitted (and to contribute to the transmittance of the slab), or reflected, so that the game starts again. We come to the conclusion, that these internal multiple reflections form a further contribution to the light transmitted through the whole slab.

Let us formulate the considerations made above in an exact quantitative manner. Obviously, the reflectance R of the whole slab may be calculated by adding up the single contributions of the primarily reflected wave and those which arise as the result of multiple internal reflections. In fact, Fig. 7.1 indicates the philosophy of the calculation. In the case of negligible damping, we find:

$$\begin{aligned} R &= R_{12} + T_{12}R_{21}T_{21} + T_{12}R_{21}^3T_{21} + \dots \\ &= R_{12} + T_{12}R_{21}T_{21}\sum_{j=1}^{\infty}\left(R_{21}^2\right)^{j-1} = R_{12} + \frac{T_{12}R_{21}T_{21}}{1-R_{21}^2} = \frac{2R_{12}}{1+R_{12}} \end{aligned} \tag{7.1}$$

In full analogy, the transmittance T of the slab is:

$$\begin{aligned} T &= T_{12}T_{21} + T_{12}R_{21}^2T_{21} + \dots \\ &= T_{12}T_{21}\sum_{j=1}^{\infty}\left(R_{21}^2\right)^{j-1} = \frac{T_{12}T_{21}}{1-R_{21}^2} = \frac{1-R_{12}}{1+R_{12}} \end{aligned} \tag{7.2}$$

It is immediately seen, that T and R sum up to 1. In the case of normal incidence, in accordance with equations (6.20) and (6.19) we have:

$$R_{12} = R_{21} = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 \tag{7.3}$$

$$T_{12} = 1 - R_{12} = \frac{4n_1n_2}{(n_2+n_1)^2} = T_{21} \tag{7.4}$$

Table 7.1. Overview on normal incidence transmittance and reflectance for interfaces or slabs on air

Problem	Equation	Glass ($n \approx 1.5$ in the VIS)	Silicon ($n \approx 3.45$ in the IR)	Germanium ($n \approx 4.0$ in the IR)
Transmittance through the surface	$4n/(n+1)^2$	0.96	0.7	0.64
Reflectance of the surface	$(n-1)^2/(n+1)^2$	0.04	0.3	0.36
Transmittance through a slab	$2n/(n^2+1)$	0.923	0.535	0.47
Reflectance of a slab	$(n-1)^2/(n^2+1)$	0.077	0.465	0.53

A further simplification is valid, when we deal with a blank transparent substrate, embedded in air. In this case, $n_1 = 1$. Let us skip the subscript for the refractive index of the slab material ($n_2 = n$) and assume normal incidence. We quickly find:

$$R|_{\varphi=0} = \frac{(n-1)^2}{n^2+1} \tag{7.5}$$

$$T|_{\varphi=0} = \frac{2n}{n^2+1} \tag{7.6}$$

The couple of equations (7.1)–(7.6) enable the reader to calculate the transmittance or reflectance of a transparent slab. Consequently, they enable one to perform the task of a *forward search*. Table 7.1 summarizes some special cases of the application of these equations.

From the measured normal incidence transmittance of a slab, the refractive index may easily be calculated inverting (7.6) according to:

$$n = T^{-1} + \sqrt{T^{-2} - 1} \tag{7.7}$$

Hence, in the case of a damping-free slab, the reverse search is also a rather trivial matter.

At normal incidence, (7.1) and (7.2) may be simply generalized to the case of absorbing slab materials. Indeed, each bouncing onto the interface formed by the slab and the ambient is preceded by a penetration of the slab. The intensity damping per penetration is proportional to $\exp(-\alpha d_S)$, where d_S is the geometrical thickness of the slab. So that the relevant expressions are obtained by means of the substitution:

$$T_{21} \to T_{21}\,\mathrm{e}^{-\alpha d_S} \qquad R_{21} \to R_{21}\,\mathrm{e}^{-\alpha d_S}$$

Of course, the Fresnel's reflection and transmission coefficients are now dependent on the real *and* imaginary parts of the refractive indices. We obtain:

$$R = \frac{R_{12}\left[1 - \mathrm{e}^{-2\alpha d_S}\,(2R_{12} - 1)\right]}{1 - R_{12}^2\,\mathrm{e}^{-2\alpha d_S}} \tag{7.8}$$

$$T = \frac{(1 - R_{12})^2\,\mathrm{e}^{-\alpha d_S}}{1 - R_{12}^2\,\mathrm{e}^{-2\alpha d_S}} \tag{7.9}$$

These expressions need some explanation.

First of all, we state that the slab thickness d_S is only relevant for the values of T and R, when the slab material is absorbing. Then, for $\alpha d_S \to \infty$, the transmittance according to (7.9) becomes zero, while the reflectance (7.8) approaches that of the first interface. In the absence of absorption, equations (7.8) and (7.9) become identical to (7.1) and (7.2) and do thus not depend on the thickness of the slab. This is a familiar matter, because everybody knows from his daily experience, that the transmittance of a window pane does not depend on its thickness.

Nevertheless, it is intuitively clear, that the derived equations cannot find application when the slab thickness becomes too small. For $d = 0$, we clearly do not have any slab at all. Correspondingly, the reflectance should become zero, and the transmittance 1. But our equations state, that even for a vanishing slab thickness, there is still a finite reflection signal, which clearly lacks any sense.

For that reason, this section is named 'Transmittance and reflectance of a *thick* slab'. The equations derived so far cannot be applied to slabs with a (vanishingly) small thickness. It will be our next task to clarify what is exactly meant by the word 'thick', and to derive a criterion for the applicability of equations (7.8) and (7.9). But before coming to this point, let us make a further remark concerning the reverse search procedures.

As we have mentioned before, (7.7) allows to calculate the refractive index of a damping-free slab material from the normal incidence transmittance. If the material shows absorption, of course, knowledge of the transmittance alone is insufficient to calculate n and K without further model assumptions. On the other hand, the pair of optical constants may be calculated from transmittance and reflectance of the slab, either numerically or inverting equations (7.8) and (7.9) to find explicit expressions for n and K as a function of T and R. The refractive index of glass as shown in Fig. 4.2 has been determined this way from experimental T- and R-data.

7.2 Thick Slabs and Thin Films

We are now in a somewhat curious situation. We have derived a seemingly exact formula for the transmittance and reflectance of a slab of a transparent material, that turned out to be independent on the slab thickness. On the other hand, we surely know that the formula leads to an incorrect result when being applied to a very (vanishingly) thin slab. That means, that somewhere

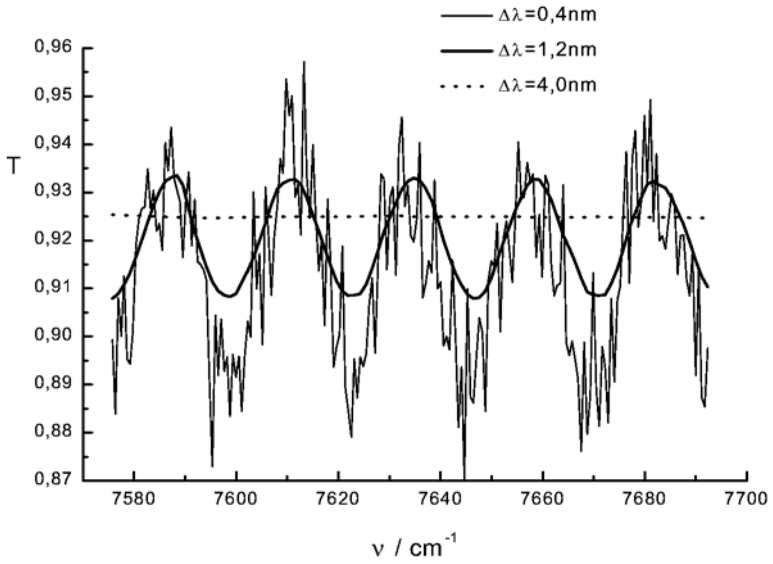

Fig. 7.2. Normal incidence transmittance of a 142 µm thick glass slab as measured at a commercial Shimadzu UV3001PC spectrophotometer with various values of the spectral bandwidth

in our derivation, we have presumed that the slab is 'sufficiently' thick. But where? And what is meant by 'sufficiently' thick?

In fact, the situation is even more worse. The applicability of (7.1) and (7.2) turns out to *not* depend on the absolute value of the thickness of the slab. This is illustrated in Fig. 7.2. Here we see a part of the measured NIR transmittance of a 142 µm thick glass slab. The measured transmittance turns out to depend on the spectral bandwidth $\Delta\lambda$ of the impinging irradiation. For a high spectral bandwidth ($\Delta\lambda = 4$ nm), the measured transmittance is more or less constant at a level of 0.925. This is fully consistent with our equations derived so far and corresponds to a quite reasonable refractive index of approximately 1.49. In this case, our equations are suitable for the description of the measurement. On the other hand, when the spectral bandwidth is smaller (higher degree of monochromaticity), there occur oscillations in the spectrum of the same sample, which cannot be explained in terms of the previous equations. In fact, $\Delta\lambda$ does not encounter into our equations at all.

So what is wrong? The periodic oscillations in the transmittance spectra from Fig. 7.2 indicate that there is an interference mechanism present in the system. Clearly, the multiply reflected waves in Fig. 7.1 may constructively or destructively interfere with each other, which would lead to a periodic modulation of the transmitted or reflected light intensity. But that would require mutually coherent light waves. On the other hand, in our derivation of (7.1) and (7.2) we could not consider any interference effects, because we

superimposed the *intensities* of the partial waves, and thus neglected the phase information necessary for any interference description.

The addition of intensities (instead of electric fields) is correct as long as the optical path difference between the superimposed partial beams is larger than the coherence length of the light (*incoherent* case). Therefore, our theoretical treatment ((7.1)–(7.9)) must be correct when the thickness of the slab exceeds the coherence length. The coherence length, in turn, is inversely proportional to $\Delta\lambda$. Therefore, for a small $\Delta\lambda$, the sample may be too thin to be described in terms of our equation, although this treatment may be correct for a higher $\Delta\lambda$. In the extreme case, when the slab thickness approaches zero, it will be smaller than any reasonable coherence length, so that, for ultrathin layers, the application of our equation makes no sense. These general considerations will allow us to formulate a quantitative criterion for the applicability of (7.1) and (7.2):

Let us regard a light beam in Fig. 7.1 travelling from the first interface to the second one and then backwards to the first. Its travelling time t will be $t = 2nd_S/c$ (normal incidence). For incoherent superposition, we shall require:

$$t = \frac{2nd_s}{c} > t_{coh}$$

with t_{coh} – coherence time of the light. In full analogy to the treatment of dephasing due to collisions in Sect. 4.2 ((4.3)), a finite coherence time causes a finite spectral bandwidth according to:

$$\Delta\omega = \frac{2}{t_{coh}}$$

while

$$\Delta\omega = \frac{2\pi c}{\lambda^2}\Delta\lambda$$

Combining the above relations, we obtain the following condition for incoherent superposition (and therefore the applicability of our equations):

$$d_s > \frac{\lambda^2}{2\pi n\Delta\lambda} \quad \text{or} \quad \Delta\lambda > \frac{\lambda^2}{2\pi n d_s} \tag{7.10}$$

Condition (7.10) is the result we wanted to obtain.

Let us check in how far this condition is consistent with the experimental observations from Fig. 7.2. We have a slab thickness of 0.142 mm and a refractive index near 1.5. The wavelength is approximately 1340 nm. For incoherent superposition (no interference pattern), according to (7.10) for the mentioned parameters we get the condition:

$$\Delta\lambda > 1.3$$

which is fully consistent with the spectra from Fig. 7.2.

In our terminology, we speak on a thick slab when condition (7.10) is fulfilled.

On the contrary, when condition (7.10) is violated, we have to expect that an interference pattern occurs. In this case, the application of (7.1) and (7.2) makes no sense, instead, these equations must be replaced by a theoretical apparatus that is based on the superposition of electric fields (including their phases), and not of light intensities.

Having defined the 'thick' slab by condition (7.10), we turn to the definition of a 'thin' film. In connection with optics, a system like shown in Fig. 7.1 is regarded as a thin film in the case, that practically *all* multiple internal reflections overlap coherently. In other words, the film thickness (for normal incidence) must be much smaller than the coherence length. So we come to the condition (7.11):

$$d \ll \frac{\lambda^2}{2\pi n \Delta\lambda} \quad \text{or} \quad \Delta\lambda \ll \frac{\lambda^2}{2\pi n d} \tag{7.11}$$

Condition (7.11) defines the thin film. To avoid confusion with the slab, the film thickness is given by the d without any subscript.

There is an intermediate case where the thickness is smaller but of the order of the coherence length, that corresponds to the interference of partially coherent light. This case is difficult to handle mathematically and will not be considered in this book. You should nevertheless note, that the oscillating spectra in Fig. 7.2 correspond to this partly coherent superposition of light.

Returning to our example and regarding an intermediate spectral bandwidth of 2 nm, from condition (7.11) we find that $d \ll 100\,\mu\text{m}$ must be fulfilled in order to regard the sample from Fig. 7.1 as an optical thin film. Note that (7.11) depends on the wavelength, so that in the UV the thickness (or the spectral bandwidth) must be considerably smaller.

7.3 Spectra of Thin Films

Having defined what is meant by the terminus 'thin film', we may now turn to the calculation of transmittance and reflectance of a thin film embedded between two media. First of all, let us remark that condition (7.11) in practice often defines a thickness less than a few micrometers. In this case, the mechanical stability of the system cannot be guaranteed, so that the film is deposited onto another solid material, which forms the substrate. Correspondingly, the first (incidence) and third (substrate) materials are usually different from each other, hence they may have different refractive indices. Such a system is given in Fig. 7.3.

The calculation of transmittance and reflectance of the system follows, in principle, the philosophy from Sect. 7.1. Instead of the intensity transmission and reflection coefficients $T_{ij}$and R_{ij} ((6.18) and (6.19)), we now have to deal

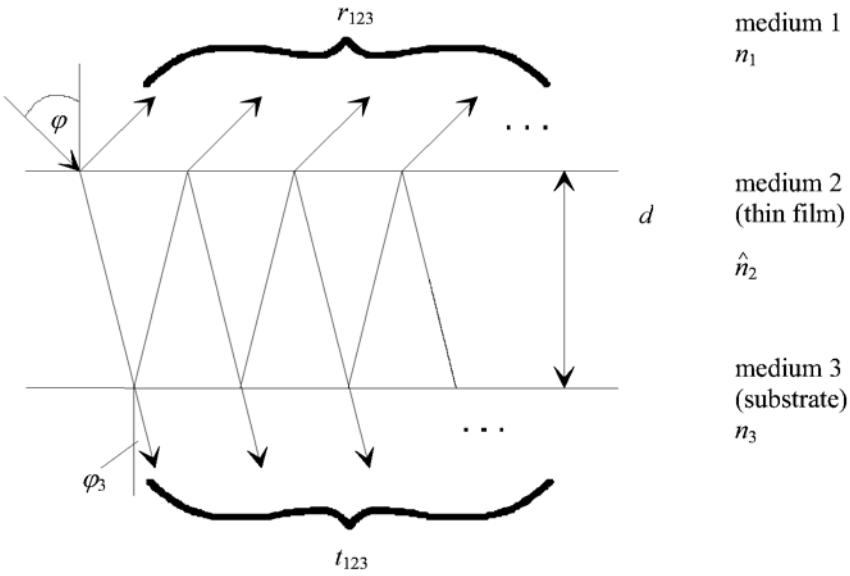

Fig. 7.3. Calculation of thin film transmittance and reflectance

with Fresnel's transmission- and reflection coefficients t_{ij} and r_{ij}, as given by (6.14)–(6.17). In general, these coefficients are complex and therefore carry an information on *both* amplitude and phase. Moreover, when the light travels through the layer, it gets a gain in phase. Let 2δ be the (possibly complex) phase gain per loop in the layer. In correspondence to our previous derivation, we obtain the following expressions for the electric field transmission coefficient t_{123} and the corresponding reflection coefficient r_{123}:

$$\begin{aligned} r_{123} &= r_{12} + t_{12}\,\mathrm{e}^{\mathrm{i}\delta} r_{23}\,\mathrm{e}^{\mathrm{i}\delta}\,t_{21} + t_{12}\,\mathrm{e}^{\mathrm{i}\delta} r_{23}\,\mathrm{e}^{\mathrm{i}\delta} r_{21}\,\mathrm{e}^{\mathrm{i}\delta} r_{23}\,\mathrm{e}^{\mathrm{i}\delta}\,t_{21} + \ldots \\ &= r_{12} + t_{12}\,r_{23}\,t_{21}\,\mathrm{e}^{2\mathrm{i}\delta}\left(1 + r_{21}\,r_{23}\,\mathrm{e}^{2\mathrm{i}\delta} + \ldots\right) \\ &= r_{12} + \frac{t_{12}\,r_{23}\,t_{21}\,\mathrm{e}^{2\mathrm{i}\delta}}{1 - r_{21}\,r_{23}\,\mathrm{e}^{2\mathrm{i}\delta}} \end{aligned} \tag{7.12}$$

$$\begin{aligned} t_{123} &= t_{12}\,\mathrm{e}^{\mathrm{i}\delta}\,t_{23}\left[1 + r_{21}\,r_{23}\,\mathrm{e}^{2\mathrm{i}\delta} + \left(r_{21}\,r_{23}\,\mathrm{e}^{2\mathrm{i}\delta}\right)^2 + \ldots\right] \\ &= \frac{t_{12}\,t_{23}\,\mathrm{e}^{\mathrm{i}\delta}}{1 - r_{21}\,r_{23}\,\mathrm{e}^{2\mathrm{i}\delta}} \end{aligned} \tag{7.13}$$

Let us shortly explain the abbreviations.

In full correspondence to Sect. 6.2, r_{123} denotes the ratio of the electric field strength in the reflected and incident waves of the system from Fig. 7.3, valid at the surface of the film. Analogously, $t_{123}$is the ratio of the transmitted field strength in medium 3 (at the film-substrate interface) and the incident one. t_{12}, t_{23}, r_{12}, and $r_{23}$are the typical interface Fresnel's coefficient between

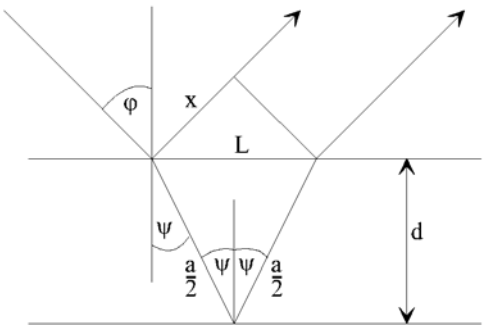

Fig. 7.4. Calculation of the phase gain for a single loop in the film

the media numbered in the subscripts. Per film penetration, the wavefunction (2.12) is multiplied with the factor $\exp(\mathrm{i}\delta)$.

From (6.14)–(6.17), we have:

$$t_{12}t_{21} = 1 - r_{12}^2$$

so that (7.12) may be written as:

$$r_{123} = \frac{r_{12} + r_{23}\,\mathrm{e}^{2\mathrm{i}\delta}}{1 - r_{21}\,r_{23}\,\mathrm{e}^{2\mathrm{i}\delta}} \tag{7.14}$$

Up to now, we cannot really work with these equations, because we have no valid expression for the phase gain. The latter may be obtained from geometrical considerations (Fig. 7.4). Let us restrict to the case $n_1 = 1$. In order to use geometrical considerations, we derive the expression for δ assuming the damping-free case (real film refractive index). The generalization to lossy film materials is then accomplished replacing the real film index in the final formula by the complex one. Using the symbols introduced in Fig. 7.4, we get:

$$\begin{aligned}
2\delta &= \frac{2\pi}{\lambda}\,(n_2 a - x) = 2\pi\nu\,(n_2 a - x) \rightarrow \delta = \pi\nu\,(n_2 a - x)\,; \\
a &= \frac{2d}{\cos\psi}\,; \quad x = L\sin\varphi \quad \text{with } L = 2d\tan\psi \\
\rightarrow \delta &= \pi\nu\left(\frac{2n_2 d}{\cos\psi} - \frac{2d}{\cos\psi}\sin\psi\sin\varphi\right) \qquad (7.15) \\
\text{with} \quad \sin\psi &= \frac{\sin\varphi}{n_2} \quad \text{and } n_2\cos\psi = n_2\sqrt{1-\sin^2\psi} \\
\rightarrow \delta &= 2\pi\nu d\sqrt{n_2^2 - \sin^2\varphi}
\end{aligned}$$

For a complex film refractive index, δ becomes complex as well, and then the phase factor $\exp(\mathrm{i}\delta)$ describes damping of the light wave in the film.

The couple of (7.13)–(7.15) allows to calculate the transmittance and reflectance of the film from figure (7.3) by using:

$$T = \frac{\mathrm{Re}\,(\hat{n}_3 \cos\varphi_3)}{\mathrm{Re}\,(\hat{n}_1 \cos\varphi)} \left|t_{123}\right|^2; \quad R = \left|r_{123}\right|^2 \tag{7.16}$$

Equation (7.16) are a natural generalization of (6.18) and (6.19) to the single film case. As we work with homogeneous materials and flat surfaces, scatter does not occur, and the absorptance becomes:

$$A = 1 - T - R$$

We are already familiar with a result of reflectance calculations based on (7.16). Indeed, in Sect. 6.4.2, we presented a calculated reflectance curve $R_p(\varphi)$ (Fig. 6.10) of the thin film system shown in Fig. 6.9. This reflectance has been calculated by means of (7.16), assuming glass as the incidence medium, a silver film of 46.2 nm thickness, and air as the third medium. It should be noted, however, that in this case, expression (7.15) must be modified, because the refractive index of the incidence medium is no further equal to 1. As already mentioned in Sect. 6.4.2, the calculation led to an excellent reproduction of the measured $R_p(\varphi)$ dependence. On the other hand, the absolute value of t_{123} represents the ratio between the electric field amplitudes in the third and first media (directly at the interfaces), and therefore gives immediate access to the electric field enhancement in surface plasmon spectroscopy.

7.4 Special Cases

7.4.1 Vanishing Damping

For the special case of vanishing damping (real Fresnel's coefficients and real δ), from (7.13)–(7.16) one obtains:

$$T = \frac{\dfrac{n_3 \cos\varphi_3}{n_1 \cos\varphi}\, t_{12}^2 t_{23}^2}{1 + r_{12}^2 r_{23}^2 + 2 r_{12} r_{23} \cos 2\delta} \tag{7.17}$$

$$R = \frac{r_{12}^2 + r_{23}^2 + 2 r_{12} r_{23} \cos 2\delta}{1 + r_{12}^2 r_{23}^2 + 2 r_{12} r_{23} \cos 2\delta} \tag{7.18}$$

It makes no sense to present illustrations of spectra calculated in terms of equations (7.17) and (7.18), because the system is still very idealized. In fact, the film is assumed to be deposited on a substrate, but the rear side of the substrate is not taken into account in our equations so far. We call this a thin film on a semi-infinite substrate. But the given equations are sufficient to get an idea on the properties of the oscillatory behaviour (the so-called interference pattern) that must be expected in thin film spectra. Clearly, as δ is proportional to the wavenumber, a spectral scan of the transmittance or reflectance must show oscillations in intensity due to the cosines present

in (7.17) and (7.18). We are already familiar with such oscillations, they appeared in the spectra shown in Fig. 7.2. Because in the absence of losses, T and R always sum up to one, it should be clear that minima in the transmittance must correspond to maxima in the reflectance and vice versa.

The practically important point is, that one may deduce the film thickness from the interference pattern. Let us focus on the extremal values of the interference pattern of a thin film. According to our equations, extrema will occur when the cosines achieve their extremal values (weak dispersion presumed). Hence, the argument of the cosines must be multiples of π. Let j be the order of the given extremum. In terms of (7.15), we have:

$$2\delta = 4\pi\nu d\sqrt{n_2^2 - \sin^2\varphi} = j\pi\,; \quad j = 0, 1, 2, \ldots$$

Then, the extrema in transmittance and reflectance (we will call them simply interference extrema in the following) occur at the wavenumbers ν_j:

$$\nu_j = \frac{j}{4d\sqrt{n_2^2 - \sin^2\varphi}} \tag{7.19}$$

In principle, from a spectrum like that shown in Fig. 7.2, the film thickness may be deduced by means of (7.19) when the film refractive index and the interference order j are known. It may however happen that the absolute order of the extremum is not known exactly. In this case, it makes no sense to apply (7.19) for the determination of the thickness. Instead, it makes sense to regard two extrema, for example adjacent extrema of the orders j and $j+1$. We then obtain:

$$d = \frac{1}{4\,(\nu_{j+1} - \nu_j)\,\sqrt{n_2^2 - \sin^2\varphi}} \tag{7.20}$$

The wavenumbers ν_j and ν_{j+1} may be obtained from the spectrum. Then, the thickness of the film may be calculated from the measured spectrum. If the interference extrema are not adjacent, we have:

$$d = \frac{\Delta j}{4\,(\nu_{j+\Delta j} - \nu_j)\,\sqrt{n_2^2 - \sin^2\varphi}} \tag{7.20a}$$

Let us regard the case of Fig. 7.2 (a glass slab at normal incidence ($\varphi = 0$)). In order to calculate the thickness of the glass slab, we apply (7.20a) and choose the interference extrema located at $7580\,\mathrm{cm}^{-1}$ (ν_j) and $7682\,\mathrm{cm}^{-1}$ ($\nu_{j+\Delta j}$). Obviously, $\Delta j = 8$, as may be simply obtained by counting the transmittance extrema in Fig. 7.2. The application of (7.20a) presumes knowledge of the refractive index, which has been estimated to be around 1.49 (Sect. 7.2). So we may calculate the thickness according to (7.20a) and obtain a value of

$$d = \frac{8}{4\,(7682 - 7588)\,\sqrt{1,49^2}}\,\mathrm{cm} = 0,01428\,\mathrm{cm} = 142.8\,\mu\mathrm{m}$$

which is quite close to the value of 142 µm that has been obtained by means of a micrometer. Hence, our theory of thin film spectra seems to be consistent with the experimental observations.

Finally, let us remark that (7.19) predicts equidistant (at the wavenumber scale) interference extrema as long as the refractive index may be regarded to be free of dispersion. If, however, dispersion is remarkable ($n = n(\nu)$), the extrema are no more equidistant, for normal dispersion their distance becomes smaller with increasing wavenumber. (7.20) or (7.20a) must then consider different refractive indices at different interference extrema, so that from (7.19) we find:

$$d = \frac{\Delta j}{4\,[\nu_{j+\Delta j}\sqrt{n_2^2\,(\nu_{j+\Delta j}) - \sin^2\varphi} - \nu_j\sqrt{n_2^2\,(\nu_j) - \sin^2\varphi}]} \qquad (7.20\mathrm{b})$$

As in practice the refractive index dispersion is often not exactly known, equations like (7.19) or (7.20b) will rather be used to estimate the refractive index dispersion when the thickness d has been previously determined by other means.

7.4.2 λ/2-Layers

Let us now turn to a very special case of a non-absorbing layer, namely the $\lambda/2$-layer (halfwave-layer). This terminus is applied to a non-absorbing layer, when the optical film thickness $d\sqrt{n_2^2 - \sin^2\varphi}$ is equal to $\lambda/2$. In this case, the term $4\pi\nu d\sqrt{n_2^2 - \sin^2\varphi}$ equals 2π. Therefore, the cosines in (7.17) and (7.18) become 1.

Let us for simplicity regard the case of normal incidence. The transmittance (7.17) becomes

$$T = \frac{n_3}{n_1} \times \frac{t_{12}^2 t_{23}^2}{(1 + r_{12}r_{23})^2}$$

Substituting the Fresnel coefficients by (6.14) and (6.15), we quickly find

$$T = \frac{4n_1n_3}{(n_1 + n_3)^2}$$

For a $\lambda/2$-layer, the transmittance (and the reflectance) do not depend on the refractive index of the film n_2! Moreover, our result is identical to (7.4), the transmittance of the air-substrate interface. In other words, such a layer has no effect on the transmittance and reflectance of the system. This result is also true for oblique incidence, as long as $4\pi\nu d\sqrt{n_2^2 - \sin^2\varphi}$ equals 2π.

It is immediately clear that a given film acts as a $\lambda/2$-layer only at certain wavelength values. The latter are determined by the condition

$$2\delta = 4\pi\nu d\sqrt{n_2^2 - \sin^2\varphi} = j\pi\,; \quad j = 0, 2, 4, 6, \ldots$$

where the mentioned cosines become 1.

7.4.3 $\lambda/4$-Layers

We deal with a quarterwave-layer ($\lambda/4$-layer), when

$$2\delta = 4\pi\nu d\sqrt{n_2^2 - \sin^2\varphi} = \pi$$

is fulfilled. In this case, the optical thickness $d\sqrt{n_2^2 - \sin^2\varphi} = \lambda/4$. Correspondingly, the cosines in (7.17) and (7.18) become -1. Regarding again normal incidence, we obtain for the transmittance:

$$T = \frac{n_3}{n_1} \cdot \frac{t_{12}^2 t_{23}^2}{(1 - r_{12}r_{23})^2} = \frac{4n_1 n_2^2 n_3}{(n_1 n_3 + n_2^2)^2} \tag{7.21}$$

This equation embodies an utmost important special case. Let us assume, that the refractive index of the film is in between those of the incident medium and the substrate, while

$$n_2 = \sqrt{n_1 n_3}$$

is fulfilled. Then, the transmittance (7.21) becomes 1, and consequently, the reflectance of the system becomes zero. Therefore, such quarterwave layers may lead to an antireflection effect and are therefore quite important in optical thin film design.

On the other hand, let us regard that the film refractive index is rather high ($n_2 > n_1$, n_3). Then, from (7.21) one obtains:

$$\frac{\partial R}{\partial n_2} = -\frac{\partial T}{\partial n_2} = \frac{8n_1 n_2 n_3}{(n_1 n_3 + n_2^2)^3}\left(n_2^2 - n_1 n_3\right) > 0$$

With an increasing film refractive index, the reflectance of a high refractive index quarterwave layer increases as well. Therefore, such quarterwave layers may be used for reflection enhancement purposes.

The same is valid for all interference extrema, where the condition

$$2\delta = 4\pi\nu d\sqrt{n_2^2 - \sin^2\varphi} = j\pi\,; \quad j = 1, 3, 5, 7, \ldots$$

is fulfilled.

The mentioned properties of the extrema of the interference pattern of a thin film are sufficient to discuss the general behaviour of the transmittance

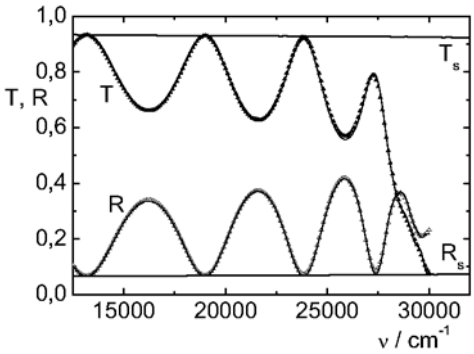

Fig. 7.5. Measured thin film spectra of a TiO_2 film

and reflectance of a damping-free single film on a substrate. Indeed, in all the extrema of the interference pattern as defined by (7.19), the film behaves either as a halfwave or as a quarterwave layer. In the halfwave points, the transmittance and the reflectance of the film-on-substrate system are identical to the values for the uncoated substrate. Therefore, the halfwave points may be easily identified in an experimental spectrum.

The remaining extrema of the interference pattern have to be regarded as quarterwave points. The transmittance in the quarterwave points is determined by (7.21), and it will depend on the relation between the refractive indices whether (7.21) defines minima or maxima of the transmittance T. Due to the lack of damping, the reflectance $R = 1 - T$.

Let us discuss the behaviour of the transmittance in the quarterwave points, restricting on the case that $n_1 < n_3$. For $n_2 = n_1$ and $n_2 = n_3$, (7.21) yields a transmittance identical to the bare ambient-substrate interface. For $n_1 < n_2 < \sqrt{n_1 n_3}$, the derivative of the transmittance with respect to the film refractive index is positive. Therefore, in this case the transmittance in the quarterwave points exceeds that of the bare substrate. The transmittance achieves its maximum value 1 for $n_2 = \sqrt{n_1 n_3}$. The derivative becomes negative for $n_2 > \sqrt{n_1 n_3}$. But the transmittance must be still higher than that of the substrate, as long as $n_2 < n_3$. A further increase in the film refractive index leads to a transmittance that is lower than that of the bare substrate, so that the quarterwave points will correspond to minima in the transmittance.

Let us look at an example. Figure 7.5 shows the normal incidence transmittance T and the reflectance R of a 337 nm thick titanium dioxide film on a quartz glass substrate. The figure also shows the transmittance T_s and reflectance R_s of the bare substrate.

We see the expected oscillatory behaviour of the spectra. For wavenumbers below $25000\,\mathrm{cm}^{-1}$, the film transmittance and reflectance sum up to 1 (no damping), so that our previous discussion is applicable. In the halfwave points, the film spectra are tangential to the substrate spectra, so that the halfwave points may be easily identified. In the quarterwave points, the mea-

sured reflectance is higher than that of the substrate, so that the film refractive index must be higher than the substrate index. Obviously, the refractive index shows normal dispersion, because the reflectance maxima become higher in reflection values with increasing wavenumber, and are not equidistant on the wavenumber scale.

In fact, our knowledge is even sufficient to perform a simple reverse search procedure. When neglecting the rear substrate surface, from (7.21) the film refractive index may be deduced in the quarterwave points. Then, we may find the film thickness from (7.20b).

7.4.4 Free-Standing Films

Let us now turn to another special case, where the thin film is not embedded on a substrate, but surrounded by air from both sides. Hence, materials 1 and 3 are identical, and our equations for the electric field transmission and reflection coefficients (7.13) and (7.14) become:

$$t_{123} = \frac{t_{12}t_{21}\mathrm{e}^{\mathrm{i}\delta}}{1 + r_{12}r_{21}\mathrm{e}^{2\mathrm{i}\delta}}$$

$$r_{123} = \frac{r_{12} + r_{21}\mathrm{e}^{2\mathrm{i}\delta}}{1 + r_{12}r_{21}\mathrm{e}^{2\mathrm{i}\delta}}$$

Let us now assume, that the film's refractive index is complex. Then, the phase gain will be complex as well and may be written as:

$$\delta = \delta' + \mathrm{i}\delta''$$

From that, we find for the transmittance (7.16):

$$T = \frac{|t_{12}|^2|t_{21}|^2\mathrm{e}^{-2\delta''}}{1 + |r_{12}|^2|r_{21}|^2\mathrm{e}^{-4\delta''} + 2\mathrm{e}^{-2\delta''}\left[\mathrm{Re}\left(r_{12}r_{21}\right)\cos 2\delta' - \mathrm{Im}\left(r_{12}r_{21}\right)\sin 2\delta'\right]} \tag{7.22}$$

Let us again check the output of this equation for $d \to 0$. Then, the sinus-terms become zero, while the cosines become 1. It is obtained:

$$T\left(d \to 0\right) = \frac{|t_{12}|^2|t_{21}|^2}{1 + |r_{12}|^2|r_{21}|^2 + 2\mathrm{Re}\left(r_{12}r_{21}\right)} = 1,$$

because $|t_{12}t_{21}|^2 = \left|1 - r_{12}^2\right|^2$ and

$$\begin{aligned}\left|1 - r_{12}^2\right|^2 &= \left|1 - \mathrm{Re}r_{12}^2 - \mathrm{i}\mathrm{Im}r_{12}^2\right|^2 = \left(1 - \mathrm{Re}r_{12}^2\right)^2 + \left(\mathrm{Im}r_{12}^2\right)^2 \\ &= 1 - 2\mathrm{Re}r_{12}^2 + \left(\mathrm{Re}r_{12}^2\right)^2 + \left(\mathrm{Im}r_{12}^2\right)^2 = 1 - 2\mathrm{Re}r_{12}^2 + \left|r_{12}^2\right|^2\end{aligned}$$

So that our thin-film equations predict a 100% transmittance and a vanishing reflectance for an infinitively thin film, which is a quite reasonable result.

Having checked this asymptotic behaviour, we will now utilize (7.22) to obtain an expression for the transmittance of a thick slab. This makes sense, because we have no expression for the transmittance of an absorbing thick slab at oblique incidence so far. On the other hand, in (7.22) both absorption and oblique incidence are automatically considered, and the corresponding equations for the thick slab might be obtained transferring (7.22) to the case of incoherent light superposition.

That may be achieved averaging (7.22) over the real part of the phase gain. In doing so, we simulate a measurement where the phase information is destroyed caused by a statistical distribution of the phases of the interfering light waves. That leads us to the transmittance of a thick slab T_s according to:

$$T_s = \frac{1}{\pi}\int_0^{\pi} \mathrm{d}\delta' \times \left\{ \frac{|t_{12}|^2 |t_{21}|^2 \mathrm{e}^{-2\delta''}}{1+|r_{12}|^2|r_{21}|^2\mathrm{e}^{-4\delta''}+2\mathrm{e}^{-2\delta''}\left[\mathrm{Re}\,(r_{12}r_{21})\cos 2\delta' - \mathrm{Im}\,(r_{12}r_{21})\sin 2\delta'\right]} \right\}$$

$$= \frac{|t_{12}|^2 |t_{21}|^2 \mathrm{e}^{-4\pi\nu d_S \mathrm{Im}\sqrt{\hat{n}_2^2 - \sin^2\varphi}}}{1 - |r_{12}|^2 |r_{21}|^2 \mathrm{e}^{-8\pi\nu d_S \mathrm{Im}\sqrt{\hat{n}_2^2 - \sin^2\varphi}}} \tag{7.23}$$

The integral has been solved according to:

$$\int \frac{\mathrm{d}x}{a + b\cos x + c\sin x}\Big|_{a^2 > b^2 + c^2} = \frac{2}{\sqrt{a^2 - b^2 - c^2}} \arctan\left[\frac{(a-b)\tan\dfrac{x}{2} + c}{\sqrt{a^2 - b^2 - c^2}}\right]$$

Equation (7.23) is identical by structure with (7.9). The corresponding equation for the reflectance may be obtained in the same way, let us write out the result:

$$R_s = |r_{12}|^2 + \frac{|t_{12}|^2 |r_{21}|^2 |t_{21}|^2 \mathrm{e}^{-8\pi\nu d_S \mathrm{Im}\sqrt{\hat{n}_2^2 - \sin^2\varphi}}}{1 - |r_{12}|^2 |r_{21}|^2 \mathrm{e}^{-8\pi\nu d_S \mathrm{Im}\sqrt{\hat{n}_2^2 - \sin^2\varphi}}} \tag{7.24}$$

The couple of equations (7.23) and (7.24) is our final result for the description of transmittance and reflectance of a thick slab. These equations are utmost important in thin film spectroscopy, because such thick slabs usually serve as a substrate for the thin films that have further to be investigated. Clearly, without knowledge on the behaviour of the bare substrate, we would not be able to correctly describe the properties of the film-on-substrate system.

7.4.5 A Single Thin Film on a Thick Substrate

We are now able to perform the final step that will enable us to calculate transmittance and reflectance of a thin film on a thick substrate.

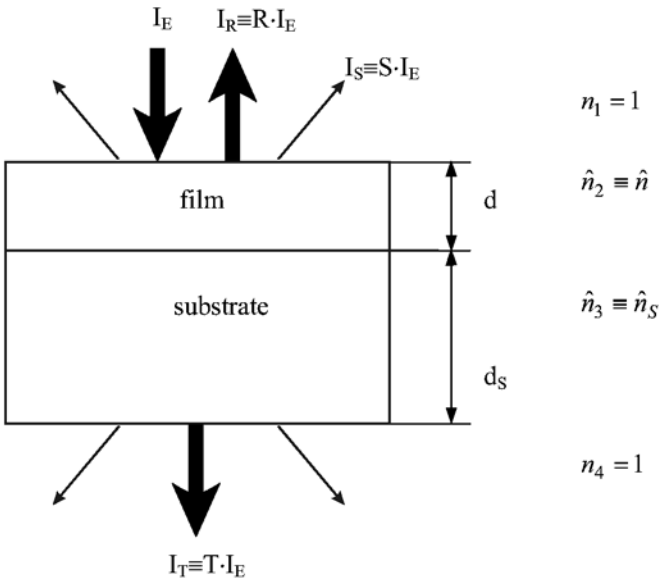

Fig. 7.6. A thin film on a substrate

First of all, let us take a closer look on the sample which will be regarded now. As shown in Fig. 7.6, we are now dealing with a realistic system in thin film spectroscopy, namely a film on a thick substrate with a finite thickness. In the following, all substrate parameters are indicated by the subscript 's'.

In order to apply our theoretical description to the system, the film and substrate thickness values shall be in the correct relation to the coherence length of the light. We shall require, that within the film, the light waves superimpose coherently, while they superimpose incoherently in the substrate. In terms of the conditions (7.10) and (7.11), that yields:

$$d \ll \frac{\lambda^2}{2\pi n \Delta\lambda} \wedge \frac{\lambda^2}{2\pi n_s \Delta\lambda} < d_S$$

For simplicity, we will assume that the surrounding medium is air ($n_1 = n_4 = 1$), while the refractive indices of the film and the substrate may be complex (for symbols see Fig. 7.6). We start our discussion from (7.23) and (7.24). These equations describe the properties of the bare substrate, the only thing that has to be done is to 'add' the film onto the substrate. That may be done in the following manner:

In (7.23), the term t_{12} in the nominator gives us the ratio between the electric field amplitudes at both sides of the air-substrate interface. This field transmission coefficient simply has to be replaced by the corresponding thin-film coefficient t_{123}. Moreover, we have to remember that the substrate material is now identical to the third material, so that instead of t_{21}, we have to write t_{31}. The same type of procedure must be applied to the denomina-

tor. Its function is to describe the multiple *internal* reflections between the substrate surfaces. So one of the Fresnel coefficients ($r_{12} = -r_{21}$) has to be replaced by the thin film field reflection coefficient r_{321}, the other one by r_{31}. An equivalent procedure has to be applied to the reflectance according to (7.24). As the result, in the terminology introduced in Fig. 7.6, we obtain the following equations for T and R:

$$T = \frac{|t_{123}|^2 |t_{31}|^2 \mathrm{e}^{(-2\mathrm{Im}(\delta_S))}}{1 - |r_{321}|^2 |r_{31}|^2 \mathrm{e}^{(-4\mathrm{Im}(\delta_S))}} \tag{7.25}$$

$$R = |r_{123}|^2 + \frac{|t_{123}|^2 |r_{31}|^2 |t_{321}|^2 \mathrm{e}^{(-4\mathrm{Im}(\delta_S))}}{1 - |r_{321}|^2 |r_{31}|^2 \mathrm{e}^{(-4\mathrm{Im}(\delta_S))}} \tag{7.26}$$

$$t_{123} = \frac{t_{12} t_{23} \mathrm{e}^{\mathrm{i}\delta}}{1 + r_{12} r_{23} \mathrm{e}^{2\mathrm{i}\delta}}$$

$$r_{123} = \frac{r_{12} + r_{23} \mathrm{e}^{2\mathrm{i}\delta}}{1 + r_{12} r_{23} \mathrm{e}^{2\mathrm{i}\delta}}$$

$$\delta_{(S)} = 2\pi\nu d_{(S)} \sqrt{\hat{n}^2_{(S)} - \sin^2\varphi}$$

This couple of equations allows us to calculate T and R for a film-on substrate system. In other words, we are now able to perform the forward search task.

Let us have a look at some examples of calculations performed within the framework of the theoretical apparatus derived so far. The first example concerns the titanium dioxide film presented in Fig. 7.5. In this figure, the single symbols correspond to measured values, while the solid lines correspond to a theoretical spectrum calculated by means of (7.25) and (7.26). In order to achieve such a good agreement between measurement and theory, the optical thin film constants and the thickness have been varied to achieve a good fit of the experimental behaviour. In the present case, this has been done assuming a multioscillator model to describe the optical constants of TiO_2. The corresponding optical constants are shown in Figs. 7.7 and 7.8.

Hence, our theoretical apparatus is able to reproduce the experimental spectra of thin films on a substrate, as long as the optical constants of the film are correctly chosen. That means, that such a spectra fit may also be used to perform a reverse search task. In fact, this is the method that has been applied to obtain the optical constants shown in the Figs. 7.7 and 7.8. Mathematically, this may be done minimizing the error function:

$$F = \sum_{j=1}^{M} \left\{ w_T(\nu_j) \left[T_{\mathrm{exp}}(\nu_j) - T_{\mathrm{calc}}(\nu_j) \right]^2 + w_R(\nu_j) \left[R_{\mathrm{exp}}(\nu_j) - R_{\mathrm{calc}}(\nu_j) \right]^2 \right\} \tag{7.27}$$

Here, the subscript 'exp' denotes measured values, while 'calc' corresponds to calculations according to (7.25) or (7.26). The w-functions represent the rela-

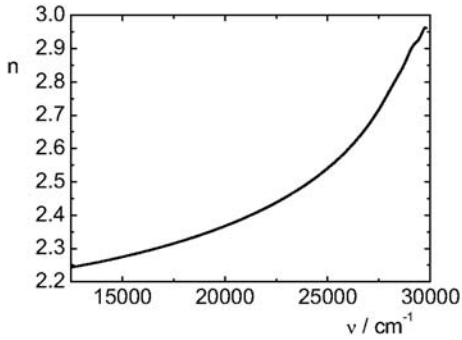

Fig. 7.7. Refractive index of TiO_2, as consistent with the spectra from Fig. 7.5

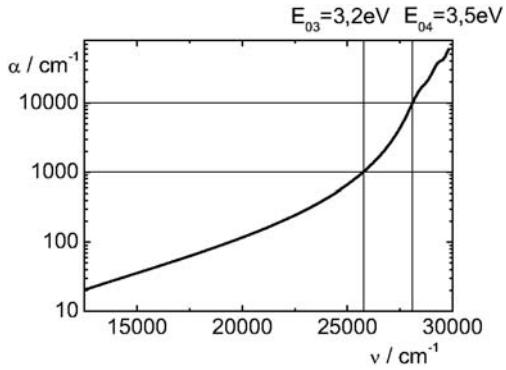

Fig. 7.8. Absorption coefficient of TiO_2, as consistent with the spectra from Fig. 7.5. (The values E_{03} and E_{04} indicate the photon energies, where the absorption coefficient reaches the values of $1000\,\mathrm{cm}^{-1}$ and $10000\,\mathrm{cm}^{-1}$, correspondingly.)

tive weight of the single error terms, it makes sense to choose them inversely proportional to the square of the measurement error. M is the number of wavenumber points considered in the calculation. However, an error function like (7.27) usually has a lot of local minima, so that any mathematical minimization procedure leads to a multiplicity of solutions. From that multiplicity, the physically correct solution must be identified with utmost care.

Let us finally look at another example. Figure 7.9 shows the normal incidence T- and R-spectra of a 157 nm thick indium tin oxide (ITO) film on glass. ITO is a material that combines transpareny in the VIS with a high electrical dc conductivity. Hence, one must expect that in the VIS, we have refractive indices above 1 and low absorption coefficients. On the other hand, the free electrons that are responsible for the dc conductivity should lead to IR-optical constants according to Drude's formula. Again, the spectra from Fig. 7.9 have been fitted by means of (7.25) and (7.26), and the corresponding optical constants (Fig. 7.10) confirm the expected behaviour.

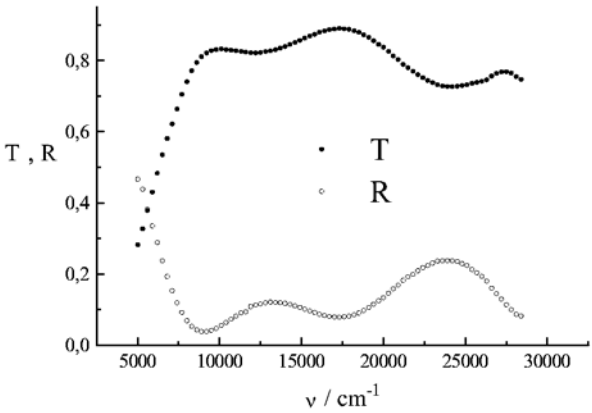

Fig. 7.9. T and R for an ITO film

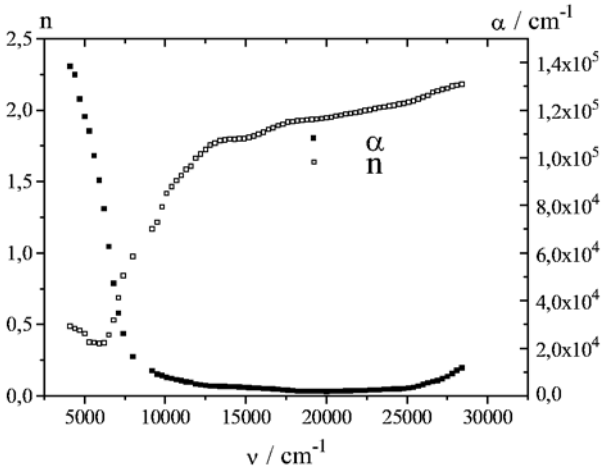

Fig. 7.10. Optical constants of ITO, consistent with the spectra from Fig. 7.9

7.4.6 Extended Detail: A Few More Words on Reverse Search Procedures

Let us make a very few additional remarks on the reverse search procedures applied to a system like shown in Fig. 7.6. As already mentioned, in the reverse search, the task is to recalculate the optical constants of the film and the film thickness from measured transmittance and reflectance data.

As shown in the previous sections, it is possible to provide explicit expressions for the spectra of a thin film sample if geometry and optical constants are known. It is however impossible to obtain explicit expressions for the optical constants as a function of the measured data. This makes the numerical side of the reverse search more complicated than the forward search, because

it is necessary to apply involved iteration procedures to find the result. As a further complication, unambiguity and numerical stability of the result may not be guaranteed.

From the formal point of view, the reverse search procedures may be classified into single wavelength methods and multiwavelength methods. The latter include the Kramers–Kronig-methods as well as curve-fitting techniques. Often, the reverse search bases on the numerical minimization of an appropriately defined error function F, as given by (7.27).

The numerical minimization of (7.27) represents a purely mathematical problem, and the corresponding skills will not be discussed here. In the ideal case, a set of optical constants may be found which generates theoretical spectra equal to those measured, so that F becomes zero. In practice this is impossible, and it makes no sense to minimize the error function (7.27) below a threshold value determined by the measurement accuracies ΔT and ΔR. Thus, we may regard that the minimization was successful when the condition:

$$F < \sum_{j=1}^{M} \left\{ w_T(\nu_j) \left[\Delta T(\nu_j)\right]^2 + w_R(\nu_j) \left[\Delta R(\nu_j)\right]^2 \right\}$$

is fulfilled. As several sets of optical constants may fulfill this criterion, we may obtain a multiplicity of mathematically acceptable solutions, from which the physically meaningful has to be selected. Especially in thin film optics, the discussion of the solution multiplicity may be a troublesome procedure.

If one has no idea on the mutual correlation of the optical constants at different frequencies, one may straightforwardly apply the minimum condition of (7.27):

$$grad\, F = \mathbf{0}$$

which reduces to a set of M equation systems:

$$[T_{\text{exp}}(\nu_j) - T_{\text{calc}}(\nu_j)] = 0$$

$$[R_{\text{exp}}(\nu_j) - R_{\text{calc}}(\nu_j)] = 0$$

in the case that no analytical dependence of the optical constants at different frequencies is assumed. These systems of equations may be solved numerically at each wavenumber of interest, which would be a typical single wavelength procedure. When the film thickness is known, we have two equations for two unknown values n and K for each wavenumber. As a disadvantage, this method often suffers from a multiplicity of solutions, which may be discontinuous with wavenumber.

There exist several methods to reduce the solutions multiplicity. First of all, a sufficient number of independent measurements or their clever choice may reduce the solutions multipicity, however, it demands the access to the

Table 7.2. *A priori* information obtained from thin film spectra at normal incidence, transparent substrat, ambient medium air

$T+R$	$T(\lambda/2)$	$T(\lambda/4)$	Further information	Interpretation
1	T_s	$> T_s$	–	$n < n_s$
	T_s	$< T_s$	–	$n > n_s$
	$< T_s$	$< T_s$	–	Positive index gradient, $\langle n \rangle > n_s$
		$> T_s$	–	Positive index gradient, $\langle n \rangle < n_s$
	$> T_s$	$< T_s$	–	Negative index gradient, $\langle n \rangle > n_s$
		$> T_s$	–	Negative index gradient, $\langle n \rangle < n_s$
< 1	–	–	–	$A + S > 0$
$\ll 1$	–	–	homogeneous layer with $S = 0$	$A \approx 1$; $d \gg \frac{\lambda}{8\pi} \frac{1-\sqrt{R}}{\sqrt{R}}$
$\ll 1$	–	–	homogeneous layer with $d \leq \frac{\lambda}{8\pi} \frac{1-\sqrt{R}}{\sqrt{R}}$	$S > 0$

corresponding measurement equipment. However, it is often impossible to increase the number of measurements because of a lack of equipment. One further way to reduce a possible solution multiplicity is given by the application of curve fitting procedures. In this case, one postulates an analytical dispersion law such as defined in the Chaps. 2–5. In minimizing (7.27), the free parameters (for example resonance frequencies of linewidth values) of the dispersion model have to be determined. Typically, the film thickness may also be found this way.

The curve fitting procedures are widely applied today, however, their successfull application demands the reliable choice of a suitable dispersion law. One of their advantages is that they may find application in quite restricted spectral regions.

No matter whether one prefers to work with single wavelength methods or with curve fitting approaches, there are a few general rules that are helpful in performing the reverse search. First of all, the solution should satisfy at least two criterions:

- It should be continuous with wavenumber.
- The dispersion should be consistent with the Kramers–Kronig-relations.

In sophisticated curve fitting procedures, these criterions are automatically fulfilled. For single wavelength methods, however, they may be a criterion to exclude physically meaningless solutions.

Additionally, it makes sense to utilize the information supplied from general properties of the measured spectrum. For the special case of a single film

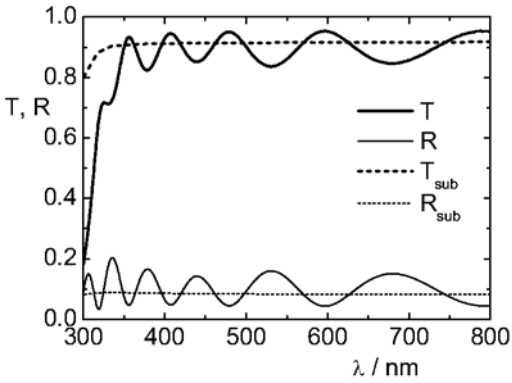

Fig. 7.11. Transmittance and reflectance of a gradient index layer

on a transparent substrate, surrounded by air, important *a priori* information may be obtained from the spectrum without any troublesome calculations. Table 7.2 summarizes important special cases.

Let us shortly comment on the information given in the table.

- In the upper part of the table ($T + R = 1$), we deal with loss-free samples which show an interference pattern. $T(\lambda/2)$ denotes the transmittance in halfwave points, and $T(\lambda/4)$ in quarterwave points. As already discussed, the behaviour in quarterwave points determines whether or not the film has a higher refractive index as the substrate. Moreover, in the absence of losses, $T(\lambda/4)$ as given by (7.25) does not depend on the film thickness. It may therefore be used to calculate the film refractive index assuming $K = 0$. Afterwards, one may calculate the film thickness from (7.19)–(7.20b).
- When the sample is free of optical losses, but the transmittance in the halfwave points differs from that of the substrate, then we have to accept that we deal with a gradient layer. That means, that the film refractive index changes smoothly with the distance from the substrate. When the film index increases with increasing distance from the substrate, we speak on a positive refractive index gradient. On the contrary, the gradient is negative when it decreases with increasing distance from the substrate. The behaviour in the halfwave points is thus helpful to identify the kind of index gradient. $\langle n \rangle$ denotes the average refractive index (averaged over the film thickness). The mentioned rules correspond to the special case of a linear refractive index gradient.

  To provide an example, Fig. 7.11 shows the spectra of a gradient index layer, prepared from a material mixture of silicon dioxide ($n \approx 1.45$) and niobium pentoxide ($n \approx 2.3$) on quartz glass ($n \approx 1.45$). Nearby the substrate, the niobium pentoxide concentration is high, while it becomes smaller with increasing distance from the substrate. Hence, we deal with a negative index gradient, so that in the halfwave points, the transmittance

exceeds that of the bare substrate. However, as the average film index is higher than that of the substrate, in the quarterwave points we have a sample transmittance lower than the substrate transmittance. This is one of the situations analysed in Table 7.2. Such spectra cannot be calculated by the theory developed in this Chap. 7, but becomes accessible to the reader which is familiar with the material from the following Chap. 8.
The three rows on the bottom of the table correspond to lossy samples. In this case, the interference pattern may disappear. Nevertheless, for the case of strong damping, asymptotic equations may be derived from (7.25) and (7.26). Some of these results are given in the table without derivation.

The *a priori* information obtained from general spectral features may, of course, be used to identify the physically meaningful solution of a reverse search procedure. Moreover, it may be utilized at the very beginning to determine a suitable initial approximation, that makes the minimization of (7.27) more efficient and faster.

8 Extended Details: Gradient Index Films and Multilayers

8.1 Gradient Index Films

8.1.1 General Assumptions

In the previous chapter, we derived the expressions for transmittance and reflectance of a single homogeneous thin film on a thick substrate. As one of the general assumptions, the refractive index of the film should not depend on the coordinates, so that $n \neq n(x, y, z)$. This is a rather special case, which may be regarded as a model to simplify the calculation of thin film spectra. In practice, any real optical thin film is (at least slightly) inhomogeneous. In this chapter we will deal with the utmost important special case, that the refractive index depends only on the z-coordinate (a so-called stratified medium). That means, that the film properties change with distance from the substrate. This may be caused, for example, by changes in the deposition conditions during the preparation of a thin film.

In order to describe the optical behaviour of such an inhomogeneous film, we have to solve Maxwell's equations with a z-dependent dielectric function.

First of all, let us remember that we exclusively deal with harmonic electric and magnetic fields. We shall write the electric and magnetic fields in the following manner:

$$\boldsymbol{E} = \boldsymbol{E}_0\,(\boldsymbol{r})\,\mathrm{e}^{-\mathrm{i}\omega t}\,; \qquad \boldsymbol{H} = \boldsymbol{H}_0\,(\boldsymbol{r})\,\mathrm{e}^{-\mathrm{i}\omega t}$$

For non-magnetic materials we have:

$$\varepsilon = \varepsilon\,(z)\;; \quad \mu = 1$$

As in the chapter on Fresnel's equations, it makes sense to discuss the particular cases of s- and p-polarization separately.

Before starting with any derivations, let us state that this Chap. 8 will be somewhat specific from both its content and the meaning of the symbols used here. We emphasize at the very beginning, that the full chapter will deal with rather complicated and sometimes tedious mathematical derivations. It is the authors personal opinion that it is worth to understand the derivation of the equations suitable for the calculation of complicated thin film systems before applying them. Therefore the derivations are included into this book.

Nevertheless, if the reader is only interested in the final "recipe" for calculations, the relevant information will be found in the Tables 8.1 and 8.2, which comprise the main results of the derivations.

Concerning the symbols, there will be an important difference to the meaning of the Fresnel-coefficients used so far. Up to now, any field transmission or reflection coefficient t_{ij}, r_{ij}, t_{ijk}, and r_{ijk} had the meaning of the ratio between *electric* fields. In this chapter, this will be true only for the case of s-polarization. For p-polarization, any field transmission or reflection coefficient will have the meaning of the ratio between the corresponding *magnetic* fields. Particularly, the Fresnel-formula for t_p ((6.15)) will not be applicable in connection with the expressions derived in this chapter and must be replaced by the relevant expression for the magnetic fields (see Sect. 8.1.4).

8.1.2 *s*-Polarization

Let us start with the case of s-polarization. Assuming a coordinate system like introduced in Fig. 6.3, we have the field components:

$$\boldsymbol{E} = \begin{pmatrix} 0 \\ E_y \\ 0 \end{pmatrix}$$

From (2.1, term 2.), we get the couple of equations (8.1) (compare with the calculations in Sect. 6.2)

$$\begin{aligned} \mathrm{i}\omega\mu_0 H_x &= -\frac{\partial}{\partial z} E_y \\ \mathrm{i}\omega\mu_0 H_y &= 0 \\ \mathrm{i}\omega\mu_0 H_z &= \frac{\partial}{\partial x} E_y \end{aligned} \tag{8.1}$$

Hence, for $\omega \neq 0$, we have:

$$\boldsymbol{H} = \begin{pmatrix} H_x \\ 0 \\ H_z \end{pmatrix}$$

So that, from (2.1, term 4.), it follows that

$$\begin{aligned} \frac{\partial}{\partial z} H_x - \frac{\partial}{\partial x} H_z &= -\mathrm{i}\omega\varepsilon\varepsilon_0 E_y \\ \frac{\partial}{\partial y} H_z = \frac{\partial}{\partial y} H_x &= 0 \end{aligned} \tag{8.2}$$

Differentiating the first and third equations in (8.1) for the second time with respect to coordinates and summing them up, we obtain the wave equation:

$$\frac{\partial^2}{\partial x^2}E_y + \frac{\partial^2}{\partial z^2}E_y = -\frac{\omega^2}{c^2}\varepsilon(z)\,E_y \tag{8.3}$$

Here, the derivatives of the magnetic field have been substituted by means of (8.2). Equation (8.3) allows a separation of variables according to:

$$E_y(x,z) = X(x)\,U(z) \tag{8.4}$$

That results in:

$$\frac{1}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2} = -\frac{1}{U}\frac{\mathrm{d}^2U}{\mathrm{d}z^2} - \frac{\omega^2}{c^2}\varepsilon(z) = \text{const.} \tag{8.5}$$

For convenience, we rewrite the constant according to:

$$\text{const.} = -k_0^2\eta^2\,; \quad k_0 \equiv \frac{\omega}{c} \tag{8.6}$$

From (8.5), it is immediately obtained that

$$X \propto \mathrm{e}^{\mathrm{i}k_0\eta x} \tag{8.7}$$

Therefore, according to (8.4) the full electric field may be written as

$$E_y = U(z)\,\mathrm{e}^{\mathrm{i}k_0\eta x} \tag{8.8}$$

Then, according to (8.1), we assume for the magnetic field:

$$\begin{aligned} H_x &= -V(z)\,\mathrm{e}^{\mathrm{i}k_0\eta x} \\ H_z &= -W(z)\,\mathrm{e}^{\mathrm{i}k_0\eta x} \end{aligned} \tag{8.9}$$

Finally, from (8.1) and (8.2), we obtain the following system of equations for the field amplitudes:

$$\begin{aligned} \frac{\mathrm{d}U}{\mathrm{d}z} &= \mathrm{i}\omega\mu_0 V \\ \frac{\mathrm{d}V}{\mathrm{d}z} &= \mathrm{i}\omega\varepsilon_0\left(\varepsilon - \eta^2\right)U \\ \mu_0 W + \frac{\eta}{c}U &= 0 \end{aligned} \tag{8.10}$$

The system of (8.10) allows us to calculate the field amplitudes at any point of the medium. Hence, it will give access to the calculation of transmittance and reflectance. As we are only interested in the intensities transmitted through or reflected from the surfaces, we will finally only need to calculate the z-component of Poynting's vector. Therefore, it will be sufficient for us to calculate the horizontal components of the fields, so that from (8.10) we will have to regard only the first two equations.

Let us finally write out the particular wave equations for U and V. Differentiating the first two equations in (8.10) with respect to the coordinates, we obtain:

$$\begin{aligned} &\frac{\mathrm{d}^2U}{\mathrm{d}z^2} + \frac{\omega^2}{c^2}(\varepsilon(z) - \eta^2)U = 0 \\ &\frac{\mathrm{d}^2V}{\mathrm{d}z^2} - \frac{1}{(\varepsilon(z) - \eta^2)}\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\frac{\mathrm{d}V}{\mathrm{d}z} + \frac{\omega^2}{c^2}(\varepsilon(z) - \eta^2)V = 0 \end{aligned} \tag{8.11}$$

We will have to return to (8.11) later when deriving the utmost important matrix method for calculating T and R for multilayer stacks. But before doing so, let us see how the corresponding equations for the p-polarization will look like.

8.1.3 *p*-Polarization

The calculations for p-polarization are similar to those for s-polarization. In the p-polarized case, we have:

$$\boldsymbol{E} = \begin{pmatrix} E_x \\ 0 \\ E_z \end{pmatrix}; \qquad \boldsymbol{H} = \begin{pmatrix} 0 \\ H_y \\ 0 \end{pmatrix}$$

We will not repeat the full calculation, but only mention the main differences and the final results. The main difference is, that it is convenient to interchange the roles of E and H when comparing with the s-case. So that instead of (8.4), we assume:

$$H_y(x, z) = X(x)\,U(z)$$

Instead of (8.8) and (8.9), we now find:

$$\begin{aligned} H_y &= U(z)\,\mathrm{e}^{\mathrm{i}k_0\eta x} \\ E_x &= V(z)\,\mathrm{e}^{\mathrm{i}k_0\eta x} \\ E_z &= W(z)\,\mathrm{e}^{\mathrm{i}k_0\eta x} \end{aligned}$$

That results in the system of equations:

$$\begin{aligned} \frac{\mathrm{d}U}{\mathrm{d}z} &= \mathrm{i}\omega\varepsilon\varepsilon_0 V \\ \frac{\mathrm{d}V}{\mathrm{d}z} &= \mathrm{i}\omega\mu_0\left(1 - \frac{\eta^2}{\varepsilon}\right)U \end{aligned} \tag{8.12}$$

The corresponding wave equations are:

$$\begin{aligned} &\frac{\mathrm{d}^2U}{\mathrm{d}z^2} - \frac{1}{\varepsilon}\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\frac{\mathrm{d}U}{\mathrm{d}z} + \frac{\omega^2}{c^2}(\varepsilon(z) - \eta^2)U = 0 \\ &\frac{\mathrm{d}^2V}{\mathrm{d}z^2} - \frac{\eta^2}{\varepsilon(z)(\varepsilon(z) - \eta^2)}\frac{\mathrm{d}\varepsilon}{\mathrm{d}z}\frac{\mathrm{d}V}{\mathrm{d}z} + \frac{\omega^2}{c^2}(\varepsilon(z) - \eta^2)V = 0 \end{aligned} \tag{8.13}$$

8.1.4 Calculation of Transmittance and Reflectance

Let us now come to the calculation of transmittance and reflectance of a stratified medium, which is deposited on a substrate. We start with some formal transformations of the equations derived so far. First of all, let us determine the physical sense of the value η. From (8.7) or (8.8), it becomes clear that the product ηk_0 must be equal to the x-component of the wavevector $\boldsymbol{k}$. We therefore have:

$$k_x = \frac{\omega}{c} n \sin\psi = \frac{\omega}{c}\eta \;\; \Rightarrow \;\; \eta = n \sin\psi$$

where ψ is the propagation angle in the stratified medium. Both ψ and n are now dependent on the z-coordinate. The condition (8.6) has the meaning of:

$$n \sin\psi = \text{const.}$$

and is therefore a generalization of Snell's law of refraction to a medium with a continuously changing refractive index. When the incidence medium has the refractive index 1, of course, $\eta = \sin\varphi$, where φ is again the angle of incidence.

We will now modify (8.10) and (8.12), in order to come to functions u and v that have the same dimension, although they represent different kinds of fields. For this purpose, we construct the following functions:

$$s\text{-polarization:} \qquad u = U\,; \quad v = \sqrt{\frac{\mu_0}{\varepsilon_0}}V \tag{8.14}$$

$$p\text{-polarization:} \qquad u = U\,; \quad v = \sqrt{\frac{\varepsilon_0}{\mu_0}}V \tag{8.15}$$

Then, instead of (8.10) and (8.12), we find the simplified equations:

$$s\text{-polarization:} \qquad \begin{aligned} \frac{\mathrm{d}u}{\mathrm{d}z} &= \mathrm{i}k_0 v \\ \frac{\mathrm{d}v}{\mathrm{d}z} &= \mathrm{i}k_0 \left(\varepsilon - \eta^2\right) u \end{aligned} \tag{8.16}$$

$$p\text{-polarization:} \qquad \begin{aligned} \frac{\mathrm{d}u}{\mathrm{d}z} &= \mathrm{i}k_0 \varepsilon v \\ \frac{\mathrm{d}v}{\mathrm{d}z} &= \mathrm{i}k_0 \left(1 - \frac{\eta^2}{\varepsilon}\right) u \end{aligned} \tag{8.17}$$

The advantage to (8.10) and (8.12) is, that the functions u and v have identical dimension. Hence, in the mathematical treatment of (8.16) and (8.17), we may regard these functions as dimensionless. Of course, the substitutions (8.14) and (8.15) cause no changes in (8.11) and (8.13).

Let us now come to the field transmission and reflection coefficients. In the case of s-polarization, u corresponds to the electric field (in fact, its y-component). The value V was connected to the x-component of the H-field. This is the full H-field, multiplied with $\cos\psi$. From (6.9) and (8.14), we see that

$$v = \sqrt{\frac{\mu_0}{\varepsilon_0}} V = \pm \hat{n} E_y \cos\psi \tag{8.18}$$

Let us assume the case of an incident medium with a refractive index n_1. In the incidence medium, the propagation angle ψ is identical to the incidence angle φ. Similar to the discussion in Chap. 6, at the ambient-film interface, we have:

$$\begin{aligned} u\,(z=0) &\equiv u_0 = E^{(e)} + E^{(r)} \\ v\,(z=0) &\equiv v_0 = \hat{n}_1 \cos\varphi \left(E^{(e)} - E^{(r)} \right) \end{aligned} \tag{8.19}$$

The "–" in the second equation reflects the fact, that the z-component of the wavevector of the reflected wave has the opposite sign than that of the impinging one. Therefore, from (8.18), the corresponding electric fields are directed antiparallel. We now use the definitions:

$$t = \frac{E^{(t)}}{E^{(e)}}\,; \quad r = \frac{E^{(r)}}{E^{(e)}}$$

Let us regard the special case $E^{(t)} = 1$. That does not change the values of t and r. We then obtain from (8.19):

$$t = \frac{2\hat{n}_1 \cos\varphi}{u_0 \hat{n}_1 \cos\varphi + v_0}\,; \quad r = \frac{u_0 \hat{n}_1 \cos\varphi - v_0}{u_0 \hat{n}_1 \cos\varphi + v_0} \tag{8.20}$$

The values of u_0 and v_0 correspond to u and v at the air/film interface.

Equations (8.20) are similar by structure to the Fresnel's coefficients obtained in Chap. 6. They allow to calculate the reflectance and transmittance through an inhomogeneous film for the case of s-polarization. However, first of all one has to solve the system of equations (8.16) or alternatively (8.11). To do so, we still need knowledge about the boundary conditions at the film/substrate boundary.

Let us for a moment assume that there is no rear substrate surface. The values t and r as calculated by (8.20) are then a simple generalization of the field transmission coefficients t_{123} and r_{123}, as introduced in Chap. 7. Consequently, in the substrate, we only have one transmitted wave with the electrical field $E^{(t)} = 1$. From there it follows, that at the film/substrate boundary, we have:

$$\begin{aligned} u &= 1 \\ v &= \hat{n}_s \cos\varphi_s \end{aligned}$$

where the second condition is again a consequence from (6.9) and (8.14), φ_s is the propagation angle in the substrate.

Finally, the transmittance and reflectance are calculated in the usual way, namely by:

$$T = \frac{\mathrm{Re}\,(\hat{n}_s \cos\varphi_s)}{\mathrm{Re}\,(\hat{n}_1 \cos\varphi)}|t|^2;\ R = |r|^2$$

In order to account for the rear side of the substrate, we use equations (7.25) and (7.26). The values t_{123} and r_{123} have to be replaced by t and r according to equations (8.20). In order to account for t_{321} and r_{321}, u and v have to be recalculated assuming the substrate medium as the incident one, and the outer space as the substrate. Then, t and r again follow from (8.20) when replacing the former incidence parameters by the substrate ones.

This might seem to be a troublesome calculation, nevertheless it offers a straightforward way to calculate the optical spectra of media with an arbitrary $n(z)$-dependence. Of course, as before, the index of refraction may be complex and depend on the wavelength, so that the method is indeed very general.

Before coming to the case of p-polarization, let us look at two examples, which correspond to normal incidence, so that the polarization state is of no significance.

1. Example

Let us return to the spectra shown in Fig. 7.11. This figure shows the T- and R-spectra of a gradient index layer, deposited on fused silica. The interesting point was, that in the case of a gradient index layer, the film's transmittance and reflectance values are not equal to those of the bare substrate in the halfwave points. Up to now, we had no theory to verify this point. By means of the theory derived in this chapter, such calculations should become accessible.

Figure 8.1 shows the result of a model calculation, where we assumed a 300 nm thick film with a z-dependent refractive index. For simplicity, both dispersion and absorption have been neglected. We regarded two cases, namely a negative and a positive gradient, while the average dielectric function should be the same for both gradients. For this model calculation, the following z-dependence of the dielectric function has been assumed:

$$\begin{aligned}\text{positive gradient}: &\quad \varepsilon(z) = 4.9 - 0.003\,\mathrm{nm}^{-1} z\\ \text{negative gradient}: &\quad \varepsilon(z) = 4.0 + 0.003\,\mathrm{nm}^{-1} z\end{aligned}$$

The result is presented in Fig. 8.1.

We see, that in the halfwave points, the transmittance is either higher or lower than the substrate transmittance, depending on the sign of the gradient. This confirms the general rules formulated in Table 7.1. On the other hand, the quarterwave points are completely insensitive to the film index gradient, they only depend on the average index, which is the same for both gradient layers.

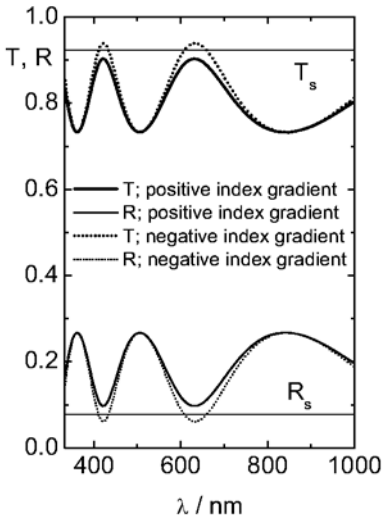

Fig. 8.1. Calculated spectra of 300 nm thick gradient index films on glass

2. Example

In the second example, we regard another case of an index gradient, namely a refractive index which depends on z according to a sinus-function. Hence, we have a periodical change in refractive index according to the law:

$$\hat{n} = \hat{n}\,(z) = \langle \hat{n} \rangle + \Delta\hat{n} \sin\left(\frac{4\pi z}{\Lambda_z}\right)$$

Note that the value Λ_z is not the period, but twice the period of the refractive index profile. $\langle \hat{n} \rangle$ is the spatially averaged index of refraction, and $\Delta\hat{n}$ determines the modulation depth. Thin film systems with a suchlike refractive index profile are called rugate filters.

Practically, such a refractive index profile is difficult to be prepared. But one may come close to such a profile when mixing two optical materials with a sinusoidal filling factor $p = p(z)$. Figure 8.2 displays the calculated T- and R-spectra of a 1500 nm thick film on fused silica. In this calculation, we assumed a nearly sinusoidal refractive index profile, obtained from (4.11) with

$$p = p\,(z) = \frac{1 + \sin\left(\frac{4\pi z}{\Lambda_z}\right)}{2}\,; \quad \Lambda_z = 300\,\mathrm{nm}$$

with silicon dioxide SiO_2 and niobium pentoxide Nb_2O_5 as mixing partners. The material is thus assumed to be a mixture from a low- and a high refractive index material with a continuously varying filling factor. The spectra in Fig. 8.2 show a well distinguished reflection peak centred at:

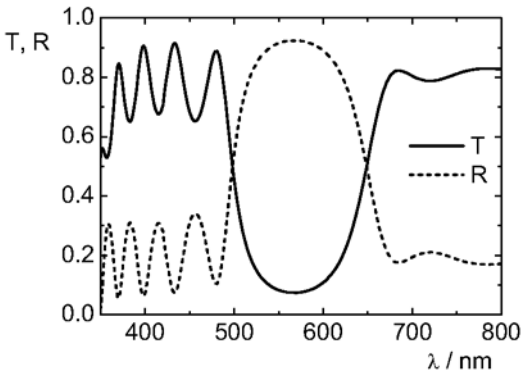

Fig. 8.2. Calculated spectra of a thin film with a nearly sinusoidal refractive index profile. The film thickness corresponds to 10 periods

$$\lambda_{reject} = \langle n \rangle \Lambda_z$$

Obviously, such systems may really find application as filters.

Let us now come to the case of p-polarization. In difference to s-polarization, the function u in (8.17) does now correspond to the magnetic field, while v is connected to the electric field. Keeping this in mind, the calculations may be carried out in an analogous way to the case of s-polarization.

Instead of (8.18), we now have:

$$v = \sqrt{\frac{\varepsilon_0}{\mu_0}} V = \pm \frac{H_y}{\hat{n}} \cos\psi$$

Correspondingly, at the ambient/film interface, we find the conditions:

$$u_0 = H^{(e)} + H^{(r)}$$
$$v_0 = \frac{\cos\varphi}{\hat{n}_1} \left(H^{(e)} - H^{(r)} \right)$$

We see, that the structure of the equations is identical to the s-polarization case, with the only difference that the terms $n \cos\varphi$ have to be replaced by $\cos\varphi / n$. Instead of the familiar electric field transmission and reflection coefficients, it is now more convenient to define magnetic field transmission and reflection coefficients according to:

$$t = \frac{H^{(t)}}{H^{(e)}}; \quad r = \frac{H^{(r)}}{H^{(e)}}$$

with the simplifying assumption of $H^{(t)} = 1$. That leads us to the result:

$$t = \frac{2\cos\varphi}{u_0 \cos\varphi + \hat{n}_1 v_0}; \quad r = \frac{u_0 \cos\varphi - \hat{n}_1 v_0}{u_0 \cos\varphi + \hat{n}_1 v_0} \tag{8.21}$$

where u and v are again taken at the film/ambient interface. So that again, one has to solve (8.17) with the boundary conditions at the film/substrate interface given by:

$$u = 1$$
$$v = \frac{\cos\varphi_s}{\hat{n}_s}$$

Having calculated u and v, the field transmission and reflection coefficients (for the magnetic fields!!) are calculated from (8.21). Coming to the intensities, one must keep in mind that one should never utilize expressions like (6.19), because they are only valid for electric field transmission coefficients. This is most easily to be seen for real refractive indices. In fact, the intensity is then proportional to:

$$I \propto n\cos\varphi|E|^2 \propto n\cos\varphi\left|\frac{H}{n}\right|^2 = \frac{\cos\varphi}{n}|H|^2$$

Therefore, whenever we deal with magnetic field transmission coefficients, we obtain for the intensity coefficients:

$$T = \frac{\mathrm{Re}\left(\frac{\cos\varphi_s}{\hat{n}_s}\right)}{\mathrm{Re}\left(\frac{\cos\varphi}{\hat{n}_1}\right)}|t|^2; \quad R = |r|^2 \tag{8.22}$$

In order to get a systematic overview on the calculation strategy, the main steps are summarized in Table 8.1.

8.2 Multilayer Systems

8.2.1 The Characteristic Matrix

We will now turn to a calculation method that is utmost important in thin film spectroscopy: the matrix method. Again, we start with the mathematical derivation of the material. The derivation will be performed for s-polarization, the p-polarization analogon will be shortly treated at the end of the section.

We start from the system of (8.16):

s-polarization:

$$\frac{\mathrm{d}u}{\mathrm{d}z} = \mathrm{i}k_0 v$$
$$\frac{\mathrm{d}v}{\mathrm{d}z} = \mathrm{i}k_0\left(\varepsilon - \eta^2\right)u$$

Let us assume that the film-air interface corresponds to the z-value $z = 0$. Our task is to find a solution of (8.16) with:

Table 8.1. Calculation of transmission and reflection of an arbitrary gradient index layer on a semininfinite substrate

	s-polarization	p-polarization
Meaning of u	Electric field	Magnetic field
Meaning of v	Magnetic field	Electric field
System of equations with $\varepsilon = \varepsilon(z)$	$\frac{\mathrm{d}u}{\mathrm{d}z} = \mathrm{i}k_0 v$ $\frac{\mathrm{d}v}{\mathrm{d}z} = \mathrm{i}k_0 \left(\varepsilon - \eta^2\right) u$	$\frac{\mathrm{d}u}{\mathrm{d}z} = \mathrm{i}k_0 \varepsilon v$ $\frac{\mathrm{d}v}{\mathrm{d}z} = \mathrm{i}k_0 \left(1 - \frac{\eta^2}{\varepsilon}\right) u$
Boundary conditions at the film/substrate interface	$u = 1$ $v = \hat{n}_s \cos\varphi_s$	$u = 1$ $v = \frac{\cos\varphi_s}{\hat{n}_s}$
Definition of field transmission and reflection coefficients	$t = \frac{E^{(t)}}{E^{(e)}}$ $r = \frac{E^{(r)}}{E^{(e)}}$	$t = \frac{H^{(t)}}{H^{(e)}}$ $r = \frac{H^{(r)}}{H^{(e)}}$
Expressions for field transmission and reflection coefficients	$t = \frac{2\hat{n}_1 \cos\varphi}{u_0 \hat{n}_1 \cos\varphi + v_0}$ $r = \frac{u_0 \hat{n}_1 \cos\varphi - v_0}{u_0 \hat{n}_1 \cos\varphi + v_0}$	$t = \frac{2\cos\varphi}{u_0 \cos\varphi + \hat{n}_1 v_0}$ $r = \frac{u_0 \cos\varphi - \hat{n}_1 v_0}{u_0 \cos\varphi + \hat{n}_1 v_0}$
Intensity coefficients for a film on a semi-infinite substrate	$T = \frac{\mathrm{Re}\left(\hat{n}_s \cos\varphi_s\right)}{\mathrm{Re}\left(\hat{n}_1 \cos\varphi\right)} \lvert t\rvert^2$ $R = \lvert r\rvert^2$	$T = \frac{\mathrm{Re}\left(\frac{\cos\varphi_s}{\hat{n}_s}\right)}{\mathrm{Re}\left(\frac{\cos\varphi}{\hat{n}_1}\right)} \lvert t\rvert^2$ $R = \lvert r\rvert^2$
Effect of rear substrate side	In full analogy to (7.25) and (7.26)	

$$\begin{aligned} u(0) &= u_0 \\ v(0) &= v_0 \end{aligned} \tag{8.23}$$

Let us further suppose that we already know two particular solutions of the system (8.16), corresponding to special boundary conditions, namely:

$$\begin{aligned} u_1(z) \quad &\text{with} \quad u_1(0) = 1 \\ v_1(z) \quad &\text{with} \quad v_1(0) = 0 \end{aligned}$$

and

$$\begin{aligned} u_2(z) \quad &\text{with} \quad u_2(0) = 0 \\ v_2(z) \quad &\text{with} \quad v_2(0) = 1 \end{aligned}$$

From (8.16) it is obvious that

$$v_1 \frac{\mathrm{d}u_2}{\mathrm{d}z} - v_2 \frac{\mathrm{d}u_1}{\mathrm{d}z} = u_1 \frac{\mathrm{d}v_2}{\mathrm{d}z} - u_2 \frac{\mathrm{d}v_1}{\mathrm{d}z} = 0$$

and hence

$$\frac{\mathrm{d}}{\mathrm{d}z}(u_1 v_2 - u_2 v_1) = 0 \;\Rightarrow\; u_1 v_2 - u_2 v_1 = \text{const.} = 1 \tag{8.24}$$

On the other hand, due to the principle of superposition, the solution of (8.16) with the boundary conditions (8.23) may be written as:

$$u(z) = u_1(z)\, u_0 + u_2(z)\, v_0$$

$$v(z) = v_1(z)\, u_0 + v_2(z)\, v_0$$

or

$$\begin{pmatrix} u(z) \\ v(z) \end{pmatrix} = \begin{pmatrix} u_1(z) & u_2(z) \\ v_1(z) & v_2(z) \end{pmatrix} \begin{pmatrix} u_0 \\ v_0 \end{pmatrix} \tag{8.25}$$

Inverting (8.25) and using (8.24), we obtain:

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = \hat{\boldsymbol{M}} \begin{pmatrix} u(z) \\ v(z) \end{pmatrix} = \begin{pmatrix} v_2(z) & -u_2(z) \\ -v_1(z) & u_1(z) \end{pmatrix} \begin{pmatrix} u(z) \\ v(z) \end{pmatrix} \tag{8.26}$$

The matrix

$$\hat{\boldsymbol{M}} = \hat{\boldsymbol{M}}(z) \equiv \begin{pmatrix} v_2(z) & -u_2(z) \\ -v_1(z) & u_1(z) \end{pmatrix} \tag{8.27}$$

is called the characteristic matrix of the film.

As it follows from (8.26), by means of the characteristic matrix it becomes possible to relate the electric and magnetic fields at $z = 0$ to those at any other z-value in the film. Of course, having the fields, the calculation of transmittance and reflectance is possible as described in Sect. 8.1. Therefore, the knowledge of the characteristic matrix is sufficient to describe the optical properties of any medium with a dielectric function $\varepsilon = \varepsilon(z)$.

For p-polarization, (8.26) and (8.27) are valid as well, with the only difference that u and v are now the solutions of (8.17).

8.2.2 Characteristic Matrix of a Single Homogeneous Film

Let us calculate the characteristic matrix of a single homogeneous film ($n \neq n(z)$). This is an important special case. For s-polarization, from (8.11) we find:

$$u'' + \frac{\omega^2}{c^2}(\varepsilon - \eta^2)u = 0$$
$$v'' + \frac{\omega^2}{c^2}(\varepsilon - \eta^2)v = 0$$

with

$$(\varepsilon - \eta^2) = \hat{n}^2 \cos^2 \psi$$

The solutions that are consistent with (8.16) may be written as:

$$\begin{aligned} u_1 &= \cos(k_0 \hat{n} z \cos\psi) \\ v_1 &= \mathrm{i}\hat{n} \cos\psi \sin(k_0 \hat{n} z \cos\psi) \\ u_2 &= \frac{\mathrm{i}}{\hat{n}\cos\psi} \sin(k_0 \hat{n} z \cos\psi) \\ v_2 &= \cos(k_0 \hat{n} z \cos\psi) \end{aligned}$$

The characteristic matrix becomes:

$$\begin{aligned} \hat{\boldsymbol{M}}(z) &\equiv \begin{pmatrix} v_2(z) & -u_2(z) \\ -v_1(z) & u_1(z) \end{pmatrix} \\ &= \begin{pmatrix} \cos(k_0 \hat{n} z \cos\psi) & -\dfrac{\mathrm{i}}{\hat{n}\cos\psi} \sin(k_0 \hat{n} z \cos\psi) \\ -\mathrm{i}\hat{n} \cos\psi \sin(k_0 \hat{n} z \cos\psi) & \cos(k_0 \hat{n} z \cos\psi) \end{pmatrix} \end{aligned} \tag{8.28}$$

Expression (8.28) is valid for s-polarization. For p-polarization, one obtains in the same way:

$$\begin{aligned} \hat{\boldsymbol{M}}(z) &\equiv \begin{pmatrix} v_2(z) & -u_2(z) \\ -v_1(z) & u_1(z) \end{pmatrix} \\ &= \begin{pmatrix} \cos(k_0 \hat{n} z \cos\psi) & -\dfrac{\mathrm{i}\hat{n}}{\cos\psi} \sin(k_0 \hat{n} z \cos\psi) \\ -\mathrm{i}\dfrac{\cos\psi}{\hat{n}} \sin(k_0 \hat{n} z \cos\psi) & \cos(k_0 \hat{n} z \cos\psi) \end{pmatrix} \end{aligned} \tag{8.29}$$

8.2.3 Characteristic Matrix of a Film Stack

Let us now assume, that instead of a single homogeneous film, we have a stack of a certain number N of homogeneous films, each of them having a thickness d_j and a refractive index n_j. Let us count the layers starting from the incident medium side. The first film extends from $z = 0$ to $z = z_1$, hence

$d_1 = z_1 - 0 = z_1$. Correspondingly, the second film extends from $z = z_1$ to $z = z_2$, hence $d_2 = z_2 - z_1$, and so on. What we obtain from (8.26) is a recursive recipe to calculate the characteristic matrix of the stack:

$$\begin{aligned}\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} &= \hat{\boldsymbol{M}}_1(z_1) \begin{pmatrix} u(z_1) \\ v(z_1) \end{pmatrix} = \hat{\boldsymbol{M}}_1(z_1)\, \hat{\boldsymbol{M}}_2(z_2 - z_1) \begin{pmatrix} u(z_2) \\ v(z_2) \end{pmatrix} = \dots \\ &= \hat{\boldsymbol{M}}_1(d_1)\, \hat{\boldsymbol{M}}_2(d_2) \dots \hat{\boldsymbol{M}}_N(d_N) \begin{pmatrix} u(z_N) \\ v(z_N) \end{pmatrix}\end{aligned}$$

We come to the result, that the whole stack is again characterized by a 2×2-matrix, which is obtained by simple multiplication of the characteristic matrices of the single films. Hence,

$$\hat{\boldsymbol{M}}_{stack} \equiv \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \prod_{j=1}^{N} \hat{\boldsymbol{M}}_j(d_j) \tag{8.30}$$

8.2.4 Calculation of Transmittance and Reflectance

In order to calculate transmittance and reflectance of a stack, the only thing that remains to be done is to relate the four elements of the characteristic matrix of the stack to the transmittance and reflectance.

s-polarization:

As in Sect. 8.1, we have at the stack/substrate border:

$$\begin{aligned} u &= u(z_N) = 1 \\ v &= v(z_N) = \hat{n}_s \cos\varphi_s \end{aligned}$$

From (8.20), it follows, that:

$$t = \frac{2\hat{n}_1 \cos\varphi}{u_0 \hat{n}_1 \cos\varphi + v_0}; \quad r = \frac{u_0 \hat{n}_1 \cos\varphi - v_0}{u_0 \hat{n}_1 \cos\varphi + v_0}$$

while from (8.26) and (8.30):

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} u(z_N) \\ v(z_N) \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 \\ \hat{n}_s \cos\varphi_s \end{pmatrix}$$

Therefore,

$$\begin{aligned} u_0 &= m_{11} + m_{12} \hat{n}_s \cos\varphi_s \\ v_0 &= m_{21} + m_{22} \hat{n}_s \cos\varphi_s \end{aligned}$$

and thus:

$$\begin{aligned} t &= \frac{2\hat{n}_1 \cos\varphi}{(m_{11} + m_{12}\hat{n}_s \cos\varphi_s)\, \hat{n}_1 \cos\varphi + m_{21} + m_{22}\hat{n}_s \cos\varphi_s} \\ r &= \frac{(m_{11} + m_{12}\hat{n}_s \cos\varphi_s)\, \hat{n}_1 \cos\varphi - (m_{21} + m_{22}\hat{n}_s \cos\varphi_s)}{(m_{11} + m_{12}\hat{n}_s \cos\varphi_s)\, \hat{n}_1 \cos\varphi + m_{21} + m_{22}\hat{n}_s \cos\varphi_s} \end{aligned} \tag{8.31}$$

The intensity coefficients are obtained in the usual way.

p-polarization:

Accordingly, for p-polarization we have:

$$u(z_N) = 1$$
$$v(z_N) = \frac{\cos\varphi_s}{\hat{n}_s}$$

and

$$t = \frac{2\frac{\cos\varphi}{\hat{n}_1}}{u_0\frac{\cos\varphi}{\hat{n}_1} + v_0}; \quad r = \frac{u_0\frac{\cos\varphi}{\hat{n}_1} - v_0}{u_0\frac{\cos\varphi}{\hat{n}_1} + v_0}$$

From

$$\begin{pmatrix} u_0 \\ v_0 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} u(z_N) \\ v(z_N) \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{\cos\varphi_s}{\hat{n}_s} \end{pmatrix}$$

we find:

$$u_0 = m_{11} + m_{12}\frac{\cos\varphi_s}{\hat{n}_s}$$
$$v_0 = m_{21} + m_{22}\frac{\cos\varphi_s}{\hat{n}_s}$$

So that t and r become:

$$\begin{aligned} t &= \frac{2\frac{\cos\varphi}{\hat{n}_1}}{\left(m_{11} + m_{12}\frac{\cos\varphi_s}{\hat{n}_s}\right)\frac{\cos\varphi}{\hat{n}_1} + \left(m_{21} + m_{22}\frac{\cos\varphi_s}{\hat{n}_s}\right)} \\ r &= \frac{\left(m_{11} + m_{12}\frac{\cos\varphi_s}{\hat{n}_s}\right)\frac{\cos\varphi}{\hat{n}_1} - \left(m_{21} + m_{22}\frac{\cos\varphi_s}{\hat{n}_s}\right)}{\left(m_{11} + m_{12}\frac{\cos\varphi_s}{\hat{n}_s}\right)\frac{\cos\varphi}{\hat{n}_1} + \left(m_{21} + m_{22}\frac{\cos\varphi_s}{\hat{n}_s}\right)} \end{aligned} \tag{8.32}$$

When calculating intensity transmission and reflection coefficients for p-polarization, one must again keep in mind that the field coefficients t and r express the relation between magnetic fields. For T- and R-calculation, one has therefore to use (8.22).

Table 8.2 resumes the main steps for calculating the optical spectra of a stack by means of the matrix method.

The matrix method offers the possibility to calculating the spectral properties of numerous practically relevant thin film systems, such as high reflectors, antireflection coatings, and others. Generally, the capacitance of the equations fixed in Table 8.2 is high enough to fill complete monographs which exclusively deal with questions of optical thin film systems design. We emphasize, that this is not the purpose of this book. The interested reader is therefore referred to the specialized literature on this subject. Nevertheless, some special systems that might be not considered in the optical coatings design literature will be addressed in Chap. 9. Some simpler examples are included into the problems to Chaps. 6–9.

Table 8.2. Calculation of transmission and reflection of an arbitrary thin film stack on a semininfinite substrate

s-polarization	p-polarization
M, single film	
$\begin{pmatrix} \cos(k_0\hat{n}d\cos\psi) & -\frac{\mathrm{i}}{\hat{n}\cos\psi}\sin(k_0\hat{n}d\cos\psi) \\ -\mathrm{i}\hat{n}\cos\psi\sin(k_0\hat{n}d\cos\psi) & \cos(k_0\hat{n}d\cos\psi) \end{pmatrix}$	$\begin{pmatrix} \cos(k_0\hat{n}d\cos\psi) & -\frac{\mathrm{i}\hat{n}}{\cos\psi}\sin(k_0\hat{n}d\cos\psi) \\ -\mathrm{i}\frac{\cos\psi}{\hat{n}}\sin(k_0\hat{n}d\cos\psi) & \cos(k_0\hat{n}d\cos\psi) \end{pmatrix}$
M, stack	
$\hat{\boldsymbol{M}}_{stack} \equiv \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} = \prod_{j=1}^{N} \hat{\boldsymbol{M}}_j(d_j)$	
Field transmission and reflection coefficients	
$t = \frac{E^{(t)}}{E^{(e)}}; \quad r = \frac{E^{(r)}}{E^{(e)}}$	$t = \frac{H^{(t)}}{H^{(e)}}; \quad r = \frac{H^{(r)}}{H^{(e)}}$
Expressions for field transmission and reflection coefficients	
$t = \frac{2\hat{n}_1\cos\varphi}{(m_{11}+m_{12}\hat{n}_s\cos\varphi_s)\hat{n}_1\cos\varphi + m_{21} + m_{22}\hat{n}_s\cos\varphi_s}$ $r = \frac{(m_{11}+m_{12}\hat{n}_s\cos\varphi_s)\hat{n}_1\cos\varphi - (m_{21}+m_{22}\hat{n}_s\cos\varphi_s)}{(m_{11}+m_{12}\hat{n}_s\cos\varphi_s)\hat{n}_1\cos\varphi + m_{21} + m_{22}\hat{n}_s\cos\varphi_s}$	$t = \frac{2\frac{\cos\varphi}{\hat{n}_1}}{\left(m_{11}+m_{12}\frac{\cos\varphi_s}{\hat{n}_s}\right)\frac{\cos\varphi}{\hat{n}_1} + \left(m_{21}+m_{22}\frac{\cos\varphi_s}{\hat{n}_s}\right)}$ $r = \frac{\left(m_{11}+m_{12}\frac{\cos\varphi_s}{\hat{n}_s}\right)\frac{\cos\varphi}{\hat{n}_1} - \left(m_{21}+m_{22}\frac{\cos\varphi_s}{\hat{n}_s}\right)}{\left(m_{11}+m_{12}\frac{\cos\varphi_s}{\hat{n}_s}\right)\frac{\cos\varphi}{\hat{n}_1} + \left(m_{21}+m_{22}\frac{\cos\varphi_s}{\hat{n}_s}\right)}$
Intensity coefficients for a film on a semi-infinite substrate	
$T = \frac{\mathrm{Re}(\hat{n}_s\cos\varphi_s)}{\mathrm{Re}(\hat{n}_1\cos\varphi)}\lvert t\rvert^2; \quad R = \lvert r\rvert^2$	$T = \frac{\mathrm{Re}\left(\frac{\cos\varphi_s}{\hat{n}_s}\right)}{\mathrm{Re}\left(\frac{\cos\varphi}{\hat{n}_1}\right)}\lvert t\rvert^2; \quad R = \lvert r\rvert^2$
Effect of rear substrate side In full analogy to (7.25) and (7.26)	

9 Special Geometries

9.1 Quarterwave Stacks and Derived Systems

This chapter will conclude the second part of this book. It will deal with a few special cases of multilayer systems that may be described in terms of the theory derived in Chap. 8. Additionally, a qualitative treatment of narrowband filters and absorbers based on the so-called Resonant Grating Waveguide Structures (GWS) will be given.

Let us start with the mathematically simple case of a quarterwave stack. As we have already mentioned in Chap. 7, at a certain wavelength, a non-absorbing layer may act as a quarterwave layer at a given reference wavelength λ_0 supposed that condition:

$$nd = \frac{\lambda_0}{4} \tag{9.1}$$

is fulfilled (normal incidence). Let us now assume, that we have a multilayer stack built from alternating high- and low-refractive index layers with corresponding refractive indices n_1 and n_2 and an optical thickness determined by (9.1). In this case, all layers will behave as quarterwave layers at the same reference wavelength λ_0. Let us see how the reflectance of such a system will look like.

If the identical sequence of the pair of high- and low-refractive index quarterwave layers is repeated for N times, then, for normal incidence, the matrix of the stack becomes:

$$\mathbf{M} = \begin{pmatrix} \left(-\frac{n_2}{n_1}\right)^N & 0 \\ 0 & \left(-\frac{n_1}{n_2}\right)^N \end{pmatrix}$$

From (8.32), we find the reflectance with air as incidence medium:

$$R = |r|^2 = \left| \frac{n_s - \left(\frac{n_1}{n_2}\right)^{2N}}{n_s + \left(\frac{n_1}{n_2}\right)^{2N}} \right|^2$$

It is obvious, that for any pair of refractive indices $n_1 \neq n_2$, we get:

$$\lim_{N\to\infty} R = 1$$

Therefore, such quarterwave stacks may find applications as high reflectors (dielectric reflectors) in the vicinity of the reference wavelength.

The previous discussion regarded the case of an even number of layers. The general conclusion on high reflectivity is also valid for an odd number of quarterwave layers, as it may easily be checked by the reader himself.

The previous derivation was of purely mathematical nature. But behind the mathematics, there is a simple physical mechanism that leads to the high reflectivity: Transmitted waves interfere destructively, while reflected waves interfere constructively. The situation is vice versa in antireflection coatings.

Having understood the working principle of a dielectric reflector, it is also easy to understand the general construction principle of a narrow bandpass filter. Let us start from a dielectric reflector built from alternating high refractive index quarterwave layers (H) and low refractive index quarterwave layers (L). Let us regard a sequence of quarterwave layers that is formally written as:

$$\text{air } |H\ (LH)^N| \text{ substrate} \qquad \text{(stack1)}$$

That means, that the multilayer stack starts with a high index quarter wave layer at the air side of the stack, followed by a pair of a low- and a high index quarterwave layer that is repeated N times. Consequently, the full number of quarterwave layers in stack1 is $2N + 1$.

Supposing that N is an even number, the same layer sequence may be written as:

$$\begin{aligned} \text{air } |H\ (LH)^N| \text{ substrate} \quad &= \text{air } |(HL)^N\ H| \text{ substrate} \\ &= \text{air } |(HL)^{N/2}\ H\ (LH)^{N/2}| \text{ substrate} \end{aligned}$$

Let us now modify the stack. We introduce a further quarterwave high index layer in the centre of the stack and obtain the design:

$$\text{air } |(HL)^{N/2}\ HH\ (LH)^{N/2}| \text{ substrate} \qquad \text{(stack2)}$$

At the reference wavelength, the combination HH is obviously a single halfwave layer. Therefore, it has no optical effect, and may be removed. So we get:

$$\begin{aligned} &\text{air } |(HL)^{N/2}\ HH\ (LH)^{N/2}| \text{ substrate} = \text{air } |(HL)^{N/2}\ (LH)^{N/2}| \text{ substrate} \\ &\text{air } |(HL)^{N/2-1}\ HLLH\ (LH)^{N/2-1}| \text{ substrate} \end{aligned}$$

Again, the combination LL is a halfwave layer and may be removed. But this removal will create a new halfwave layer, and the process may be continued until there is no layer at all remaining. Consequently, at the reference wavelength, the system stack2 has the same transmission and reflection as

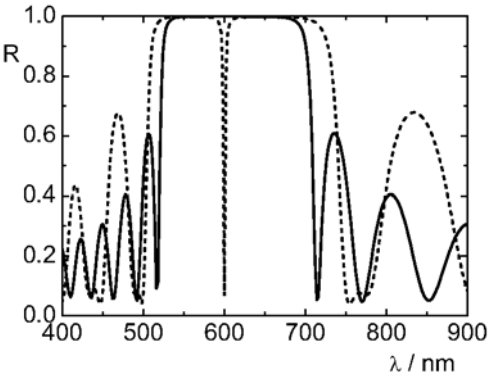

Fig. 9.1. Calculated reflection spectra of stack1 (*solid line*) and stack2 (*short dash*). In both cases, the reference wavelength is 600 nm

the air-substrate-interface. For a typical glass substrate, we therefore have to expect a transmittance in the region of 0.92.

So far, the considerations on high reflectance of stack1 and high transmittance of stack2 only concern the T and R values at the reference wavelength λ_0. Apart from this wavelength, T and/or R may be calculated in terms of the theoretical apparatus derived in Chap. 8.

Figure 9.1 shows the calculated reflectance of two model systems. In the first case (solid line), we have assumed a quarterwave stack according to stack1 with $N = 10$ (21 layers). The simulation has been carried out neglecting dispersion and absorption. We assumed a high refractive index of $n_H = 2.3$ and a low refractive index of $n_L = 1.5$ and a reference wavelength $\lambda_0 = 600\,\text{nm}$. The calculation confirms the expected high reflection in the vicinity of the reference wavelength. This region of high reflectance is sometimes called the rejection band.

The second spectrum (short dash) corresponds to a system like stack2, all parameters are the same as for stack1. The spectra look qualitatively similar, but at the reference wavelength, stack2 shows a sharp drop in the reflectance, corresponding to a narrow region of high transmittance. Hence, stack1 may work as a primitive version of a broadband reflector, while stack2 represents a crude narrowline transmission filter, which transmits light at 600 nm but blocks the radiation in the vicinity of the reference wavelength.

In practice, it may become highly desirable to suppress the sidelobes outside the high reflection regions shown in Fig. 9.1. For example, that might be necessary when the stack should work as an edge filter.

Although this is definitely not a book on thin film design, let us demonstrate the simplest method to flatten the reflectance characteristic at the long wavelength side of the rejection band. What we will obtain then is a longpass filter: It reflects the radiation at shorter wavelength, while the long wavelength region is transmitted. Again, we will start from stack1, but modify

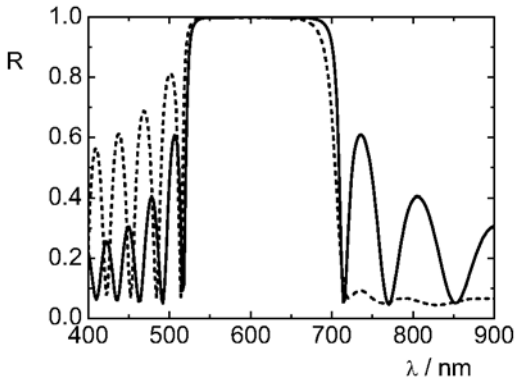

Fig. 9.2. Calculated reflection spectra of stack1 (*solid line*) and stack3 (*short dash*). In both cases, the reference wavelength is 600 nm

the first and the last layer of the stack: instead of quarterwave layers, the thickness of the outer layers will be chosen according to:

$$n_H d_H = \frac{\lambda_0}{8}$$

Then, instead of stack1, we obtain the design stack3:

$$\text{air } |0.5H\ (LH)^{N-1}\ L\ 0.5H| \text{ substrate} \qquad \text{(stack3)}$$

This system has a reflectance like that shown in Fig. 9.2 in short dash. In particular, one sees that the reflection sidelobes at the short wavelength edge of the spectrum are magnified, while those at the long wavelength edge are nearly completely suppressed. Therefore such as system may work as a long-wave pass filter.

Concluding this section, let us finally remark that the existence of a rejection band in a quarterwave stack is connected to the fundamental principles of wave propagation in periodic systems. In fact, any real quarterwave stack represents a truncated periodic arrangement with a periodic modulation of the refractive index. In such systems, destructive interference forbids the propagation of waves in certain spectral regions. Therefore, in these 'forbidden' zones, the transmittance approaches zero. For reasons of energy conservation, the reflectance must then approach the value 1. The same effect may be obtained for any periodic layer structure with a continuous refractive index profile, as it has been demonstrated in the reflectance curve from Fig. 8.2 where we dealt with rugate filters.

9.2 Extended Detail: Remarks on Resonant Grating Waveguide Structures

9.2.1 General Idea

In the previous subsection we have demonstrated, that destructive interference in transmission and constructive in reflection may lead to a stack reflectivity that approaches 100% when the number of layers becomes infinitive. In fact, the same high reflectivity may be achieved by a single layer design in a more subtle way, when the single layer is combined with a diffraction grating. This leads us to the so-called resonant Grating Waveguide Structures (GWS). You may note, that due to the grating, the system geometry is again periodic.

In its simplest version, a GWS is built up by a single high-refractive index layer (the waveguide layer) with a one-dimensional diffraction grating on top (see Fig. 9.3). It is essential (as will be shown below), that the film (waveguide) refractive index is higher than the indices of the ambient and the substrate. Then, for a sufficiently large propagation angle of the light, it will suffer total internal reflection at both film interfaces, so that the wave cannot leave the film, instead, it propagates in the waveguide.

We will restrict ourselves to the case that the incidence medium is air. Let us now try to obtain a qualitative understanding of the general function of a GWS.

As seen in Fig. 9.3, the incident irradiation impinges onto the diffraction grating. In the general case, this leads to the appearance of several modes of diffracted waves corresponding to different orders of diffraction. They will

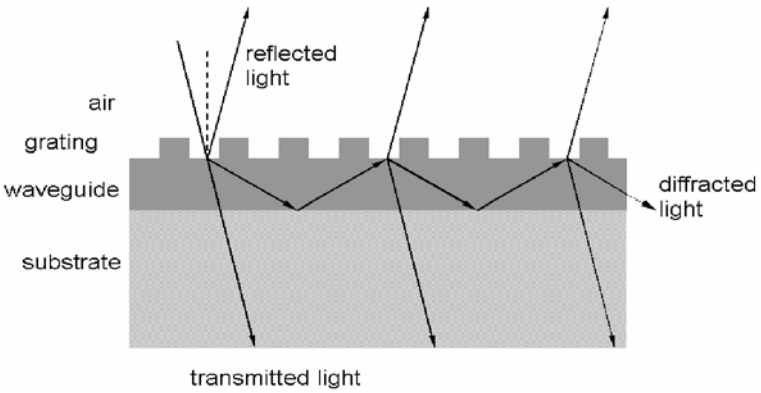

Fig. 9.3. Principal structure of a GWS. In the high-refractive index film, both zero- and first-order diffracted waves may propagate. The first-order diffracted wave suffers total internal reflection at the film boundaries. The performance of the system is determined by interference of multiply reflected as well as diffracted and re-diffracted waves

occur in both reflection and transmission. The propagation angle ψ_M of the M-th order diffracted wave may be calculated according to the following equation (9.2) (please consult textbooks on optics for the derivation):

$$\begin{aligned} &\text{reflection:} \qquad \sin\psi_M = \sin\varphi + \frac{M\lambda}{\Lambda}; \quad M = 0, \pm 1, \pm 2, \ldots \\ &\text{transmission:} \qquad \sin\psi_M = \frac{\sin\varphi}{n} + \frac{M\lambda}{n\Lambda}; \quad M = 0, \pm 1, \pm 2, \ldots \end{aligned} \tag{9.2}$$

M is the order of diffraction. For reflected modes, the wave propagates in air, so that $n = 1$. For transmitted modes, n equals the refractive index of the film material. For the zero's diffraction order, (9.2) becomes identical to Snell's law of refraction. Λ is the period of the grating.

9.2.2 Propagating Modes and Grating Period

As in the case of total internal reflection, the modes can only propagate into the film (or back into the ambient) when $\sin\psi_M < 1$ is fulfilled. Otherwise the wave is evanescent. Particularly, it is possible to chose the geometrical parameters such that no diffracted waves occur in reflection, while in transmission, at least the first order diffracted wave is allowed to propagate. Let us for simplicity regard only the case $M = +1$. We have:

$$\begin{aligned} &\text{reflection:} \qquad \sin\psi_1 > 0 \Rightarrow \sin\varphi + \frac{\lambda}{\Lambda} > 1 \Rightarrow \Lambda < \frac{\lambda}{1 - \sin\varphi} \\ &\text{transmission:} \qquad \sin\psi_1 < 1 \Rightarrow \frac{\sin\varphi}{n} + \frac{\lambda}{n\Lambda} < 1 \Rightarrow \Lambda > \frac{\lambda}{n - \sin\varphi} \end{aligned}$$

These conditions are fulfilled for:

$$\frac{\lambda}{n - \sin\varphi} < \Lambda < \frac{\lambda}{1 - \sin\varphi} \tag{9.3}$$

Condition (9.3) defines a range of grating periods suitable for our idea, as long as $n > 1$ is fulfilled.

A grating period chosen according to (9.3) guarantees, that an incident wave 'creates' at least three propagating modes: the specularly reflected wave ($M = 0$ in reflection), the usual transmittance ($M = 0$ in transmission), and a diffracted wave ($M = 1$ in transmission) that propagates into the film.

Once the diffracted wave has been generated by the grating, it may be re-diffracted into the zero's order at next bouncing onto the grating and thus contribute to the usual transmittance and reflectance. When the phase relations are suitable, the specularly reflected wave may be enhanced, while the transmission becomes suppressed – similar to what we have seen in the multilayer stack. In order to get 100% reflection, of course, no light should be allowed to leave the system into the substrate. This is guaranteed when the

diffracted wave suffers total internal reflection at the film-substrate-interface. In terms of (6.24) this leads us to the condition:

$$\sin\psi_1 = \frac{\sin\varphi}{n} + \frac{\lambda}{n\Lambda} > \frac{n_s}{n}$$

or

$$\Lambda < \frac{\lambda}{n_s - \sin\varphi} \tag{9.4}$$

For $n_s > 1$, (9.3) and (9.4) finally yield:

$$\frac{\lambda}{n - \sin\varphi} < \Lambda < \frac{\lambda}{n_s - \sin\varphi} \tag{9.5}$$

For the minus first order ($M = -1$), we find in analogy:

$$\frac{\lambda}{n + \sin\varphi} < \Lambda < \frac{\lambda}{n_s + \sin\varphi} \tag{9.6}$$

It turns out, that the refractive index of the film must be higher than that of the substrate.

Because in oblique incidence the plus first and minus first orders are physically distinct, in the general case we have two types of first order diffracted waves.

9.2.3 Energy Exchange Between the Propagating Modes

Having clarified the relation between grating period and refractive indices, let us now try to understand the energy exchange between the incident, transmitted, reflected, and first-order diffracted waves in a qualitative manner. Imagine a wave front impinging onto the grating. A part of the intensity will be specularly reflected, while the other part is either ordinarily transmitted or diffracted. The diffracted wave suffers total reflection at the film-substrate boundary, and bounces onto the grating for a second time, but now from the film side. Again, it may be reflected (remaining in the same diffracted mode) or suffer a second diffraction process, which spreads the intensity into one of the other allowed modes. Note that a wave primarily diffracted into the plus first order wave may be diffracted into the minus first order propagating mode at second bouncing.

Let us now assume, that we have chosen a particular film thickness so that the diffracted waves interfere constructively when having performed one loop in the film. The phase gain while propagating through the film has been calculated previously (equation (7.15)). We have:

$$2\delta = \frac{4\pi}{\lambda} nd\cos\psi$$

This expression is obtained from (7.15) when re-substituting the incidence angle φ by the refraction angle ψ, which corresponds to the propagation angle relevant for the system discussed here. Consequently, constructive overlapping of the diffracted wave trains occurs, when the condition:

$$\frac{4\pi}{\lambda} nd\cos\psi + 2\delta_{21} + 2\delta_{23} = 2j\pi; \quad j = 0, 1, 2, \ldots \tag{9.7}$$

is fulfilled. $2\delta_{21}$ is the phase shift of the wave when being reflected at the grating side of the film, and $2\delta_{23}$ that at the film-substrate side. The factor 2 has been introduced for mathematical convenience. As before, j is the interference order.

In constructive interference conditions, the wave is expected to grow in intensity. It cannot loose energy at the film-substrate interface. The only energy-loss channel accessible for the diffracted wave is the re-diffraction at the grating. This loss grows proportionally to the intensity of the diffracted wave. In stationary conditions, the loss at the grating must compensate the energy input from the incident irradiation. The only question is: Which of the energy loss mechanisms of the diffracted wave is the dominant one, zero order transmission or reflection?

In fact, in the discussed case of coherent superposition of the diffracted wave, the incident intensity in the stationary case will only contribute to the systems reflectance. As long as there is energy re-diffracted into the zero-order transmitted wave, part of this intensity comes back to the grating as the result of reflection at the film-substrate interface, and again contributes to the diffracted wave. Hence, in this case we have no stationary regime, because the diffracted wave still grows in intensity as the result of the mentioned feedback mechanism. On the other hand, any intensity going into the zero-order reflected wave leaves the system forever. Therefore, the intensity of the diffracted (guided) wave will increase until the intensity of the zero-order reflected light completely compensates the energy input caused by the incident irradiation. In other words, we get 100% reflection. Then, the transmittance must be zero, as the result of destructive interference between multiple internal reflections of zero-and first order diffracted waves and their mutual energy exchange.

9.2.4 Analytical Film Thickness Estimation for a GWS

Of course, the mentioned constructive overlapping of the diffracted wave trains is only possible when the wavelength is consistent with (9.7). Let us further refer to this wavelength as the resonance wavelength λ_0. According to (9.7), there are several film thickness values $\{d_j\}$ corresponding to a desired resonance wavelength. They may be explicitly calculated according to:

$$d = d_j = \frac{\lambda_0 \Lambda}{2\pi} \frac{j\pi + \delta_{21} + \delta_{23}}{\sqrt{n^2\Lambda^2 - (\Lambda\sin\varphi \pm \lambda_0)^2}} \tag{9.8}$$

which is obtained from (9.7) when $\cos\psi$ is substituted by:

$$\cos\psi = \sqrt{1 - \frac{\left(\Lambda \sin\varphi \pm \lambda_0\right)^2}{n^2\Lambda^2}}$$

as it follows from (9.2).

The phase shift δ_{23} is easily calculated from Fresnel's coefficients assuming internal total reflection conditions (see problem 1 to Chaps. 6–9). On the other hand, δ_{21} cannot be calculated this way due to the finite profile depth of the grating itself. However, when the grating depth is much smaller than the wavelength, it may be neglected, and then the phase shift may again be approached by the expression following from Fresnel's coefficients at the film-air interface. This leads to the following equations:

$$\begin{aligned}
&\tan\delta_{21,s} = \frac{1}{n^2}\tan\delta_{21,p} = \sqrt{\frac{\left(\Lambda\sin\varphi \pm \lambda_0\right)^2 - \Lambda^2}{n^2\Lambda^2 - \left(\Lambda\sin\varphi \pm \lambda_0\right)^2}}\,; \\
&\tan\delta_{23,s} = \frac{n_s^2}{n^2}\tan\delta_{23,p} = \sqrt{\frac{\left(\Lambda\sin\varphi \pm \lambda_0\right)^2 - n_s^2\Lambda^2}{n^2\Lambda^2 - \left(\Lambda\sin\varphi \pm \lambda_0\right)^2}} \qquad (9.9) \\
&\text{`}+\text{'}:\ \Lambda \in \left[\frac{\lambda_0}{n - \sin\varphi}, \frac{\lambda_0}{n_s - \sin\varphi}\right];\ \text{`}-\text{'}:\Lambda \in \left[\frac{\lambda_0}{n + \sin\varphi}, \frac{\lambda_0}{n_s + \sin\varphi}\right]
\end{aligned}$$

The subscripts s and p denote s- or p-polarization of the incident light, n_s is the substrate refractive index. The different signs in (9.8)–(9.9) correspond to the first (+) and minus first (−) diffraction orders, which are not equivalent for oblique light incidence.

When applying (9.8)–(9.9), one should keep in mind that these equations have been obtained neglecting the finite profile depth of the grating. That leads to a systematic error in the film thickness estimation, which is of the order of the grating depth itself. On the other hand, in (9.8)–(9.9) any absorption losses are neglected. In practice, absorption will cause a certain absorption loss A in the waveguide layer. Additionally, the presence of absorption destroys total internal reflection at the interfaces of the waveguide, which leads to a residual transmission T even in resonance conditions. Both circumstances reduce the peak reflectance $R_{\max}$ achieved by the system.

GWS are candidate systems for extreme narrow line reflection filters, because in resonance conditions the reflectance may theoretically reach 100%, while the system is only merely reflective off-resonance. Consequently, the reflection spectrum is expected to show narrow peaks of nearly ideal reflection, which suggest applications as narrowband reflection filters.

The exact theoretical treatment of a GWS is more complicated, because the real profile shape and depth of the grating have to be taken into account. In this case, our theoretical apparatus as derived so far is clearly at stake. In practice, calculations of this kind are performed today utilizing commer-

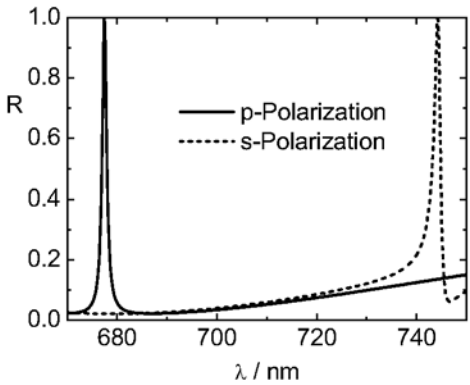

Fig. 9.4. Model calculation of the normal incidence reflectance of a GWS. The film refractive index is 2.3, the substrate index 1.37. The assumed grating period is 475 nm, the grating depth 40 nm, and the film thickness 285.2 nm

cial grating solver software, which accomplishes these calculations, for example, within the Rigorous Coupled Wave Approach (RCWA). Here, Maxwell's equations are rigorously solved expanding the electric and magnetic fields into a series of Bragg modes.

Figure 9.4 shows the thus calculated normal incidence reflection spectrum of a GWS with parameters as defined in the figure caption. As the assumed one-dimensional grating structure is clearly laterally anisotropic, even at normal incidence, the reflection behaviour is dependent on the polarization. At oblique incidence, each of the reflection lines splits into two maxima due to the different behaviour of the plus and minus first diffraction orders.

9.2.5 Remarks on GWS Absorbers

With slight modifications in the design of Fig. 9.3, GWS may be used as spectrally selective absorbers. Figure 9.5 (left) depicts one of the possible realizations of a GWS based absorber. It consists of the waveguide film and a metal layer, separated from each other by the diffraction grating. Due to the (sufficiently thick) metal layer, the transmittance is automatically zero, so that the light is either reflected by the system or absorbed in the metal fraction. Again, the diffracted wave is in total internal reflection conditions at the waveguide-air interface, so that diffracted waves suffer multiple bouncing onto the metal surface, which leads to an enhancement of light absorption. In resonance conditions, one may achieve destructive interference in reflectance, so that the light is effectively absorbed, especially in the case of *p*-polarization, because in this case the guided mode may couple to a surface plasmon polariton at the waveguide-metal interface.

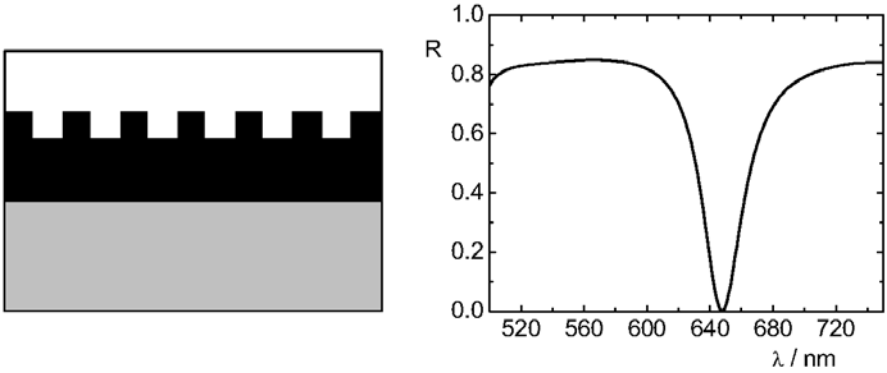

Fig. 9.5. Left: GWS-absorber, *grey*: substrate, *black*: metal, *white*: waveguide material; **right**: calculated reflectance for p-polarization (RCWA); $d = 250\,\text{nm}$; grating profile depth 50 nm; $\Lambda = 475\,\text{nm}$; $n = 1.37$. The calculation has been carried out supposing aluminum as the metal

Figure 9.5 (right) shows the calculated reflectance of a model system with vanishing reflectance at nearly 650 nm. The theoretical absorptance is consequently close to 100%.

Although this calculation has again been performed in terms of the RCWA, an analytical estimation of the waveguide thickness by (9.8) is possible as well. However, the phase shift $2\delta_{23}$ now has to be calculated from the complex Fresnel's coefficient at the waveguide-metal interface, considering both the real and imaginary parts of the metals index of refraction.

9.3 Resume from Chapters 6–9

9.3.1 Overview on Main Results

Today's optical instrumentation becomes more and more complex. In order to guarantee durability and high optical performance of any optical component, its surfaces have to be over-coated with specially designed thin film stacks to achieve tailored optical properties as well as surface protection. Most of the coatings produced today are built from optically (nearly) isotropic and homogeneous materials. Therefore, it is utmost important to understand the theory of the optical properties of homogeneous and isotropic thin solid films and multilayer stacks built from them. The Chaps. 6–9, which form the second part of this book, were mainly dedicated to this kind of system.

In particular, we obtained the following results:

– The reflection at ideally smooth surfaces or interfaces as well as the transmission of light through the interfaces may be described in terms of Fresnel's equations. Metallic reflection and total internal reflection of light have been discussed as special cases of Fresnel's equations.

- We established the dispersion relation for propagating surface plasmons.
- Explicit expressions have been derived to calculate both transmittance and reflectance of a thick slab. The derived theory allows to consider the effects of absorption as well as oblique incidence.
- The corresponding equations for a single film have also been derived. They include free standing films, films on a semiinfinite substrate, as well as films on a possibly absorbing substrate of finite thickness. The important special cases of quarterwave and halfwave layers have been addressed as well.
- Finally, we derived the matrix formalism to calculate transmittance and reflectance of multilayer stacks. Some simple examples of high reflectors, narrowband filters as well as an edge filter have been demonstrated.

The previous results have been derived to deal with the optical behaviour of any thin film stack built from optically homogeneous and isotropic materials. This will be sufficient to handle a large amount of practically important thin film systems. On the other hand, there is extensive theoretical and experimental research today pursuing on the implementation of optically inhomogeneous and/or anisotropic coating materials to manufacture coatings with principally novel optical properties. To comply with these trends, we also regarded special systems that go beyond the homogeneous and isotropic films mentioned so far:

We demonstrated the effect of uniaxial optical anisotropy on Fresnel's reflection coefficients. On this basis, important effects of the new field of Giant Birefringent Optics (GBO) could be explained.

A mathematical apparatus has been derived to calculate the optical properties of inhomogeneous films with a refractive index that changes along the film axis. As examples, we regarded the special cases of linear gradient layers and rugates. As in the quarterwave stack, the periodic modulation of the refractive index leads to the appearance of transmission stopbands (and correspondingly to regions of high reflection).

As the last example, we provided a qualitative discussion of the behaviour of resonant Grating Waveguide Structures (GWS). In the context of the previous statements, these systems combine optical inhomogeneity with anisotropy. Indeed, regarding the diffraction grating as a thin laterally textured film, the latter appears to be laterally inhomogeneous with a periodic modulation of the refractive index. On the other hand it is clearly anisotropic, therefore exhibiting polarizing properties even at normal light incidence. Again, these systems show reflection maxima, which are equivalent to transmission stopbands. As in the case of quarterwave stacks or rugates, the existence of a stopband is caused by the periodicity of the system, here resulting from the assumed grating structure.

Table 9.1 gives an overview of the above mentioned coating types and their mutual relation with respect to homogeneity and isotropy. The typical dielectric coatings, as composed from homogeneous and isotropic layer

Table 9.1. Schematic overview on the discussed research fields in thin film optics

<table>
<tr><td rowspan="2">Optically homogeneous</td><td rowspan="2">Optically isotropic</td><td colspan="2">Pure materials or nanoscopically homogeneous mixtures</td><td colspan="2">Composites or porous layers</td></tr>
<tr><td>Non-absorbing</td><td>Absorbing</td><td>Non-absorbing</td><td>absorbing</td></tr>
<tr><td rowspan="2">yes</td><td>yes</td><td>Conventional dielectric Coatings (Sect. 8.2, 9.1)</td><td>Conventional (selective) absorbers, metal films</td><td>composite dielectric coatings</td><td>Cermets, metal island films</td></tr>
<tr><td>no</td><td>Giant birefringent optics (Sect. 6.5)</td><td>Polarizer foils</td><td colspan="2">Metal island films (Sect. 4.5)</td></tr>
<tr><td rowspan="2">no</td><td>yes</td><td colspan="2" rowspan="2">Rugates, Gradient index layers (Sect. 8.1)</td><td colspan="2">Rugates, Gradient index layers</td></tr>
<tr><td>no</td><td colspan="2">Photonic crystals and Plasmonics</td></tr>
<tr><td></td><td></td><td>Grating waveguide structures (reflectors) (Sect. 9.2.1–9.2.4)</td><td>Grating waveguide structures (absorbers) (Sect. 9.2.5)</td><td colspan="2"></td></tr>
</table>

materials, are represented in the left upper corner in the table. Anisotropy or inhomogeneity of the coating materials (moving *downwards* in the table) leads to such important classes of novel coatings like Giant Birefringent Optics-devices and rugate filters. Finally, the Resonant Grating Waveguide Structures (GWS) combine lateral inhomogeneity with anisotropy.

On the contrary, starting from conventional coatings and moving *to the right* in Table 9.1 will lead us to nanoscopically heterogeneous coating materials, which may however be optically homogeneous due to the small characteristic size of the structural units. In this way it is possible to manipulate optical material properties, offering more flexibility in the choice of optical constants for design tasks. In Chap. 4, metal island films have been discussed as a prominent example.

The most complicated case, namely the presence of anisotropy, absorption, and heterogeneity on different length scales in combination will finally lead to the fields of photonics and plasmonics, but is not the subject of this book.

9.3.2 Examples

In order to illustrate the theoretical results obtained in the second part of this book, let us look at a few experimental examples. Figure 9.6 shows the

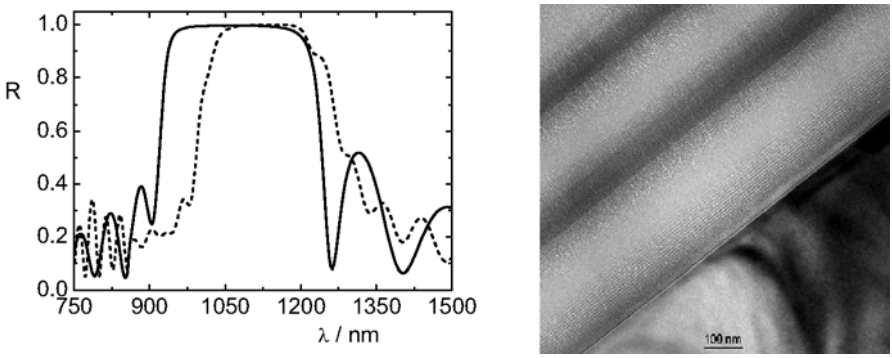

Fig. 9.6. Left: measured reflectance of a quarterwave stack (*solid*) and a rugate filter (*dash*); **right**: TEM cross sectional image of the first two periods of the rugate filter (Courtesy of Ute Kaiser, FSU Jena, Germany)

measured reflectance of a quarterwave stack that is built from alternating niobium pentoxide and silicon dioxide layers (solid line). In sum 17 layers have been deposited with a thickness corresponding to a central rejection wavelength of 1064 nm. In addition, the figure shows the reflectance of a rugate filter with 20 periods, consisting of the same materials (dash). The mismatch in the central wavelength results from practical difficulties in monitoring the rugate filter period during deposition.

Figure 9.6 (right) shows a cross-sectional image of the first two periods of the mentioned rugate filter, as obtained by transmission electron microscopy. The concentration (and refractive index) profile obviously changes in a continuous manner with increasing distance from the substrate. A more detailed analysis shows that the profile is indeed close to a sinusoidal one.

The second example concerns a dielectric grating waveguide structure.

To produce a grating wave guide structure of the type as discussed in Sect. 9.2.1, one would principally have to deposit a waveguide layer with a high refractive index onto a low refractive index transparent substrate. After that, the grating on top of the film might be produced by a suitable lithographic technique.

In fact, one may choose a somewhat different way. The grating might be etched on the top of the substrate, and after that, the waveguide layer may be deposited. In this case, instead of the geometry from Fig. 9.3, we have to expect a sample geometry like shown in Fig. 9.7.

Let us have a look at a grating waveguide structure that has been produced this way. In experimental practice, sample preparation started with lithographic writing of the grating on the bare substrate surface. As substrate, a fused silica wafer has been used. Figure 9.8 shows the SEM image of a substrate surface with a rectangular grating (groove depth $t = 57$ nm, period $\Lambda = 330$ nm).

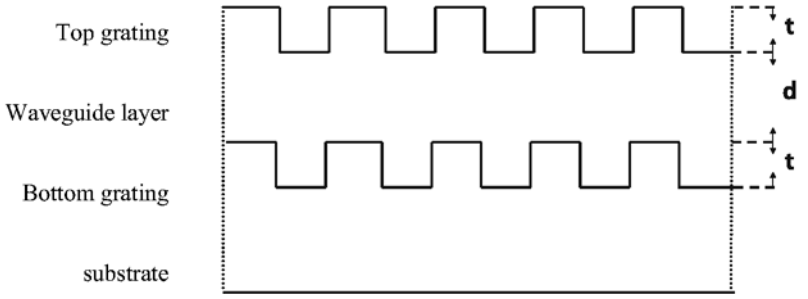

Fig. 9.7. Alternative sample geometry of a grating waveguide structure

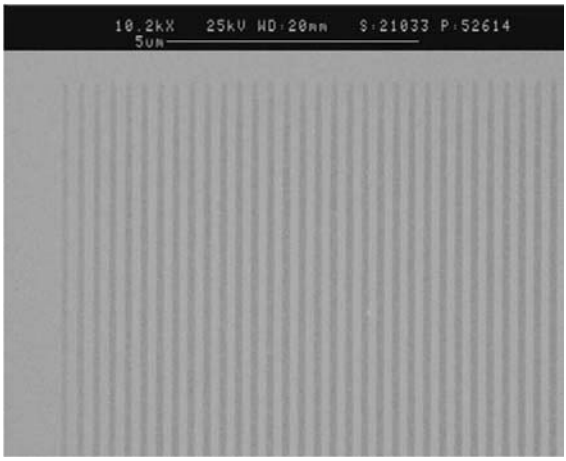

Fig. 9.8. SEM image of the grating on fused silica (Courtesy of FSU (IAP) in Jena, Germany)

Before depositing the waveguide layer, the normal incidence transmittance of the grating-on-substrate system has been measured by means of a Zeiss microscope spectrophotometer for both polarizations. One had to use a microscope photometer because of the small sample area. The corresponding transmittance spectra are depicted in Fig. 9.9 and show two characteristic so-called Wood's anomalies at 330 nm and 480 nm, corresponding to $\lambda = \Lambda$ and $\lambda = n_s \times \Lambda$. TE denotes s-polarization, in grating theory that means that the electric field vector is parallel to the grooves of the grating. Correspondingly, in the TM wave (p-polarization), the electric field vector is perpendicular to the grooves.

After that, the waveguide layer (titanium dioxide) has been deposited by electron beam evaporation in a Balzers BAK 640 deposition system. Consequently, we have to expect that the final sample structure rather resembles the geometry shown in Fig. 9.7 than that in Fig. 9.3.

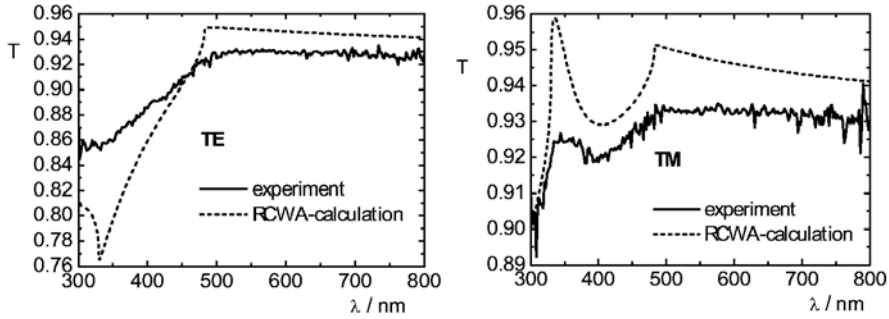

Fig. 9.9. Transmission spectra of the textured silica substrate from Fig. 9.8

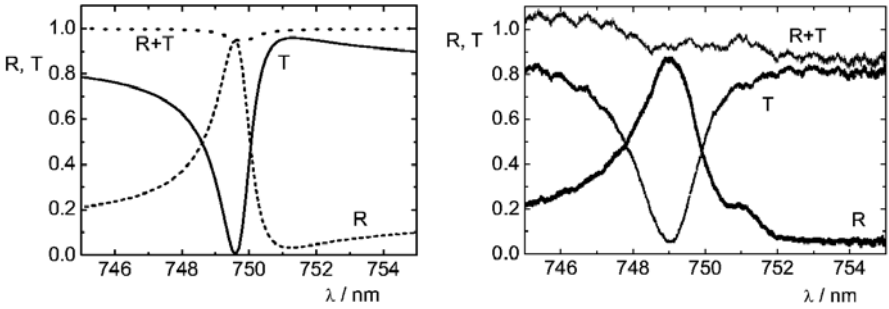

Fig. 9.10. Left: Calculated transmittance and reflectance for a GWS (TE wave); **right**: experimental spectra

Because transmission and reflection spectra of GWS structures are extremely sensitive to the angle of incidence, the high numeric aperture of the microscope photometer made it impossible to record the spectra by means of this kind of equipment. Instead, one had to use a laser source for sample illumination that guarantees nearly parallel incident light at the sample surface, which had an area of only 1 mm × 1 mm. For the present study, a transmission and reflection measurement set-up at the Laser Laboratorium Göttingen, Germany has been utilized. The incident light was provided by femtosecond laser pulses of a titanium-sapphire laser. By means of this laser system, transmission and reflection spectra at nearly normal light incidence could be recorded around $\lambda = 750\,\text{nm}$ in a spectral range limited by the spectral bandwidth of the laser pulse. Practically, the incidence angle was set 10 deg.

Figure 9.10 (left) shows the TE wave transmittance and reflectance calculated in terms of the RCWA for a GWS, built from a TiO_2 film on SiO_2. In the given spectral region, the TiO_2 refractive index was set 2.216, while the extinction coefficient is around 6×10^{-5}. The calculation is carried out for an assumed film thickness of 400 nm and a grating thickness of 50 nm, the grating period being 328 nm. On both sides of the film, we assumed a

rectangular grating profile such as shown in Fig. 9.7 with a filling factor of 0.5. The assumed angle of incidence is 9.8 deg.

On the right hand sight of Fig. 9.10, one sees corresponding experimental spectra, recorded at an angle of incidence of 10 deg. The evaporated TiO_2-mass coverage corresponds to a thickness of 451 nm, so that, according to Fig. 9.7, $d \approx 401$ nm.

As seen from the figure, the experimental maximum reflectance in resonance reaches approximately 87% , while the rejection band FWHM is around 2 nm. The reflectance is thus somewhat lower than the theoretical one, which is around 95% in resonance. As the sum of transmittance T and reflectance R tends to exceed the 100%-value in several parts of the spectrum in a significant manner, it is clear that the inaccuracy of the T- and R-measurements must be of the order of a few percent.

9.3.3 Problems

1. Derive explicit expressions for the phase shift occuring at an interface at total internal reflection (real refractive indices only).

   Answer:

   s-polarization: $\arg r_s = 2 \arctan \left(- \dfrac{\sqrt{\sin^2 \varphi - \frac{n_2^2}{n_1^2}}}{\cos \varphi} \right)$

   p-polarization: $\arg r_p = 2 \arctan \left(- \dfrac{n_1 \sqrt{\frac{n_1^2}{n_2^2} \sin^2 \varphi - 1}}{n_2 \cos \varphi} \right)$

   Remark: Use expressions (6.14) and (6.16) and consider that $\cos \psi$ is purely imaginary. The solution is then immediately obtained calculating the phase of the Fresnel's coefficients. In order to obtain expressions (9.9), the angle ψ has to be expressed as a function of grating period and wavelength according to (9.2).

2. Repeat the same for an air-metal interface at oblique incidence, assuming $\omega \to 0$!

   Answer:
   p-polarisation: no phase shift
   s-polarization: the phase shift is π

   Remark: For a vanishing frequency, the refractive index of a metal becomes infinitively large by modulus. Therefore, from Fresnel's formulae we immediately obtain $r_p \to 1$ and $r_s \to -1$. At any angle of incidence, in s-polarization the electric field vectors at the metal surface are therefore equal by modulus, but antiparallel, so that the resulting field strength is

zero. On the other hand, according to Fig. 6.3, the mutual orientation of the field strength vectors for p-polarization depends on the angle of incidence. For normal incidence, we have the same situation as in the case of s-polarization, hence the vectors are equal by modulus and antiparallel. On the other hand, at grazing incidence, the vectors are nearly parallel, so that the field strength vectors sum up to a higher resulting field strength. A weakly absorbing adsorbate layer at the metal surface may therefore be detected in the reflectance spectrum of the p-polarized light at grazing incidence, while the reflectance of s-polarized light is insensitive to the adsorbate. This effect is frequently used in the infrared spectral region for the detection of adsorbates at metal interfaces. The corresponding spectroscopy method is commonly called Infrared Reflection Absorption Spectroscopy IRAS. Due to the frequency limitations (IR), in IRAS one detects vibrational degrees of freedom of the adsorbates. As in p-polarization the resulting $\boldsymbol{E}$-vector is normal to the metal surface, it may only excite molecular vibrations normal to the surface. Hence, the resulting spectra may be used to identify the adsorbate molecules, and to determine their orientation with respect to the surface.

3. Calculate Brewster's angle for the surfaces air-glass and glass-air. Compare the results. Assume $n_{glass} = 1.45$!

   Answer:
   Air-glass: $\varphi_B = 55.4°$
   Glass-air: $\varphi_B = 34.6°$

   Both angles are connected to each other by Snell's law of refraction.

4. At Brewsters angle, calculate the polarization degree for the transmitted light at the surfaces from problem 3. The incident light is supposed to be unpolarized. Then, calculate the polarization degree of light transmitted through a glass plate at Brewster's incidence angle!

   Answer: The polarization degree in transmission is defined as

   $$\left|\frac{T_s - T_p}{T_s + T_p}\right|.$$

   In the absence of absorption, it may be written as

   $$\left|\frac{R_p - R_s}{2 - R_s - R_p}\right|.$$

   At Brewsters angle, $R_p = 0$. A direct application of (6.16) and (6.18) yields a polarization degree of 0.067 at the air-glass interface and the same value at the glass-air interface. The polarization degree of a thick glass plate may be calculated utilizing (7.1). We obtain 0.126 or 12.6%.

5. How many glass plates should be arranged in sequence to achieve a polarization degree of 99.9% in transmission, when they are illuminated

with unpolarized light at Brewster's angle? Neglect the effect of multiple reflections of the s-polarized light.

Answer: 30 plates

6. Calculate the characteristic matrices for a quarterwave- and a halfwave layer!

   Answer:
   quarterwave layer: $M = \begin{pmatrix} 0 & -\frac{\mathrm{i}}{n} \\ -\mathrm{i}n & 0 \end{pmatrix}$

   halfwave layer: $M = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$

   Remark: In correspondence to the already discussed properties of a halfwave layer, its characteristic matrix does not contain any information about the refractive index of the film.

7. Imagine a normal incidence thin film spectrum like shown in Fig. 7.5. Which qualitative changes in the spectrum are expected at oblique incidence?

   Answer: The interference pattern shifts to shorter wavelength or higher wavenumbers. For checking, calculate

   $$\left.\frac{\partial\lambda}{\partial\varphi}\right|_{\delta=\text{const.}}$$

   from (7.15). You should find:

   $$\left.\frac{\partial\lambda}{\partial\varphi}\right|_{\delta=const.} = -\lambda\frac{\sin\varphi\cos\varphi}{n^2-\sin^2\varphi} < 0.$$

   This wavelength shift does not depend on the film thickness and may in principle be utilized to estimate the refractive index of a single thin film.

   Remark: As a consequence of the shift to shorter wavelength, thin film interference colours depend on the angle of incidence. An interference filter changes its colour in direction to the blue-violet when being inclined.

8. In (8.10), (8.12), (8.14)–(8.16): Make sure that the dimensions in these equations are correct. Keep in mind, that the meaning of U and V is different for different polarizations.

9. In (9.8), an infinitively large number of thickness values is expected to cause the same resonance frequency of a GWS. Hence, as long as the effects of the finite grating profile thickness are neglected, the resonance wavelength does not depend on the interference order j. Consider the effect of the interference order on the FWHM of the reflectance peak of a GWS!

   Answer: $FWHM \propto \frac{1}{d_j}$

Remark: In full analogy to what we have learned in Chap. 4, the FWHM is inversely proportional to the 'lifetime' of a photon in the waveguide. As long as there is no absorption in the waveguide, the only chance for a photon to escape from the waveguide is to suffer diffraction when bouncing onto the grating. The FWHM is therefore proportional to the bouncing rate of a guided photon onto the grating, while the latter is inversely proportional to the thickness of the film.

Part III

Semiclassical Description of the Interaction of Light with Matter

10 Einstein Coefficients

10.1 General Remarks

Starting from Chap. 10, we will turn to a more precise description of optical thin film spectra. It will now be our purpose to develop a semiclassical theory of the interaction between light and matter. In this picture, the matter will be described in terms of quantum mechanical models, while the electromagnetic field is described, as before, in terms of Maxwell's equations. For our particular subject of thin film spectroscopy, such a treatment has a few important consequences, namely:

- The dispersion models developed in Chaps. 2–4 need to be modified. Instead of Newton's equations of motion (classical treatment of matter), we will now have to solve Schrödinger's equation to calculate microscopic dipole moments.
- Having calculated microscopic dipole moments in quantum mechanical terms, the dielectric function may be calculated in terms of (3.20)–(3.22d). This will lead us to quantum mechanical expressions for the optical constants, which may then be used to solve Maxwell's equations.
- The theoretical apparatus developed in Chaps. 6–9 remains valid, because it is based on Maxwell's theory only.

In order to develop the mentioned theoretical description, some basic knowledge on quantum mechanics will be absolutely necessary to the reader. This concerns Schrödinger's equation, general properties of the wave function, as well as simple quantum mechanical models such as the harmonic oscillator, and perturbation theory. It will be our purpose to apply these theoretical tools to the treatment of the interaction of electromagnetic irradiation with matter.

10.2 Phenomenological Description

First of all, we must formulate a suitable physical model. The very simplest possibility to deal with the radiation-matter subject in quantum mechanics is to regard a so-called two-level system. The idea is to neglect the multiplicity of energy levels a real material system might have, and to restrict the attention

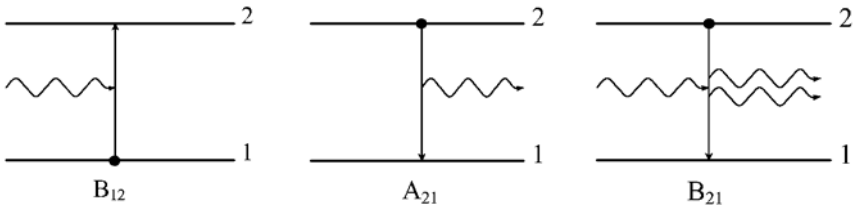

Fig. 10.1. Introduction of Einstein coefficients

to two energy levels only. This makes sense, when the electromagnetic wave has a frequency close to the eigenfrequency of the subsystem described by the two energy levels.

A very popular and transparent treatment of the interaction of such a two-level system with electromagnetic irradiation may be performed in terms of the so-called Einstein coefficients. It will be the purpose of this chapter to deal with this treatment.

Let us have a look at Fig. 10.1. It shows two discrete energy levels, E_1 and E_2. The level 2 corresponds to a higher energy state of the system than level 1. Simply for unambiguity, let us call the first level a ground state, and the second one an excited state. In order to describe the interaction between radiation and the two-level system in terms of Einstein coefficients, we have to consider three phenomena: absorption, spontaneous emission, and stimulated emission of light by the two-level system.

Let us assume, that the system is in the first, low energy state. When the radiation source is switched on, and the radiation frequency is close to the eigenfrequency of the system, an absorption process of light is expected to transfer the system from energy level 1 to the level 2 (the quantum system becomes excited). Due to energy conservation, this energy gain of the two-level system must be accompanied by an energy loss of the radiation field, hence the energy is transferred from the electromagnetic field to the two-level system. This absorption process becomes more probable when the electromagnetic irradiation is more intense. The transition rate from level 1 to level 2 by absorption of light is therefore expected to be proportional to:

- The intensity of the irradiation
- The statistical probability to find the system in state 1

Let us now assume, that the electromagnetic field interacts with a large amount of such two-level systems. Then, a considerably large part of the energy of the field may be transferred to the assembly of two-level systems. However, the energy loss of the field is always equal to an integer multiple of the excitation energy of the two-level systems. In our simple description, these single portions of light energy that may be absorbed will be called photons.

We will now consider the case, that the system is in the second (excited) state. From our experience we know, that within a certain time any excited

system tends to loose its energy, thus returning back into the ground state. In order to loose energy, our system has to perform a process that is reverse to the absorption of light, namely the emission of light. Let us postulate, that an excited quantum system may loose energy without any stimulation from outside by the emission of an energy portion of light that exactly corresponds to the energy difference between the two energy levels. In this case, we speak on the spontaneous emission of a photon. The transition rate from level 2 to level 1 by spontaneous emission will be proportional to:

– The statistical probability to find the system in state 2

There is a second mechanism to turn the system from the excited state into the ground state. We postulate that the system may also perform a so-called stimulated emission process. This has to be understood as an emission of light activated by the impinging electromagnetic wave. The transition rate from state 2 to state 1 caused by this process should be proportional to:

– The intensity of the irradiation
– The statistical probability to find the system in state 2

Of course, any of these elementary processes enters into the resulting full transition rate with a specific proportionality coefficient. If any of the processes considered so far turns out to be not necessary, in the further derivations the corresponding proportionality coefficient would become zero.

As the reader will already have guessed, the mentioned proportionality coefficients are nothing else than the so-called Einstein-coefficients. It is a common practice to use the following symbols for Einstein coefficients:

– A_{21} for spontaneous emission $(2 \to 1)$
– B_{21} for stimulated emission $(2 \to 1)$
– B_{12} for absorption $(1 \to 2)$

The next section will deal with a mathematical treatment of Einstein coefficients, and it will be our purpose to derive – step by step – exact expression for Einstein coefficients in the dipole approximation.

Finally, Fig. 10.1 gives a schematic representation of all the mentioned elementary processes. Here, the vertical arrows correspond to the transitions between ground and excited states, while the sinusoidal structures demonstrate annihilation or creation of a photon.

10.3 Mathematical Treatment

Commonly, in the philosophy of Einstein coefficients, the electromagnetic field is characterized by the so-called spectral density of the radiation field defined as:

$$u \equiv \frac{\mathrm{d}E}{V \mathrm{d}\omega} \tag{10.1}$$

This equation defines the spectral density as the field energy per angular frequency interval and per volume. In quantum mechanics, it is generally accepted to use the symbol E for the energy. This may lead to confusion with the electric field strength, and we will try to avoid any misinterpretations using suitable subscripts when necessary.

Let us further assume, that we have an assembly of N_0 two-level systems, interacting with the radiation field. Let N_1 be the number of systems in the ground state, and N_2 in the excited one. Obviously,

$$N_1 + N_2 = N_0 = \text{const.} \tag{10.2}$$

Due to the radiation field, the population of the excited state may be changed. In terms of the mechanisms proposed in Sect. 10.2, the corresponding rate equation is:

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = N_1 B_{12} u - N_2 B_{21} u - N_2 A_{21} \tag{10.3}$$

Corresponding to Fig. 10.1, the first term describes absorption, which leads to an increase in the population of the excited state. The second term corresponds to stimulated emission, and the third one to spontaneous emission, both resulting in a decrease of the population of the excited state. Of course, here and throughout this section we can only regard the spectral density at the frequency corresponding to the eigenfrequency of the two-level system. Up to now, we have no information how this frequency is connected to the excitation energy of the system.

Of course, this treatment is not very helpful as long as we are unable to give explicit expressions for Einstein coefficients. Let us therefore turn to their determination.

First of all, some interesting information may be obtained regarding the special case of thermodynamic equilibrium between radiation and matter. In this situation, $\mathrm{d}N_2/\mathrm{d}t = 0$, and from (10.3) it follows:

$$\text{equilibrium:} \quad \frac{N_1}{N_2} = \frac{B_{21} u + A_{21}}{B_{12} u} \tag{10.4}$$

On the other hand, in equilibrium conditions Boltzmann's statistics hold, resulting in

$$\text{equilibrium:} \quad \frac{N_1}{N_2} = \mathrm{e}^{\frac{E_2 - E_1}{k_\mathrm{B} T}} \tag{10.5}$$

where T is now the absolute temperature. From (10.4) and (10.5) in combination, we obtain an expression for the spectral density of the radiation field in equilibrium conditions as:

$$\text{equilibrium:} \quad u = \frac{A_{21}}{B_{12}\left(\mathrm{e}^{\frac{E_2 - E_1}{k_\mathrm{B} T}} - \dfrac{B_{21}}{B_{12}}\right)} \tag{10.6}$$

It is useful to discuss some special cases resulting from (10.6). Let us consider the case of $T \to 0$. Clearly, in this case, the radiation field at the eigenfrequency of the two-level system is vanishing in intensity. In the other extreme case ($T \to \infty$), it makes sense to assume that the radiation density becomes infinitively large. If so, from (10.6) we must demand that:

$$B_{12} = B_{21} \tag{10.7}$$

Hence, the postulation of the stimulated emission appears to be absolutely necessary to suffice thermodynamics.

In fact, we do not need to rely on our feeling of an infinitively large spectral density at infinitively large temperatures. Condition (10.7) will be obtained independently as a result of the following perturbation theory treatment of quantum transitions.

10.4 Extended Detail: Perturbation Theory of Quantum Transitions

In order to get information about the mathematical structure of Einstein coefficients, it becomes now necessary to apply the mathematical apparatus of quantum mechanics to the interaction of light with matter. In quantum mechanics, instead of a Hamilton function familiar from classical theoretical mechanics, we deal with the Hamilton operator. It is obtained from the classical Hamilton function substituting coordinates and moments by the corresponding quantum mechanical operators. The behaviour of the system is described by a wavefunction Ψ, obtained as the solution of Schrödinger's equation:

$$\mathrm{i}\hbar \frac{\partial}{\partial t} \Psi(\boldsymbol{r}, t) = \boldsymbol{H} \Psi(\boldsymbol{r}, t) \tag{10.8}$$

where the wavefunction Ψ depends on the coordinates and on the time. If the Hamilton operator (or *Hamiltonian*) does not explicitly depend on time, then the time independent Schrödinger's equation follows from (10.8) after substituting:

$$\Psi(\boldsymbol{r}, t) = \mathrm{e}^{-\frac{\mathrm{i}}{\hbar} E t} \times \psi(\boldsymbol{r})$$

We obtain:

$$\boldsymbol{H} \psi(\boldsymbol{r}) = E \psi(\boldsymbol{r}) \tag{10.9}$$

where E is again the energy. Equation (10.9) represents an eigenvalue problem, and the time independent eigenfunctions $\psi_n(\boldsymbol{r})$ as well as the eigenvalues E_n may be obtained solving (10.9). The eigenvalues E_n have to be regarded as the allowed energy levels of the system. For simplicity, throughout this

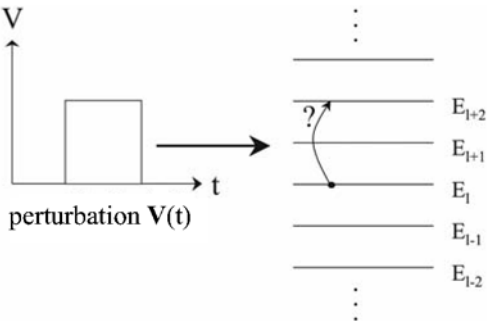

Fig. 10.2. Quantum transitions

chapter we will assume that the energy levels are generally discrete and non-degenerated. The quantum number n will often be used to count the energy levels and wavefunctions, a confusion with the refractive index should not occur. Furtherly, let us recall that the wavefunctions (or eigenfunctions of the Hamiltonian) are mutually orthogonal and normalized to the value 1.

The particular problem which will be considered now is sketched in Fig. 10.2. Imagine a time-independent Hamiltonian $\boldsymbol{H_0}$ with a set of eigenfunctions and corresponding eigenvalues $\{E_n\}$. Consider further, that at a certain moment, say at $t = 0$, the system is definitely in the l-th quantum state and has the energy E_l.

In this case, the wavefunction $\Psi_l(\boldsymbol{r}, t)$ suffices Schrödinger's equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\Psi_n(\boldsymbol{r}, t) = \boldsymbol{H_0}\Psi_n(\boldsymbol{r}, t) \tag{10.10}$$

while the energy level E_l is a solution of the eigenvalue problem

$$\boldsymbol{H_0}\psi_n(\boldsymbol{r}) = E_n\psi_n(\boldsymbol{r}) \tag{10.11}$$

and

$$\Psi_l(\boldsymbol{r}, t) = \mathrm{e}^{-\mathrm{i}\frac{E_l}{\hbar}t}\psi_l(\boldsymbol{r}) \tag{10.12}$$

We will now make the situation more complicated. Let us illuminate the regarded system with light. The light source will be switched on at the moment $t = 0$. We will now have a completely different situation. The system is no more described by the time-independent Hamiltonian $\boldsymbol{H_0}$. Instead, the full Hamilton operator is now given by:

$$\boldsymbol{H} = \boldsymbol{H}(t) = \boldsymbol{H_0} + \boldsymbol{V}(t) \tag{10.13}$$

where the certainly time-dependent perturbation operator $\boldsymbol{V}$ describes the interaction between the light and the system.

Finally, let us switch off the light at $t = t_0$. Again, (10.10)–(10.12) are valid for the system. The question is: Is there any chance to find the system

now in a quantum state m different from that which has been occupied at $t = 0$? If yes, then we will state that the perturbation $\boldsymbol{V}$ has caused a quantum transition between the states l and $m \neq l$. It is our task now to understand the conditions necessary for such a transition.

First of all, let us introduce the common terminology:

- If the probability of the transition $l \rightarrow m$ is equal to zero, then the transition is called *forbidden* with respect to the given perturbation $\boldsymbol{V}$.
- If the probability of the transition $l \rightarrow m$ is larger than zero, then the transition is called *allowed* with respect to the given perturbation $\boldsymbol{V}$.
- The recipe which classifies any transition as allowed or forbidden is called a *selection rule*.

Clearly, Einstein's coefficient B_{12} must be correlated to the transition probability. In particular, for a forbidden transition, the Einstein coefficient should be zero.

Let us now turn to the mathematics. We tackle the interesting time interval:

$$0 < t < t_0 \,.$$

Because the perturbation may be time-dependent, we have to regard the time-dependent Schrödinger's Equation:

$$\mathrm{i}\hbar \frac{\partial}{\partial t} \Psi(\boldsymbol{r}, t) = \boldsymbol{H} \Psi(\boldsymbol{r}, t) \tag{10.14}$$

with the Hamiltonian (10.13). In order to find the solution, the unknown wavefunction is expanded into a series of eigenfunctions of the unperturbated Hamiltonian $\boldsymbol{H_0}$:

$$\Psi(\boldsymbol{r}, t) = \sum_n a_n(t) \Psi_n(\boldsymbol{r}, t) \tag{10.15}$$

where the expansion coefficients a_n may also depend on time. According to the normalization condition, we have:

$$\sum_n |a_n(t)|^2 = 1 \tag{10.16}$$

The system as described by (10.15) is in a quantum superposition state. Following the usual interpretation of quantum mechanics, the value

$$|a_n(t)|^2$$

has to be understood as the probability to find the system in the n-th quantum state, when the superposition state (10.15) is destroyed as the result of a measurement procedure.

It is therefore the time evolution of the expansion coefficients $\{a_n\}$ that is most interesting for us. Substituting the wavefunction in (10.14) by (10.15) yields:

$$\begin{aligned}
\mathrm{i}\hbar\frac{\partial}{\partial t}\Psi\left(\boldsymbol{r},t\right) &= \mathrm{i}\hbar\frac{\partial}{\partial t}\left[\sum_n a_n\left(t\right)\Psi_n\left(\boldsymbol{r},t\right)\right] \\
&= \mathrm{i}\hbar\sum_n \Psi_n\left(\boldsymbol{r},t\right)\frac{\partial}{\partial t}a_n\left(t\right) + \mathrm{i}\hbar\sum_n a_n\left(t\right)\frac{\partial}{\partial t}\Psi_n\left(\boldsymbol{r},t\right) \\
&= \boldsymbol{H}\Psi\left(\boldsymbol{r},t\right) = \boldsymbol{H}\sum_n a_n\left(t\right)\Psi_n\left(\boldsymbol{r},t\right) \qquad (10.17)\\
&= \boldsymbol{H_0}\sum_n a_n\left(t\right)\Psi_n\left(\boldsymbol{r},t\right) + \boldsymbol{V}\sum_n a_n\left(t\right)\Psi_n\left(\boldsymbol{r},t\right) \\
&\Rightarrow \mathrm{i}\hbar\sum_n \Psi_n\left(\boldsymbol{r},t\right)\frac{\partial}{\partial t}a_n\left(t\right) = \boldsymbol{V}\sum_n a_n\left(t\right)\Psi_n\left(\boldsymbol{r},t\right)
\end{aligned}$$

because for every n we have:

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\Psi_n\left(\boldsymbol{r},t\right) = \boldsymbol{H_0}\Psi_n\left(\boldsymbol{r},t\right)$$

Let us now see whether or not the perturbation is able to transfer the system from state l to state m. We multiply (10.17) from the left side with the conjugate complex function $\Psi_m^*\left(\boldsymbol{r},t\right)$ and integrate over all coordinates. Due to the normalization and orthogonality of wavefunctions, we have:

$$\int \Psi_m^*\left(\boldsymbol{r},t\right)\Psi_n\left(\boldsymbol{r},t\right)\mathrm{d}\boldsymbol{r} = \delta_{mn}$$

and therefore, from (10.17) we find:

$$\mathrm{i}\hbar\dot{a}_m = \sum_n a_n\int \Psi_m^*\left(\boldsymbol{r},t\right)\boldsymbol{V}\Psi_n\left(\boldsymbol{r},t\right)\mathrm{d}\boldsymbol{r} \qquad (10.18)$$

As following from (10.12), we may write:

$$\begin{aligned}
\Psi_m^*\left(\boldsymbol{r},t\right) &= \mathrm{e}^{\mathrm{i}\frac{E_m}{\hbar}t}\psi_m^*\left(\boldsymbol{r}\right) \\
\Psi_n\left(\boldsymbol{r},t\right) &= \mathrm{e}^{-\mathrm{i}\frac{E_n}{\hbar}t}\psi_n\left(\boldsymbol{r}\right)
\end{aligned}$$

Let us introduce the transition angular frequency according to:

$$\omega_{mn} \equiv \frac{E_m - E_n}{\hbar} \qquad (10.19)$$

Equation (10.18) may then be rewritten as:

$$\mathrm{i}\hbar\dot{a}_m = \sum_n a_n V_{mn}\mathrm{e}^{\mathrm{i}\omega_{mn}t} \qquad (10.20)$$

where the so-called matrix element V_{mn} is defined as:

$$V_{mn} \equiv \int \psi_m^* (\boldsymbol{r})\, \boldsymbol{V} \psi_n (\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \tag{10.21}$$

Let us now formulate the initial conditions. At $t = 0$, we require:

$$|a_l| = 1\,; \quad a_{n \neq l} = 0$$

Particularly, at $t = 0$ we have $a_m = 0$. As long as a_l is close to one by modulus, it may be regarded as constant, and a population of the m-th state would then require:

$$i\hbar \dot{a}_m|_{t \geq 0} = a_l V_{ml} \mathrm{e}^{\mathrm{i}\omega_{ml} t} \neq 0 \tag{10.22}$$

To fulfil (10.22), it is absolutely necessary that the matrix element V_{ml} is different from zero. What we have found this way is the general formulation of a selection rule: *Given a perturbation* $\boldsymbol{V}$*, it can only cause a quantum transition between the states* l *and* m *when the corresponding matrix element of the perturbation operator* V_{ml} *is different from zero.*

Let us now regard the concrete case of the interaction of a microscopic quantum system with light. When the spatial extension of the system is much smaller than the wavelength, we can neglect the spatial structure of the wave and regard a homogeneous but oscillating electric field. In that dipole approximation, the perturbation operator may be written as:

$$\boldsymbol{V} = -\boldsymbol{pE} = -\boldsymbol{pE_0} \cos \omega t = -\frac{1}{2} \left(\boldsymbol{pE_0}\, \mathrm{e}^{-\mathrm{i}\omega t} + \boldsymbol{pE_0}\, \mathrm{e}^{\mathrm{i}\omega t} \right) \tag{10.23}$$

with $\boldsymbol{E}$ as the electric field vector and $\boldsymbol{E_0}$ its amplitude. From (10.22) we find:

$$a_m (t) = a_l \frac{\boldsymbol{p}_{ml} \boldsymbol{E_0}}{2\hbar} \left\{ \frac{\mathrm{e}^{\mathrm{i}(\omega_{ml} - \omega)t} - 1}{\omega_{ml} - \omega} + \frac{\mathrm{e}^{\mathrm{i}(\omega_{ml} + \omega)t} - 1}{\omega_{ml} + \omega} \right\} \tag{10.24}$$

which is valid as long as $|a_l| \approx 1$. $\boldsymbol{p}_{ml}$ is the matrix element of the dipole operator. From (10.24) we see that for a dipole transition, it is the matrix element of the dipole operator that needs to be different from zero. Moreover, we recognize that the transition frequency ω_{ml} plays the role of the resonance frequency: The closer the frequency of the electric field to one of the transition frequencies ω_{ml} or ω_{lm} is, the more probable the transition becomes.

In order to compare this result to (10.3), we will again assume that we deal with an assembly of quantum systems (for example atoms or molecules), while N_l is the number of systems in the l-th quantum state. Moreover, the transition rate between the states l and m is given by the expression:

$$\frac{\mathrm{d}N_m}{\mathrm{d}t} \propto \frac{\mathrm{d}}{\mathrm{d}t} |a_m|^2 \propto |p_{ml}|^2 |E_0|^2 N_l \tag{10.25a}$$

Interchanging the indices, we obtain

$$\frac{\mathrm{d}N_l}{\mathrm{d}t} \propto \frac{\mathrm{d}}{\mathrm{d}t} |a_l|^2 \propto |p_{lm}|^2 |E_0|^2 N_m \tag{10.25b}$$

Obviously, when comparing with (10.3), one sees that one of the above relations (10.25) should correspond to absorption, and the other one to stimulated emission of light, depending on whether $E_m > E_l$ or vice versa. But the proportionality factors are completely identical, because, due to the hermitecity of the dipole operator, we have:

$$|p_{ml}|^2 = |p_{lm}|^2$$

For that reason, the Einstein's coefficient B_{12} and B_{21} must be identical. On the other hand, from (10.25) it turns out, that

$$B_{12} \propto |p_{12}|^2 \tag{10.26}$$

In fact, this is the most important conclusion for our further treatment of Einstein coefficients. The second conclusion is, that, according to (10.19), resonance of the radiation with the two-level system is expected to occur when the condition:

$$\omega = \omega_{21} \equiv \frac{E_2 - E_1}{\hbar} \tag{10.27}$$

is fulfilled. Correspondingly, the energy of the photon as defined before must be equal to $\hbar\omega_{21}$.

Finally, let us rewrite (10.6) taking our new findings into account. We shall write:

$$\text{equilibrium:} \quad u(\omega_{21}) = \frac{A_{21}}{B_{12}\left(\mathrm{e}^{\frac{\hbar\omega_{21}}{k_\mathrm{B}T}} - 1\right)} \tag{10.28}$$

10.5 Extended Detail: Planck's Formula

10.5.1 Idea

Equation (10.28) describes the spectral density of irradiation in equilibrium with an assembly of two-level-systems, held at a temperature T. The purpose of this section is to independently derive an alternative expression for this spectral density, which might be compared to (10.28), and will therefore give us an expression for the ratio between the coefficients A_{21} and B_{21}. The formula we will obtain is well known as the famous Planck's formula. We start with the definition (10.1):

$$u \equiv \frac{\mathrm{d}E}{V\mathrm{d}\omega}$$

The energy per angular frequency interval may be expressed as the energy per photon ($\hbar\omega$) with an angular frequency ω, multiplied with the average

number of photons expected to be excited in a corresponding quantum state ($\langle N \rangle$), again multiplied with the number of those quantum states per angular frequency interval at the relevant frequency (the density of states $\frac{\mathrm{d}Z}{\mathrm{d}\omega}$). Hence, we make use of the approach:

$$\frac{\mathrm{d}E}{\mathrm{d}\omega} = \hbar\omega \langle N \rangle \frac{\mathrm{d}Z}{\mathrm{d}\omega} \tag{10.29}$$

It is now our task to calculate the single terms encountering into (10.29). Let us start with the average number of photons $\langle N \rangle$.

10.5.2 Planck's Distribution

In order to obtain an expression for $\langle N \rangle$ in the equilibrium case, let us calculate the energy accumulated in a quantum state when a number of N photons each with an energy $\hbar\omega$ is excited. Obviously, its energy will be $N\,\hbar\omega$. In equilibrium, the probability w to find the state with N photons excited is given by Boltzmann's factor:

$$w(N) = \frac{x^N}{\sum\limits_N x^N} \quad \text{with} \quad x \equiv \mathrm{e}^{-\frac{\hbar\omega}{k_\mathrm{B}T}}$$

The average number of photons excited in the quantum state is now calculated in the usual way:

$$\begin{aligned} \langle N \rangle &= \sum_N N w(N) = \frac{\sum\limits_N N x^N}{\sum\limits_N x^N} = x \frac{\mathrm{d}}{\mathrm{d}x} \ln \sum_N x^N \\ &= x \frac{\mathrm{d}}{\mathrm{d}x} \ln (1-x)^{-1} = \frac{x}{1-x} \Rightarrow \\ \langle N \rangle &= \frac{1}{\mathrm{e}^{\frac{\hbar\omega}{k_\mathrm{B}T}} - 1} \end{aligned} \tag{10.30}$$

Equation (10.30) is known as Planck's distribution for photons. It is a special case of Bose–Einstein statistics.

10.5.3 Density of States

What remains to determine is the density of states. Let us start with a simple one-dimensional problem, namely the one-dimensional movement (for example along the x-axis) of a particle between two impermeable walls. The corresponding wavefunction is sketched in Fig. 10.3. Of course, we will further assume that the moving 'particle' is nothing else than a photon.

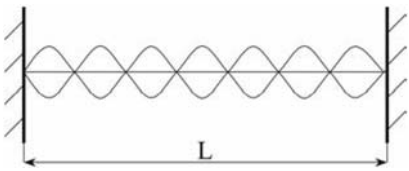

Fig. 10.3. Standing wave as solution of Schrödinger's Equation for a particle between two walls

When the separation between the two walls is L, it is known from quantum mechanics that the allowed eigenstates of the system correspond to standing waves as shown in Fig. 10.3, so that the allowed wavelength-values become:

$$\lambda_n = \frac{2L}{n_x};\ n_x = 1, 2, 3, \ldots \tag{10.31}$$

That corresponds to allowed values of the wavevector:

$$k_x = \frac{2\pi}{\lambda} = \frac{\pi n_x}{L} \tag{10.32}$$

where n_x is now a quantum number.

Let us generalize our result to the three-dimensional case. Instead of the system like shown in Fig. 10.3, we should now imagine a hollow cube with a volume $V = L^3$, and count the allowed states inside the cube. Accordingly, we obtain for the wavevector:

$$k^2 = k_x^2 + k_y^2 + k_z^2 = \left(\frac{\pi}{L}\right)^2 \left(n_x^2 + n_y^2 + x_z^2\right) \equiv \left(\frac{\pi}{L} n\right)^2 \tag{10.33}$$

Note that the thus defined value n is not a quantum number. For suffciently high n_x, n_y, and n_z, n may be regarded as a continuous function. In particular, it is possible to determine the number of states in an n-interval $\frac{\mathrm{d}Z}{\mathrm{d}n}$. To do so, let us have a look at Fig. 10.4. It visualizes the n-space occupied by the states corresponding to n-values between zero and a given maximum value of n. Each state corresponds to a certain triple of n_x, n_y, and n_z, and therefore occupies a cube of the volume 1 in the n-space. The volume of the sphere with the radius n in Fig. 10.4 would therefore correspond to the full number of states, however, due to the circumstance that n_x, n_y, and n_z shall not be negative, we should only count the states in the first octant. Hence, we find for the full number of states Z:

$$Z = \frac{1}{8} \times \frac{4\pi}{3} n^3 \times 2$$

The factor 2 has been introduced to account for the degeneracy of photons with respect to their polarization direction.

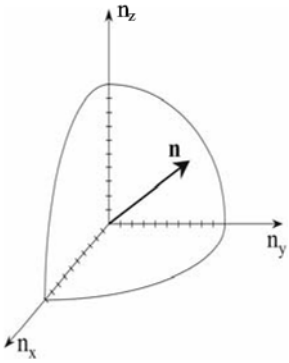

Fig. 10.4. To the calculation of the number of quantum states

We therefore have:

$$Z = \frac{\pi}{3} n^3 \Rightarrow \frac{\mathrm{d}Z}{\mathrm{d}n} = \pi n^2 \tag{10.34}$$

From (10.33) it turns out that

$$k^2 = \frac{\omega^2}{c^2} = \left(\frac{\pi}{L} n\right)^2 \Rightarrow n^2 = \left(\frac{L\omega}{\pi c}\right)^2 \tag{10.35}$$

and

$$\frac{\mathrm{d}n}{\mathrm{d}\omega} = \frac{L}{\pi c} \tag{10.36}$$

Equations (10.34)–(10.36) in combination finally yield:

$$\frac{\mathrm{d}Z}{\mathrm{d}\omega} = \frac{\mathrm{d}Z}{\mathrm{d}n}\frac{\mathrm{d}n}{\mathrm{d}\omega} = \pi \left(\frac{L\omega}{\pi c}\right)^2 \frac{L}{\pi c} = \frac{L^3\omega^2}{c^3\pi^2} = V\frac{\omega^2}{c^3\pi^2} \tag{10.37}$$

We have now calculated all values that encounter into (10.29). In sum, we find:

$$\frac{\mathrm{d}E}{\mathrm{d}\omega} = \hbar\omega\langle N\rangle\frac{\mathrm{d}Z}{\mathrm{d}\omega} = \hbar\omega\frac{1}{\mathrm{e}^{\frac{\hbar\omega}{k_\mathrm{B}T}} - 1} V\frac{\omega^2}{c^3\pi^2} = V\frac{\hbar\omega^3}{c^3\pi^2}\frac{1}{\mathrm{e}^{\frac{\hbar\omega}{k_\mathrm{B}T}} - 1}$$

and

$$u\left(\omega, T\right) = \frac{\hbar\omega^3}{c^3\pi^2}\frac{1}{\mathrm{e}^{\frac{\hbar\omega}{k_\mathrm{B}T}} - 1} \tag{10.38}$$

Equation (10.38) represents Planck's famous formula for the so-called black-body irradiation.

10.6 Extended Detail: Expressions for Einstein Coefficients in the Dipole Approximation

Equation (10.38) describes the spectral density of electromagnetic irradiation in a hollow cube with walls held at a temperature T in the equilibrium case. Hence, for any selected angular frequency ω, the absorption of photons caused by the atoms or molecules of the wall is compensated by the processes of spontaneous and stimulated emission of photons by the same atoms or molecules. This is exactly the situation for which (10.28) has been derived. Therefore, comparing (10.28) and (10.38), we find an important relationship between Einstein coefficients, namely:

$$\frac{A_{21}}{B_{21}} = \frac{\hbar\omega^3}{\pi^2 c^3} \tag{10.39}$$

Therefore, only one of Einstein's coefficient remains to be determined. In the forthcoming, we will directly calculate the coefficient A_{21} making use of the correspondence principle. But before doing so, let us make one remark concerning (10.39):

The relation between A_{21} (efficiency of spontaneous emission processes) and B_{21} (efficiency of stimulated processes) turns out to be strongly frequency-dependent. At low frequencies, the stimulated processes dominate, while at higher frequencies spontaneous processes become more efficient. The practical conclusion is, that in infrared spectroscopy (low frequencies) one usually works with stimulated processes (absorption spectroscopy), while in VIS and UV-analytics (much higher frequencies), fluorescence spectroscopy has become an utmost important spectroscopic tool.

Let us now come to the derivation of A_{21} in the dipole approximation.

First of all, we shall remember that the process of spontaneous emission of light occurs regardless of the presence or absence of an incident light wave. However, in the philosophy of Sect. 10.3, equation (10.24), no quantum transitions are allowed to occur when the field strength of the exciting wave is zero. Hence, our theory as developed so far does generally not allow for any spontaneous transition processes. Nevertheless these transitions occur in real life. So what is the reason for the discrepancy between our theory and experiment?

It turns out, that it is the ansatz for the perturbation operator (10.23) that is incompatible with the existence of spontaneous quantum transitions resulting in the emission of a photon. In classical electrodynamics, the energy of the electromagnetic field is zero in the case of a vanishing field strength. A complete quantum mechanical treatment (including the quantization of the field itself) will lead to a somewhat different result. As in the case of the quantum mechanical harmonic oscillator, the electromagnetic field is expected to have eigenvalues of the energy given by $\hbar\omega\,(N + 1/2)$. As in Sect. 10.5, N is the number of photons. Note that this expression is different from what has been assumed deriving (10.30), but one may easily check that this will

not influence the results obtained in Sect. 10.5.2. In particular, it turns out that in the absence of photons, the field still has a ground-state-energy of $\hbar\omega/2$. These "zero-oscillations" of the electromagnetic field in the absence of photons are the quantum mechanical perturbations that cause the spontaneous effects in optical spectroscopy, among them the spontaneous emission of light.

As mentioned before, we will not derive the mathematical apparatus of the second quantization in the frames of this book. Instead, we will use a somewhat different approach that is based on the correspondence principle. Recalling our knowledge on the structure of Einstein coefficients, from (10.26) and (10.39) and considering that $p = qx$ we may write:

$$A_{21} = C|x_{21}|^2 \tag{10.40}$$

C is a constant, and it may be determined regarding any special case that is accessible to analytic calculations. Having found C, we may write down the final expression for Einstein coefficients. We choose the particular case of a harmonic oscillator and calculate the decay rate of the energy accumulated in the oscillator, when the latter is allowed to dissipate as a result of the spontaneous emission of photons. From the correspondence principle, we may write:

$$E \to \infty : \left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_{\text{classics}} = \left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_{\text{quantum mechanics}} \tag{10.41}$$

That means, that for sufficiently high energies, the quantum mechanical expressions shall become identical to the classical ones. From classical electrodynamics, we know that:

$$\left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_{\text{classics}} = \frac{q^2}{6\pi\varepsilon_0 c^3}\left\langle \ddot{x}^2 \right\rangle\Big|_t \tag{10.42}$$

The average is taken over a relevant time period, say the duration of one oscillation, the latter being performed along the x-axis. For the motion of a classical harmonic oscillator, we have:

$$x = x_0 \cos\omega t \quad \text{with} \quad E = \frac{m\omega^2 x_0^2}{2} \tag{10.43}$$

When the amplitude of the oscillation x_o does not seriously change during one period, after averaging from (10.42) and (10.43) we find:

$$\left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_{\text{classics}} = \frac{q^2\omega^2}{6\pi\varepsilon_0 c^3 m}E \tag{10.44}$$

Let us now turn to the quantum mechanical case. The energy of the harmonic oscillator is given by:

$$E_n = \hbar\omega\left(n + \frac{1}{2}\right) \Rightarrow E_{n\to\infty} \approx \hbar\omega n \tag{10.45}$$

Any quantum transition from level n to $n-1$ leads to an energy decay per time interval given by (compare (10.3) and (10.40)):

$$\left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_{\text{quantum mechanics}} = \hbar\omega A_{n.n-1} = \hbar\omega C|x_{n,n-1}|^2 \tag{10.46}$$

The matrix elements for the coordinate x of a harmonic oscillator are well known from quantum mechanics. They are given by:

$$|x_{n,n-1}|^2 = \frac{\hbar n}{2\omega m} \approx \frac{E}{2\omega^2 m} \tag{10.47}$$

That leads us to:

$$\left.\frac{\mathrm{d}E}{\mathrm{d}t}\right|_{\text{quantum mechanics}} = \frac{\hbar C E}{2\omega m} \tag{10.48}$$

The constant C may now easily be found combining (10.41), (10.44) and (10.48). We obtain:

$$C = \frac{q^2\omega^3}{3\varepsilon_0\pi\hbar c^3}$$

So that from (10.40) Einstein's coefficient A_{21} is found as:

$$A_{21} = \frac{q^2\omega^3|x_{21}|^2}{3\varepsilon_0\pi\hbar c^3} = \frac{\omega^3|p_{21}|^2}{3\varepsilon_0\pi\hbar c^3} \tag{10.49}$$

The other coefficients follow from (10.39) and (10.7):

$$B_{21} = B_{12} = \frac{\pi|p_{21}|^2}{3\varepsilon_0\hbar^2} \tag{10.50}$$

Expressions (10.49) and (10.50) represent the final expressions for Einstein coefficients as introduced ins Sect. 10.3, (10.3), in the dipole approximation.

As this was a rather long and complex derivation, it might be helpful to provide an overview that recalls the main steps of the derivation of Einstein coefficients. This is done in Table 10.1.

In finishing this section, let us make two final remarks:

Firstly, our derivation was based on the perturbation operator (10.23), which describes the electrical dipole interaction. Therefore, the resulting Einstein coefficients are only valid in the (electric) dipole approximation. If, for any reason, that dipole transition is forbidden, nevertheless quantum transitions may occur as the result of other types of interaction – for example magnetic dipole interaction, electric quadrupole interaction and so on. The corresponding Einstein coefficients may be derived analogously. However, when the electric dipole transition is allowed, in the non-relativistic case it is usually much stronger than the other interaction terms, so that it

Table 10.1. Derivation of Einstein coefficients

Section	Equation	Result
10.3	(10.3)	Introduction/Definition of Einstein coefficients A_{21}, B_{21}, and B_{12}
10.3	(10.7)	$B_{12} = B_{21}$ from thermodynamical considerations
10.4	(10.26)	From perturbation theory it follows, that $B_{12} \propto \lvert p_{12}\rvert^2$
10.5		Derivation of Planck's formula
10.6	(10.39)	As a consequence of Planck's formula, we find $\frac{A_{21}}{B_{21}} = \frac{\hbar\omega^3}{\pi^2 c^3}$
10.6	(10.49)	Calculation of A_{21} by means of the correspondence principle, basing on the assumption: $A_{21} = C\lvert x_{21}\rvert^2$
$\Rightarrow$ Final Expressions: $A_{21} = \frac{\omega^3 \lvert p_{21}\rvert^2}{3\varepsilon_0 \pi \hbar c^3}$; $B_{21} = B_{12} = \frac{\pi \lvert p_{21}\rvert^2}{3\varepsilon_0 \hbar^2}$		

is often sufficient to regard this first term in the multipole expansion of the interaction of an electromagnetic wave with matter.

In the absence of incident irradiation ($u = 0$), the population of the excited quantum state decays according to (compare (10.3):

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -N_2 A_{21}$$

so that we find:

$$N_2 = N_{20}\mathrm{e}^{-A_{21}t} \equiv N_{20}\mathrm{e}^{-\frac{t}{\tau}} \tag{10.51}$$

Therefore, the reciprocal value of A_{21} may be interpreted as the lifetime of an excited quantum level. Note the similarity between (10.51) and (4.1), (4.2). In fact, expression (10.49) enables us to estimate the energy decay time as introduced in Sect. 4.1, and consequently the natural linewidth as long as the latter is determined by radiative relaxation processes only.

From (10.49) and (10.51), the radiative lifetime of the excited state in a two-level system in the dipole approximation turns out to be equal to:

$$\tau = \frac{3\varepsilon_0 \pi \hbar c^3}{q^2 \omega^3 |x_{21}|^2} = \frac{3\varepsilon_0 \pi \hbar c^3}{\omega^3 |p_{21}|^2} \tag{10.52}$$

For well allowed dipole transitions, (10.52) yields lifetimes of the order of 10^{-8}s. A measurement of these relaxation processes requires the application of ultrafast spectroscopic tools. On the other hand, the lifetime becomes infinitively large, when the matrix element of the dipole operator vanishes. When we excite (by any means) a quantum level that cannot relax into the ground state via dipole irradiation, then such a state may remain excited for a considerably long time. Of course, this time will be not infinitively large, because in fact there are still other relaxation channels than the electric dipole irradiation. But the lifetime may easily extend for minutes or hours, and this is the reason for the phosphorescent behaviour of various materials.

10.7 Lasers

10.7.1 Population Inversion and Light Amplification

Let us now come to an utmost important practical application of the theoretical apparatus derived in this chapter so far. From (10.2), (10.3) and (10.7) we find, that

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = (N_1 - N_2)\,(2B_{12}u + A_{21}) - N_0 A_{21} \tag{10.53}$$

In the stationary case, when $\mathrm{d}N_2/\mathrm{d}t = 0$, we obtain the stationary solution:

$$(N_1 - N_2) = \frac{N_0 A_{21}}{2B_{12}u + A_{21}} > 0 \quad \forall u \tag{10.54}$$

No matter how intense the field is, as long as we deal with a two-level system, the stationary population of the ground state will always be higher than that of the excited one. Of course, this conclusion is only true as long as the excited level may exclusively be populated by optical pumping from the ground state. In this case, it is impossible to achieve a stationary population inversion ($N_2 > N_1$). On the contrary, for a sufficiently high u, N_1 and N_2 become nearly equal to each other. In this case, the transition $1 \to 2$ is called to be saturated.

On the other hand, a population inversion (if it could be achieved anyway) would offer prospective physical effects. Let us for a moment assume, that we prepared the system in a way that $N_2 > N_1$ is fulfilled. From (10.53) we find, that

$$N_2 > N_1 \Rightarrow \frac{\mathrm{d}N_2}{\mathrm{d}t} < 0.$$

As long as we have population inversion in a two level system, absorption and emission processes in sum tend to transfer the system from the excited to the ground state. This conclusion is particularly true when the spontaneous

emission processes may be neglected. Consequently, the energy is transferred from the two-level systems to the radiation field. An incident light beam may therefore be magnified in intensity when travelling through a medium with population inversion. The preparation of population inversion is therefore important for the construction of light amplifiers.

This simple discussion leads us to a very important conclusion on the structure of the absorption coefficient in quantum mechanics. We must assume, that the absorption coefficient explicitly depends on the population difference $N_1 - N_2$. Particularly, the sign of the population difference is crucial for the decision whether the material is absorbing (positive absorption coefficient) or amplifying (negative absorption coefficient). In saturation conditions, one should expect that absorption and stimulated emission processes compensate each other, hence in this case a light beam would travel through the medium without any damping or amplification. Such a medium appears to be transparent, and the corresponding absorption coefficient is zero.

10.7.2 Feedback

Let us now assume, that we are able to prepare a two-level system with population inversion. In practice, the population inversion may be achieved for example by collisions of electrons with atoms, which transfer the atoms into an excited state. For example, gas lasers such as the Helium-Neon-laser work on this principle. Another way is to use three- or four level systems with optical pumping, as shown in Fig. 10.5.

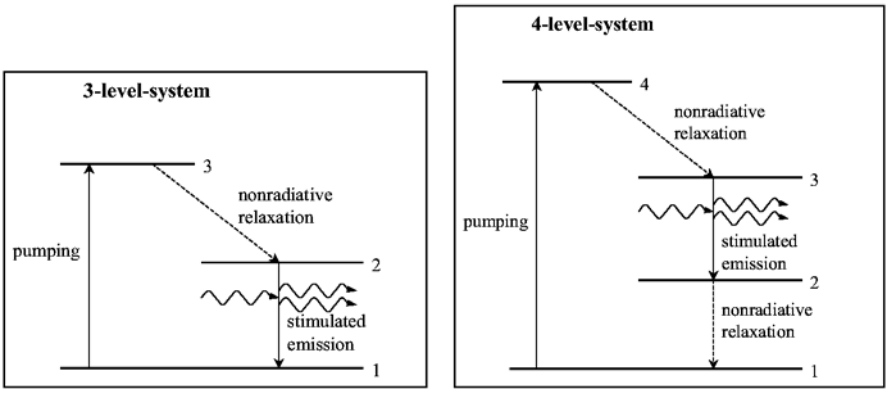

Fig. 10.5. Three- and four-level systems that allow to create population inversion by optical pumping. In the three-level-system, population inversion is accomplished between the first and second state. This is what is done in the ruby laser. In the four-level-system, one achieves population inversion between the second and third state (for example in the Neodymium-YAG laser)

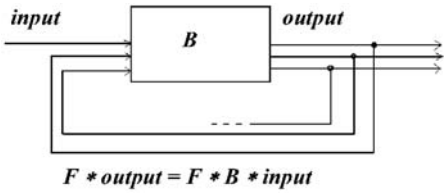

Fig. 10.6. An amplifier with feedback

In our further discussion we will simply assume, that the population inversion has been achieved anyway. In the terminology of laser physics, such a medium is called an *active medium*. Keeping in mind that the absorption coefficient of an active medium is negative, from Lambert's law (2.19) we find:

$$I = I_0 \, \mathrm{e}^{-\alpha x} = I_0 \, \mathrm{e}^{|\alpha| x} \tag{10.55}$$

Hence, in an active medium, the intensity of a propagating wave is expected to exponentially grow in intensity. What we obtain this way is an amplifier of electromagnetic waves.

In complete analogy to electronics, we only need to add a positive feedback to an amplifier in order to build a generator of electromagnetic waves. That kind of light amplification due to stimulated emission, combined with a feedback mechanism, leads us to a specific kind of light source that is called a laser.

Let us have a look at Fig. 10.6. Figure 10.6 sketches the idea of combining an amplifying element with a feedback mechanism. Let us start from the left side. We assume an external input, for example an electromagnetic wave. The action of the amplifying element shall be to magnify the input by a complex factor B. After passing the amplifier for one time, we obtain an output according to:

$$\text{output} = B \times \text{input}\,, \quad |B| > 1 \tag{10.56}$$

Let us now discuss what happens when the mentioned feedback mechanism comes into play. A part of the output (say, the nominal output multiplied with a constant F, where F is again a possibly complex number with $|F| < 1$) is transferred back to the input side, again amplified and so on. Then, instead of the simple output as given by (10.56) we get an effective output obtained as the result of an infinite number of loops through the amplifier caused by the feedback mechanism. Mathematically, this may be expressed in the following manner:

$$\begin{aligned} \text{effective output} &= \left[(1-F) + (1-F)\,BF + (1-F)\,B^2F^2 + \ldots\right] \text{output} \\ &= B\,(1-F) \sum_{j=0}^{\infty} (BF)^j \ \text{input} \equiv B_{\text{eff}} \ \text{input} \end{aligned} \tag{10.57}$$

with $$B_{\text{eff}} = (1-F)\,B \sum_{j=0}^{\infty} (FB)^j = \frac{(1-F)\,B}{1-FB} \ \text{if} \ |FB| < 1.$$

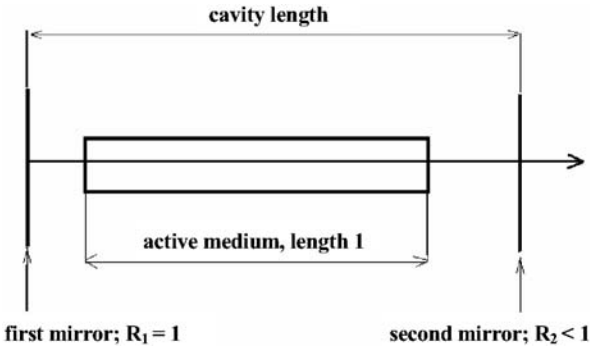

Fig. 10.7. Simple laser geometry

Particularly, (10.57) leads to an infinitively large effective enhancement factor in the limit $BF \to 1$. In this case, two conditions must be fulfilled:

$$\begin{aligned} |B| &\to \left|F^{-1}\right| \\ \varphi_B + \varphi_F &= 2j\pi \quad \text{where } j \text{ is an integer} \\ \text{assuming:} \quad B &\equiv |B|\,\mathrm{e}^{\mathrm{i}\varphi_B}\,; \quad F \equiv |F|\,\mathrm{e}^{\mathrm{i}\varphi_F} \end{aligned} \tag{10.58}$$

In the case that the effective enhancement becomes infinitively large, an arbitrarily small input (for example a single photon, accidentally emitted as the result of a spontaneous emission process) may lead to a finite effective output of the system (in our case to electromagnetic irradiation). There is no contradiction to energy conservation, because the amplifying element (in our case the active medium) is pumped by an external energy source. Such a system works as a generator of light and is called a laser. Of course, it will also generate light when the light amplification defined by B is larger than the threshold value defined by (10.58).

Consequently, we have to fulfil two conditions in order to construct a generator of light. First of all, we have to take care that the light amplification is large enough to compensate any losses of light that leave the system. Technically, this is accomplished through a sufficiently high population inversion in the active medium, and the corresponding mathematical criterion is called the *laser condition*, but it will not be derived here.

The necessary feedback is usually achieved placing the active medium into a resonator (the cavity), which may be built up by two parallel mirrors. This situation looks similar to what has been sketched previously in Fig. 10.3. Let L be the cavity length. Usually, the active medium does not fill the full cavity, but extends for a smaller length l. This situation is shown in Fig. 10.7.

Let us now rewrite (10.58) for the special case of the geometry from Fig. 10.7. Imagine a light wave, that performs one loop in the cavity. Let us perform the discussion of the light amplification in terms of the light intensity, and not of the electric field. Performing one loop in the cavity, the

wave crosses the active medium for two times, and its intensity will grow by a factor $\mathrm{e}^{2|\alpha|l}$. On the other hand, at the second mirror some light will escape from the cavity, so that after one loop, the intensity of the light wave becomes:

$$I\,(\text{after one loop}) = I_0\,\mathrm{e}^{2|\alpha|l} R_2 \tag{10.59}$$

Equation (10.59) describes an extremely simplified case, because no loss mechanisms in the cavity have been taken into account. But the principle is nevertheless clear. The value $\sqrt{R_2 \mathrm{e}^{2|\alpha|l}}$ is nothing else than the absolute value of the product BF as fixed in (10.58). Consequently, for light generation it is necessary that

$$\sqrt{R_2 \mathrm{e}^{2|\alpha|l}} \geq 1 \Rightarrow \mathrm{e}^{2|\alpha|l} \geq (R_2)^{-1} \tag{10.60}$$

is fulfilled.

The condition on the phases may be written in a similarly simple way. After one loop, according to (10.58), the phase of the wave may have changed only for an integer multiple of 2π. We thus have:

$$\varphi\,(\text{after one loop}) = \varphi_0 + 2j\pi \tag{10.61}$$

The phase gain is thus equal to $2j\pi$. On the other hand, in (7.15) we have already calculated the phase gain for a single loop of a light wave between two plane interfaces, which turned out to be equal to $4\pi L/\lambda$ in the case that the refractive index is equal to one. This is clearly a rough simplification, but it still allows to highlight the main principles of the laser action. Moreover, possible phase shifts occurring upon reflection at the mirrors will also not be taken into account.

We therefore find the condition:

$$2j\pi = \frac{4\pi L}{\lambda} \Rightarrow \lambda = \lambda_j = \frac{2L}{j} \tag{10.62}$$

The conditions (10.60) and (10.62) have to be fulfilled together in order to get the laser working. Let us therefore analyse their common solutions.

We will start from condition (10.60). It obviously defines a threshold value for the amplification coefficient which must be exceeded in order to get the laser work. As the amplification coefficient is wavelength-dependent (instead of the familiar absorption line, we now have an amplification line), one may expect that (10.60) defines one or several spectral ranges where the amplification coefficient is sufficiently large to achieve light generation. In order to handle this in a more convenient mathematical way, let us rewrite (10.60) in the more symbolic manner:

$$|\alpha| = |\alpha\,(\omega)| \geq \text{threshold} \tag{10.63}$$

where the generation threshold according to (10.60) is defined by the reflectivity of the second mirror and the length of the active medium. In general,

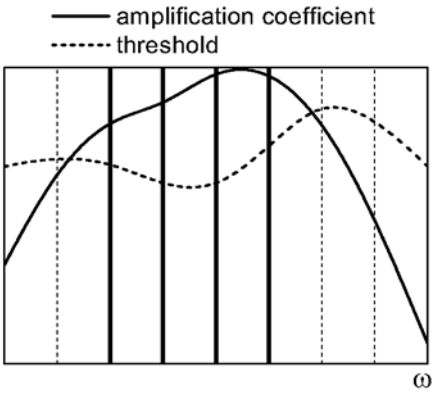

Fig. 10.8. Frequency spectrum of a laser. The *vertical lines* represent the longitudinal resonator modes. Generation may only occur when the amplification coefficient (*solid curve*) exceeds the generation threshold (*dash*). Hence, from the full set of longitudinal resonator modes only the bold ones constitute the laser spectrum

other loss mechanisms may also be present in the cavity. In this case the threshold will be enhanced, but the general formulation of criterion (10.63) remains the same. Of course, the threshold value itself may also depend on the frequency.

On the other hand, (10.62) defines a series of discrete wavelength values that are equidistant at the frequency scale as long as the refractive index is considered to be constant (in our case it is 1 regardless of the frequency). Indeed, from (10.62) the allowed frequency values may be calculated according to:

$$\omega_j = \frac{2c\pi}{\lambda_j} = j\frac{c\pi}{L} \tag{10.64}$$

where $c\pi/L$ is the line spacing in angular frequency units. Equation (10.64) defines the set of allowed so-called longitudinal resonator modes.

The light frequencies that suffice both criteria (10.63) and (10.64) are given by a set of discrete lines confined to the frequency region defined by (10.63). Figure 10.8 sketches this situation in a simplified manner.

In general, we thus have to expect that a laser generates light at a certain set of different frequencies.

We come to the conclusion, that the laser is a somewhat flexible light source that may generate light at different frequencies according to its geometrical specifics and the kind of active material. Particularly, a laser may be specifically designed to meet quite different specifications. If we are interested in highly monochromatic light, we need to design a laser which generates at only one well-defined longitudinal resonator mode. If we are, however, interested in a laser source that generates very short pulses of light, we cannot work with a single longitudinal mode. The reason is that a short light pulse

has a broad frequency spectrum (compare with the discussion in Sect. 4.1 and 4.2), and the laser must supply this broad spectrum of light. So that in practice one needs quite different types and designs of lasers depending on the concrete application.

The problem is, that the extreme cases of one single mode (high monochromaticity) and of a broad spectrum of equidistant resonator modes (necessary for a short pulse laser) are difficult to be accomplished. On his own, the laser tends to generate light at several (not necessary adjacent at the frequency scale) longitudinal modes, which may quickly change with time. The reason is, that the modes are not independent from each other, because all they are fed from the same active medium (supposed that the amplification line is not inhomogeneously broadened). Therefore, the modes compete, and it is some kind of Darwinian selection between the longitudinal modes that leads to the "survival of the fittest". In a resonator like shown in figure 10.7, usually several modes may survive. This is caused by the standing wave pattern observed in such a resonator. As seen from figure 10.3, in such a standing wave there are nodes and antinodes of the electric field strength. In the spatial region of antinodes, a strong longitudinal mode tends to transfer the atoms from the excited level to the ground state due to stimulated emission, thus destroying the population inversion. As a consequence, many other longitudinal modes cannot survive. On the other hand, the mode does not affect the population inversion in the node region, thus giving a survival chance to several modes that have their antinodes in these regions. Such a resonator construction, which leads to the formation of standing waves, is therefore not suitable for the construction of a single mode laser.

Single mode lasers are therefore usually prepared as ring lasers, where the resonator modes do not form standing waves, but propagate as travelling waves. In this case, and in the absence of any inhomogeneous amplification line broadening mechanisms, the competition between the modes may lead to the survival of only one single mode, such supplying a highly monochromatic light.

On the other hand, when one has to construct a shortpulse laser, it is necessary that a large amount of mutually adjacent longitudinal modes becomes excited. That may be done modulating the resonator properties with a frequency that exactly corresponds to the mode spacing $c\pi/L$. The modulation creates sidebands to the already excited longitudinal modes, which again become modulated, so that new sidebands are created until the process comes to an end when a broad spectrum of adjacent modes is excited.

The picture developed in this section is quite simplified, but nevertheless it should give a qualitative understanding of what is going on in a laser. There remains one question: What is the connection to thin film optics?

In fact, the resonator shown in Fig. 10.7 is completely analogue to a single thin film with a negative absorption coefficient. Equations (10.59)–(10.62) are easily obtained from the thin film equation (7.13), demanding a transmission $T \to \infty$ (finite output while no input) and setting $d = l = L$ and $n = 1$.

11 Semiclassical Treatment of the Dielectric Function

11.1 First Suggestions

In the previous chapter, we've got a first idea about the specifics of the quantum mechanical treatment of the interaction between light and matter. In particular, we found that the efficiency of that interaction is determined by at least three factors:

- The square of the absolute value of the matrix element of the perturbation (in our case the electric dipole) operator (which defines the *oscillator strength*).
- The relation between the frequency of the impinging electromagnetic wave and the energy spacing between the energy levels of the unperturbated material system which define the resonance frequencies (*resonance condition*).
- The population difference of the energy levels involved into the quantum transition.

In order to get a convenient quantum mechanical description of refraction and absorption processes in thin film materials, as necessary for any calculations in thin film optics, the task is now to find a semiclassical expression for the dielectric function. From that, refractive index and absorption coefficient would follow according to (2.18).

Qualitatively, we may already guess the correct structure of the quantum mechanical expression for the dielectric function. We may start from (4.6):

$$\beta = \frac{q^2}{\varepsilon_0 m} \sum_{j=1}^{M} \frac{f_j}{\omega_{0j}^2 - \omega^2 - 2\mathrm{i}\omega\Gamma_j} = \frac{3}{N} \frac{\hat{n}^2 - 1}{\hat{n}^2 + 2} = \frac{3}{N} \frac{\varepsilon - 1}{\varepsilon + 2}$$

This equation determines the dielectric function in the framework of the classical theory. In quantum mechanics, we will have to expect that the resonance frequencies from (4.6) are replaced by the transition frequencies (10.19), while the f-factors depend on the transition matrix elements as well as on the population difference. Therefore, starting from (4.6), the structure of the polarizability and accordingly of the dielectric function may be guessed according to:

$$\beta \propto \frac{q^2}{\varepsilon_0 m} \sum_l \sum_{n>l} \frac{|x_{nl}|^2 [W(l) - W(n)]}{\omega_{nl}^2 - \omega^2 - 2\mathrm{i}\omega\Gamma_{nl}} = \frac{3}{N} \frac{\varepsilon - 1}{\varepsilon + 2} \tag{11.1}$$

Here, it is assumed for simplicity, that the quantum states are counted in a way that $E_n > E_1$ is fulfilled. The values $W(l)$ are the statistical probabilities that the l-th energy level is populated.

To some extent, (11.1) may already be regarded as the final result of this chapter. Of course, our guess cannot be regarded as a serious derivation, so that we need to prove that (11.1) is really correct. Even more important, expression (11.1) cannot be used for absolute calculations, as long as the proportionality constant is unknown. So that it remains necessary to provide a relevant derivation of the expression for the dielectric function, and this will be done in the next subchapters. For those readers who do not want to go into these details, (11.1) might be sufficient, and the following sections may be skipped.

In finishing Sect. 11.1, let us make a few remarks concerning (11.1). The first important fact is, that in (11.1), all allowed quantum transitions may principally contribute to the polarizability as long as the participating energy levels are populated and the transition is not saturated. Particularly, in the case of population inversion we get negative contributions to the polarizability (light amplification instead of absorption).

Another remark concerns (3.17). In linear optics ((2.4)), the polarizability as defined by (3.17) should not depend on the electric field strength (otherwise (3.17) would define a nonlinear dependence of the polarization on the field strength). Therefore, (11.1) defines a linear polarizability only in the case of sufficiently low field strength values, so that the impinging light wave does not affect the population probabilities $W(l)$. As soon as the population of the quantum states becomes altered by the light intensity, we leave the field of linear optics and enter that of nonlinear optics.

11.2 Extended Detail: Calculation of the Dielectric Function by Means of the Density Matrix

11.2.1 The Interaction Picture

For those readers which are willing to proceed with reading of this chapter in order to find a complete derivation of the semiclassical expression of the linear polarizability, we will now have to perform some purely mathematical work in advance that shall make the further derivation easier and more compact. Let us therefore introduce the so-called interaction picture.

The formal purpose of this section is to find another writing of Schrödinger's equation which is more convenient for our particular purposes. As in Sect. 10.4, we write the Hamiltonian in the following manner:

$$\boldsymbol{H} = \boldsymbol{H_0} + \boldsymbol{V}$$

Where $\boldsymbol{V}$ describes the interaction between the material system and the irradiation, while $\boldsymbol{H_0}$ describes the unperturbated Hamiltonian of the material system. The solution of Schrödinger's Equation:

$$\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = \boldsymbol{H}\Psi$$

supplies us with wavefunctions Ψ that are called wavefunctions in the Schrödinger's picture. Let us now define the operators $\boldsymbol{U}_0$ and $\boldsymbol{U}_0^{-1}$ via

$$\begin{gathered}\boldsymbol{U}_0 \equiv \mathrm{e}^{-\mathrm{i}\frac{\boldsymbol{H_0}t}{\hbar}}; \quad \boldsymbol{U}_0^{-1} \equiv \mathrm{e}^{\mathrm{i}\frac{\boldsymbol{H_0}t}{\hbar}} \\ \mathrm{e}^{\mathrm{i}\frac{\boldsymbol{H_0}}{\hbar}t} \equiv 1 + \mathrm{i}\frac{\boldsymbol{H_0}}{\hbar}t + \tfrac{1}{2}\left(\mathrm{i}\frac{\boldsymbol{H_0}}{\hbar}t\right)^2 + \ldots\end{gathered} \tag{11.2}$$

The wavefunction

$$\Psi_w \equiv \boldsymbol{U}_0^{-1}\Psi \tag{11.3}$$

is (per definition) called the wavefunction in the interaction picture. We will understand later why this definition is so convenient.

Of course, in the interaction picture, the wavefunctions shall be orthonormalized as well as in Schrödinger's picture. We therefore find:

$$\int \Psi^*\Psi \,\mathrm{d}^3\boldsymbol{r} = 1 = \int \Psi_w^*\Psi_w \,\mathrm{d}^3\boldsymbol{r}$$

So that

$$\Psi_w^* = \Psi^*\boldsymbol{U}_0 \tag{11.4}$$

holds.

We already know, that matrix elements of several operators will be essential to describe the radiation-with-matter interaction. Let us regard any arbitrary operator $\boldsymbol{A}$ which acts onto the wavefunction Ψ. In the quantum state described by Ψ, the quantum mechanical expectation of $\boldsymbol{A}$ is given by:

$$\langle A\rangle = \int \Psi^*\boldsymbol{A}\Psi \,\mathrm{d}^3\boldsymbol{r} \tag{11.5}$$

That expectation corresponds to the result of a physical measurement and shall therefore be independent on the concrete quantum picture. We therefore find:

$$\langle A_w\rangle = \langle A\rangle = \int \Psi^*\boldsymbol{A}\Psi \,\mathrm{d}^3\boldsymbol{r} = \int \Psi_w^*\boldsymbol{A}_w\Psi_w \,\mathrm{d}^3\boldsymbol{r} = \int \Psi^*\boldsymbol{U}_0\boldsymbol{A}_w\boldsymbol{U}_0^{-1}\Psi \,\mathrm{d}^3\boldsymbol{r}$$

which holds when $\boldsymbol{A}$ reads in the interaction picture as:

$$\boldsymbol{A}_w = \boldsymbol{U}_0^{-1}\boldsymbol{A}\boldsymbol{U}_0 \tag{11.6}$$

Let us finally write down Schrödinger's equation itself in the interaction picture. We find:

$$\begin{aligned} \mathrm{i}\hbar\frac{\partial}{\partial t}\Psi_w &= -\boldsymbol{H_0}\boldsymbol{U}_0^{-1}\Psi + \boldsymbol{U}_0^{-1}\times\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} = \boldsymbol{U}_0^{-1}\left(-\boldsymbol{H_0}\Psi + \boldsymbol{H}\Psi\right) = \\ &= \boldsymbol{U}_0^{-1}\boldsymbol{V}\Psi = \boldsymbol{U}_0^{-1}\boldsymbol{V}\boldsymbol{U}_0\boldsymbol{U}_0^{-1}\Psi = \\ &= \boldsymbol{V}_w\Psi_w \end{aligned}$$

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\Psi_w = \boldsymbol{V}_w\Psi_w \tag{11.7}$$

This is clearly a more compact equation than in Schrödinger's picture. Particularly, the wavefunction is only time-dependent when the interaction potential $\boldsymbol{V}$ is different from zero.

11.2.2 Introduction of the Density Matrix

The general idea of calculating the dielectric function is completely analogous to what has been done in the classical theory. There we started from the calculation of microscopic dipole moments. The obtained expressions have been compared to (3.17) to find the polarizability. Finally, the dielectric function has been found from the Lorentz-Lorenz formula.

The entire difference to the quantum mechanical theory is in the method of calculating the dipole moment. In classical theory, we simply solved Newton's equation of motion for an oscillating charge with a finite mass. In quantum mechanics, the system may be characterized by a wavefunction Ψ, and the expected dipole moment $\langle p\rangle$ has to be calculated according to the recipe:

$$\langle p\rangle = \int \Psi^*\boldsymbol{p}\Psi\,\mathrm{d}^3\boldsymbol{r} \tag{11.8}$$

$\boldsymbol{p}$ is again the operator of the dipole moment. The integration must be performed over all coordinates of the system, which encounter into the wavefunction Ψ.

In fact, the situation is usually more complicated. Let us assume, that the system which is in interaction with light is nothing else than a molecule without a permanent dipole moment. Then, $\langle p\rangle$ describes the induced dipole moment of the molecule. The problem is, that the molecule itself is not only interacting with the external light, but also with its ambient. It is therefore not obvious that there exists a wavefunction that describes the molecule + radiation system in terms of only the molecular coordinates and the electromagnetic fields. Quite possible, that there exists a more general wavefunction which additionally depends on a variety of coordinates describing the properties of the surrounding medium, but such a wavefunction is not helpful for calculations by means of (11.8).

For that reason, in the semiclassical theory of the interaction of light with matter it becomes necessary to find another type of description. This is also obvious from purely thermodynamical considerations: Any highly excited molecule tends to loose its energy in order to come into thermodynamical equilibrium with its surrounding. But as the quantum states described by (10.11) are not time dependent, relaxation processes cannot be taken into account this way. As a consequence, the energy levels described by (10.11) are absolutely sharp, so that optical transitions would be expected to cause absorption or emission lines with an infinitesimally small linewidth. These facts are clearly in contradiction with reality, so that relaxation processes have to be incorporated into our description anyway.

We will not develop a general theory here, but focus on our particular task. Let us assume at the beginning, that we deal with a material system that may be described by a wavefunction Ψ depending on the systems coordinates only. Such a system is called to be in a *pure* quantum state. The mentioned wavefunction may be expanded into a series of eigenfunctions of the unperturbated Hamiltonian of the system leading to:

$$\Psi(\boldsymbol{r}, t) = \sum_n a_n(t)\, \psi_n(\boldsymbol{r}) \tag{11.9}$$

The quantum mechanical expectation of the dipole operator follows from (11.8) and may be written according to:

$$\begin{aligned} \langle A \rangle &= \int \Psi^* \boldsymbol{A} \Psi \, \mathrm{d}^3\boldsymbol{r} = \sum_n \sum_m a_n^* a_m \int \psi_n^* \boldsymbol{A} \psi_m \, \mathrm{d}^3\boldsymbol{r} = \\ &= \sum_n \sum_m a_n^* a_m A_{nm} \equiv \sum_n \sum_m \sigma_{mn} A_{nm} = \sum_n (\boldsymbol{A}\boldsymbol{\sigma})_{nn} \\ &= Tr(\boldsymbol{A}\boldsymbol{\sigma}) \end{aligned} \tag{11.10}$$

The values σ_{nm} are called elements of the density matrix of the system in the given (pure) quantum state. According to (11.10), knowledge of the density matrix again allows to calculate the quantum mechanical expectation value for an operator in a given quantum state. Particularly, from the special case $\boldsymbol{A} = 1$, we obtain:

$$1 = Tr\boldsymbol{\sigma} = \sum_n \sigma_{nn} \tag{11.11}$$

Per definition (11.10), the diagonal elements of the density matrix σ_{nn} are identical to the probability to find the system in the n-th quantum state as a result of a measurement procedure.

Let us now find the equation that describes the evolution of the elements of the density matrix with time. From Schrödinger's equation, we have:

$$
\begin{aligned}
\mathrm{i}\hbar\frac{\partial\Psi}{\partial t} &= \boldsymbol{H}\Psi = \mathrm{i}\hbar\sum_m \frac{\partial}{\partial t}a_m\psi_m = \sum_m a_m\boldsymbol{H}\psi_m \;\Big|\times\psi_n^*\,;\; \int \mathrm{d}^3\boldsymbol{r} \\
\rightarrow \mathrm{i}\hbar\frac{\partial}{\partial t}a_n &= \sum_m H_{nm}a_m \\
\mathrm{i}\hbar\frac{\partial}{\partial t}a_n^* &= -\sum_m H_{nm}^*a_m^* = -\sum_m a_m^*H_{mn} \\
\mathrm{i}\hbar\frac{\partial}{\partial t}\sigma_{mn} &= \mathrm{i}\hbar a_n^*\frac{\partial}{\partial t}a_m + \mathrm{i}\hbar a_m\frac{\partial}{\partial t}a_n^* = \sum_l\left(H_{ml}\underbrace{a_la_n^*}_{\sigma_{ln}} - H_{ln}\underbrace{a_l^*a_m}_{\sigma_{ml}}\right) \\
\rightarrow \mathrm{i}\hbar\frac{\partial}{\partial t}\sigma_{mn} &= \sum_l\{H_{ml}\sigma_{ln} - H_{ln}\sigma_{ml}\} = \{\boldsymbol{H}\boldsymbol{\sigma} - \boldsymbol{\sigma}\boldsymbol{H}\}_{mn}
\end{aligned}
$$

What we have found is the so-called Liouville's equation for the elements of the density matrix:

$$
\mathrm{i}\hbar\frac{\partial}{\partial t}\sigma_{mn} = \{\boldsymbol{H}\boldsymbol{\sigma} - \boldsymbol{\sigma}\boldsymbol{H}\}_{mn} \tag{11.12}
$$

In the operator language, (11.12) may be written as:

$$
\mathrm{i}\hbar\frac{\partial}{\partial t}\boldsymbol{\sigma} = [\boldsymbol{H},\boldsymbol{\sigma}] \tag{11.13}
$$

Let us now come to our concrete problem. We turn to the more complicated case, that our regarded system interacts with the electromagnetic field of the light wave as well as with another material system which may be regarded as the surrounding. In order to simplify the task, we will assume that the surrounding medium itself does not interact with the electromagnetic wave.

This situation is sketched in Fig. 11.1. It shows the regarded system S (for example a molecule) and its ambient (system U). The perturbation $\boldsymbol{V}$ only interacts with the system S. Additionally, system S interacts with its ambient U via the interaction Hamiltonian $\boldsymbol{H}_{SU}$. The individual systems S and U themselves are regarded to be described by the Hamiltonians $\boldsymbol{H}_{0S}$ and $\boldsymbol{H}_{0U}$. The full problem shall be described by the Hamiltonian:

$$
\boldsymbol{H} = \boldsymbol{H}_{0U} + \boldsymbol{H}_{0S} + \boldsymbol{H}_{SU} + \boldsymbol{V}
$$

where only the operator $\boldsymbol{V}$ shall explicitly depend on the time.

We will assume, that the pure states of the complete system $(S+U)$ may be described by the wavefunction $\Psi^{(j)}$ or density matrices $\boldsymbol{\sigma}^{(j)}$, while j is a quantum number that counts the quantum states of the complete system. $\boldsymbol{\sigma}^{(j)}$ suffices (11.13):

$$
\mathrm{i}\hbar\frac{\partial\boldsymbol{\sigma}^{(j)}}{\partial t} = \left[\boldsymbol{H},\boldsymbol{\sigma}^{(j)}\right] \tag{11.14}
$$

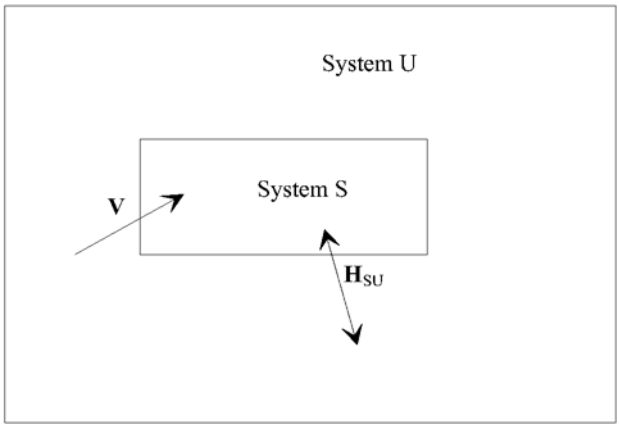

Fig. 11.1. System S interacting with the perturbation V and its ambient U

We will moreover assume, that we know the eigenfunctions of the Hamiltonian $\boldsymbol{H}_{0S}$, which are solutions of:

$$\boldsymbol{H}_{0S}\psi_n(r) = E_n\psi_n(\boldsymbol{r}) \tag{11.15}$$

Multiplying (11.14) by ψ_n^* from the left and by ψ_m from the right and integrating over all coordinates, we obtain:

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\boldsymbol{\sigma}_{nm}^{(j)} = \left[\boldsymbol{H}, \boldsymbol{\sigma}^{(j)}\right]_{nm} \tag{11.16}$$

In the general case we will not know, what is the actual quantum state of the complete system $(S+U)$. We might be able to control the behaviour of our subsystem S, but the behaviour of the environment is difficult to be handled. Nevertheless, one may find a satisfying mathematical treatment of the problem intermixing our quantum mechanical treatment with a classical averaging procedure.

This kind of treatment is particularly evident when we assume that the system $S+U$ is large enough to be considered as a macroscopic system. Such systems are successfully described by classical statistical mechanics. We will therefore assume that we may determine a particular classical probability $w^{(j)}$ to find the whole system in the quantum state j. The density matrix of system S, which is definitely not in a pure state (but in a so-called mixed state), is now defined by the expression:

$$\rho_{nm} \equiv \sum_j w^{(j)}\sigma_{nm}^{(j)} \tag{11.17}$$

Equations (11.17) and (11.16) together yield:

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\rho_{nm} = [\boldsymbol{H}, \boldsymbol{\rho}]_{nm} \tag{11.18}$$

This is again Liouville's equation, applied to the density matrix of the system S in a mixed state. By means of the operators

$$\boldsymbol{U}_0 = \mathrm{e}^{-\frac{\mathrm{i}}{\hbar}(\boldsymbol{H}_{0S}+\boldsymbol{H}_{0U})t}$$

$$\boldsymbol{U}_0^{-1} = \mathrm{e}^{\frac{\mathrm{i}}{\hbar}(\boldsymbol{H}_{0S}+\boldsymbol{H}_{0U})t}$$

Equation (11.18) may be transformed into the interaction picture. In complete analogy to the previous subsection, this procedure leads to the elimination of the Hamiltonians $\boldsymbol{H}_{0S}$ and $\boldsymbol{H}_{0U}$ from Liouville's equation. We find:

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\rho_{w_{nm}} = [\boldsymbol{V}_w, \boldsymbol{\rho}_w] + [\boldsymbol{H}_{SU_w}, \boldsymbol{\rho}_w]_{nm} \tag{11.19}$$

In the forthcoming we will assume that all operators are written in the interaction picture, so that we skip the index w for simplicity in most cases.

Let us have a closer look at (11.19). Obviously it is the first commutator that describes the interaction of system S with the electromagnetic irradiation. Hence, it should contain information on the light induced quantum transitions which determine the optical behaviour of our system as if it would be isolated. On the other hand, the second commutator describes the interaction of the system with the ambient U, which is expected to act as a thermal reservoir. Hence, it is responsible for the relaxation of the perturbated density matrix back to thermal equilibrium.

Let us come to the calculation of expectation values. From (11.10) we find:

$$\langle A\rangle = Tr\,(\boldsymbol{A}\boldsymbol{\sigma}) = \sum_n (\boldsymbol{A}\boldsymbol{\sigma})_{nn}$$

Any of the pure states j is connected to a statistical probability $w^{(j)}$. Performing a classical averaging procedure, we therefore obtain:

$$\langle A\rangle = \sum_j w^{(j)} \sum_n \left(\boldsymbol{A}\boldsymbol{\sigma}^{(j)}\right)_{nn} = \sum_n (\boldsymbol{A}\boldsymbol{\rho})_{nn} \tag{11.20}$$

Again, knowledge on the density matrix will enable us to calculate the necessary expectation values.

Finally, let us understand the sense of the diagonal elements of the density matrix. From

$$\rho_{nn} = \sum_j w^{(j)} \sigma_{nn}^{(j)}$$

it follows, that ρ_{nn} is the probability to find the system in the n-th quantum state after having performed a corresponding measurement. If the system is in thermodynamical equilibrium with its environment, that probability will be given by:

$$\rho_{nn} = \frac{\mathrm{e}^{-\frac{E_n}{k_\mathrm{B}T}}}{\sum\limits_n \mathrm{e}^{-\frac{E_n}{k_\mathrm{B}T}}} \tag{11.21}$$

If the system is in or close to equilibrium with its environment that is held at temperature T, we may use (11.21) to describe the diagonal elements of the density matrix.

In the forthcoming, it will be our task to solve (11.19) for selected systems in order to get knowledge on the density matrix. Having calculated the density matrix, the expectation value for the dipole operator p is calculated according to (11.20). After that, we obtain the polarizability according to (3.17). We may then write down the expression for the dielectric function, so that our previously formulated task to find a quantum mechanical description of the linear optical constants will be solved at this point.

11.2.3 Semiclassical Calculation of the Polarizability

We start with the simplest quantum mechanical model system, namely the two-level system previously discussed in Chap. 10. It will now be our purpose to find a semiclassical expression for the linear polarizability of such a system. According to (11.20), the expectation value for the dipole moment is given by:

$$\langle p \rangle = Tr\left(\boldsymbol{p}\boldsymbol{\rho}\right) \tag{11.22}$$

The elements of the density matrix may be obtained as the solution of Liouville's (11.19):

$$\mathrm{i}\hbar\frac{\partial}{\partial t}\rho_{nm} = \left[\boldsymbol{V}, \boldsymbol{\rho}\right]_{nm} + \left[\boldsymbol{H}_{SU}, \boldsymbol{\rho}\right]_{nm} \tag{11.23}$$

where the term $[\boldsymbol{H}_{SU}, \boldsymbol{\rho}]$ describes the interaction between the two-level system and its material surrounding. For the two-level system, the operators $\boldsymbol{p}$ and $\boldsymbol{\rho}$ may be written as:

$$\boldsymbol{p} = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}; \qquad \boldsymbol{\rho} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}$$

The diagonal elements of the dipole operator are assumed to be zero, in order to exclude permanent dipole moments in the medium. So that from (11.22) it follows:

$$Tr\left(\boldsymbol{p}\boldsymbol{\rho}\right) = \langle p \rangle = p_{12}\rho_{21} + p_{21}\rho_{12} \tag{11.24}$$

We find, that it will be necessary to calculate the non-diagonal elements of the density matrix in order to find an expression for the expectation value of the dipole moment. Moreover, that expectation value will be definitely zero

when the matrix element p_{12} becomes zero. So that the selection rule for electric dipole transitions previously derived in Chap. 10 is a natural conclusion from the more sophisticated treatment applied in the present chapter.

In order to solve (11.23), we have to find a suitable expression for the terms $[\boldsymbol{H}_{SU}, \boldsymbol{\rho}]$. In accordance to the assumptions on relaxation processes in the classical picture, we will assume an exponential damping of the free polarization of the medium. That will be consistent with the assumption:

$$[\boldsymbol{H}_{SU}, \boldsymbol{\rho}]_{nm} \equiv -\frac{\mathrm{i}\hbar\rho_{nm}}{T_{2_{(nm)}}}$$

where T_2 is the so-called transversal relaxation time. Let us mention in this connection, that a corresponding relaxation time may be introduced for the diagonal elements of the density matrix as well. It will be responsible for the relaxation of the population of the quantum levels back to equilibrium and is called the longitudinal relaxation time T_1. In relation to the classical treatment, T_2 results in the homogeneous linewidth according to (4.4), while T_1 is responsible for the natural linewidth according to (4.2).

We may now apply (11.23) to our two level case. We have to calculate two non-diagonal elements of the density matrix, and the corresponding equation become:

$$\begin{aligned} \frac{\partial}{\partial t}\rho_{21} + \frac{\rho_{21}}{T_2} &= -\frac{\mathrm{i}}{\hbar}[\boldsymbol{V}, \boldsymbol{\rho}]_{21} \\ \frac{\partial}{\partial t}\rho_{12} + \frac{\rho_{12}}{T_2} &= -\frac{\mathrm{i}}{\hbar}[\boldsymbol{V}, \boldsymbol{\rho}]_{12} \end{aligned} \tag{11.25}$$

As before, the interaction operator $\boldsymbol{V}$ is given by:

$$\boldsymbol{V} = -\boldsymbol{p}\boldsymbol{E} = -pE = -pE_0\,\mathrm{e}^{-\mathrm{i}\omega t} \tag{11.26}$$

with $\boldsymbol{E}$ as the electric field strength, optical isotropy assumed. The products $\boldsymbol{V}\boldsymbol{\rho}$ and $\boldsymbol{\rho}\boldsymbol{V}$ are then obtained as:

$$\boldsymbol{V}\boldsymbol{\rho} = -E\begin{pmatrix} p_{12}\rho_{21} & p_{12}\rho_{22} \\ p_{21}\rho_{11} & p_{21}\rho_{12} \end{pmatrix}$$

$$\boldsymbol{\rho}\boldsymbol{V} = -E\begin{pmatrix} \rho_{12}p_{21} & \rho_{11}p_{12} \\ \rho_{22}p_{21} & \rho_{21}p_{12} \end{pmatrix}$$

so that we find:

$$\begin{aligned} [\boldsymbol{V}, \boldsymbol{\rho}]_{21} &= -Ep_{21}\,(\rho_{11} - \rho_{22}) \\ [\boldsymbol{V}, \boldsymbol{\rho}]_{12} &= -Ep_{12}\,(\rho_{22} - \rho_{11}) \end{aligned}$$

Consequently, (11.25) may be rewritten as:

$$\begin{aligned}\frac{\partial}{\partial t}\rho_{21}+\frac{\rho_{21}}{T_2}&=\frac{\mathrm{i}}{\hbar}Ep_{21}\left(\rho_{11}-\rho_{22}\right)\\ \frac{\partial}{\partial t}\rho_{12}+\frac{\rho_{12}}{T_2}&=\frac{\mathrm{i}}{\hbar}Ep_{12}\left(\rho_{22}-\rho_{11}\right)\end{aligned}\tag{11.27}$$

According to (11.6), the matrix elements of the dipole operator must be time-dependent in the interaction picture. From the transformation recipe:

$$\boldsymbol{A}_w=\boldsymbol{U}_0^{-1}\boldsymbol{A}\boldsymbol{U}_0=\mathrm{e}^{\mathrm{i}\frac{\boldsymbol{H}_{0S}t}{\hbar}}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\frac{\boldsymbol{H}_{0S}t}{\hbar}}$$

it follows, that the matrix elements of an operator $\boldsymbol{A}$ in the interaction picture may be written as:

$$A_{w_{nm}}=\int\psi_n^*\mathrm{e}^{\mathrm{i}\frac{\boldsymbol{H}_{0S}t}{\hbar}}\boldsymbol{A}\mathrm{e}^{-\mathrm{i}\frac{\boldsymbol{H}_{0S}t}{\hbar}}\psi_m\,\mathrm{d}^3\boldsymbol{r}=\mathrm{e}^{\mathrm{i}\omega_{nm}t}A_{nm}\tag{11.28}$$

when ψ_n and ψ_m are eigenfunctions of $\boldsymbol{H}_{0S}$ as previously postulated. Therefore, the nondiagonal matrix elements of p that occur in (11.27) have a time-dependence according to (11.28). Moreover, the electric field is also time-dependent according to (11.26). It makes therefore sense to assume, that the non-diagonal elements of the density matrix are oscillating in time following the dependence:

$$\begin{aligned}\rho_{12}&=P_{12}\mathrm{e}^{\mathrm{i}(\omega_{12}-\omega)t}\\ \rho_{21}&=P_{21}\mathrm{e}^{\mathrm{i}(\omega_{21}-\omega)t}\end{aligned}$$

where P_{12} and P_{21} are constants. This approach is reasonable as long as the field is sufficiently weak, so that it does not alter the diagonal elements of the density matrix, which determine the population of levels 1 and 2. In other words, the diagonal elements shall not change with time. We then obtain for the density matrix:

$$\begin{aligned}\rho_{21}&=\frac{Ep_{21}}{\hbar}\frac{\rho_{11}-\rho_{22}}{\omega_{21}-\omega-\mathrm{i}\Gamma}\\ \rho_{12}&=\frac{Ep_{12}}{\hbar}\frac{\rho_{11}-\rho_{22}}{\omega_{21}+\omega+\mathrm{i}\Gamma}\\ \Gamma&\equiv T_2^{-1}\end{aligned}$$

and for the dipole moment

$$\begin{aligned}\langle p\rangle&=p_{12}\rho_{21}+p_{21}\rho_{12}\\ &=\frac{|p_{12}|^2E}{\hbar}\left(\rho_{11}-\rho_{22}\right)\left[\frac{1}{\omega_{21}-\omega-\mathrm{i}\Gamma}+\frac{1}{\omega_{21}+\omega+\mathrm{i}\Gamma}\right]\\ &=\frac{|p_{12}|^2E\cdot 2\omega_{21}\left(\rho_{11}-\rho_{22}\right)}{\hbar}\cdot\frac{1}{\omega_{21}^2+\Gamma^2-\omega^2-2\mathrm{i}\omega\Gamma}\end{aligned}\tag{11.29}$$

Caused by the assumed exponential decay of the free polarization, we find the familiar Lorentzian in the expression for the dipole moment.

We have not yet discussed the diagonal elements of the density matrix occurring in expression (11.29). We only postulated that they are time-independent. On the other hand, as we are searching for an expression for the linear polarizability, we shall require that (11.29) is linear in the electric field strength. Consequently, the diagonal elements of the density matrix shall not depend on the electric field applied. It is therefore reasonable to assume that they are equal to the equilibrium values which would be valid when no electromagnetic field is applied and the system is in equilibrium with its environment. We denote these values by a superscript '(0)', they may be calculated in terms of (11.21).

From the material equation:

$$p = \varepsilon_0 \beta E$$

we find:

$$\beta = \frac{|p_{12}|^2}{\varepsilon_0 \hbar} \cdot 2\omega_{21} \frac{\left(\rho_{11}^{(0)} - \rho_{22}^{(0)}\right)}{\omega_{21}^2 + \Gamma^2 - \omega^2 - 2\mathrm{i}\omega\Gamma} \tag{11.30}$$

where $\left(\rho_{11}^{(0)} - \rho_{22}^{(0)}\right)$ denotes a field-independent population difference. Note the similarity between (11.30) and the previously guessed expression (11.1). If more than two energy levels have to be considered, (11.30) may be generalized according to:

$$\beta = \sum_l \sum_{n>l} \frac{|p_{nl}|^2}{\varepsilon_0 \hbar} \times 2\omega_{nl} \times \frac{\left(\rho_{ll}^{(0)} - \rho_{nn}^{(0)}\right)}{\omega_{nl}^2 + \Gamma_{nl}^2 - \omega^2 - 2\mathrm{i}\omega\Gamma_{nl}} \tag{11.31}$$

Equation (11.31) gives the general semiclassical expression for the polarizability of a quantum system with discrete energy levels. Given the polarizability, the dielectric function and the optical constants follow from (3.25).

In Chap. 5, problem 10, we obtained a temperature dependent refractive index of an optical material with voids, that may partially be filled with water depending on the temperature. That was a rather extrinsic temperature effect. On the other hand, (11.31) in combination with (11.21) describes a rather intrinsic temperature dependence of an optical material, because the population of the individual energy levels will be influenced by the temperature. Moreover, the linewidth values in (11.31) are also temperature dependent, usually the lines become broader with increasing temperature. Hence, in a real optical thin film material, there may be several physical mechanisms that alter the optical constants with temperature, and the question whether the refractive index increases or decreases with temperature will depend on which of the mechanisms is the dominant one.

12 Solid State Optics

12.1 Formal Treatment of the Dielectric Function of Crystals (Direct Transitions)

It cannot be the purpose of the present chapter to provide the reader with the complete theory of electronic properties of solids, and to derive the theory of optical transitions in solids from there. For that, the reader is referred to the textbooks on solid state physics. Instead, we will assume that the reader is familiar with the general ideas of solid state physics. Particularly, basic knowledge on the band structure of crystalline solids is assumed as well as on phonons, or excitons.

In the first section, we will try to apply the treatment from the previous chapter to the specifics of optical constants of solids. We will find an expression for the dielectric function that appears to be a special case of the general expressions found in Chap. 11. From there we have (11.31):

$$\beta = \frac{2}{\varepsilon_0 \hbar} \sum_l \sum_{n>l} |p_{nl}|^2 \omega_{nl} \frac{[\rho_{ll}^{(0)} - \rho_{nn}^{(0)}]}{\omega_{nl}^2 + \Gamma_{nl}^2 - \omega^2 - 2\mathrm{i}\omega\Gamma_{nl}}.$$

Let us start with the case of a crystal. As it is known from solid state physics, a single electron moving in a periodic potential (single electron approximation) has a continuous spectrum of energy eigenvalues instead of the discrete energy levels discussed so far. Moreover, the electron energy is a continuous function of the wavevector of the electron $\boldsymbol{k}$. This is a direct consequence of the translational symmetry in the atomic arrangement, as it is per definition characteristic for a crystal. In crystal physics, one therefore speaks on *energy bands* instead of energy levels. The general theoretical considerations given below are applicable to different kinds of crystals, no matter whether they represent insulators, semiconductors, or metals. Nevertheless, we will often use a terminology which is typically applied in the field of semiconductors. The reason is simple. The electronic transitions which are detected in optical spectroscopy do usually occur between the valence and the conduction bands of a crystal. Good insulators have usually a broad energy spacing between these bands, so that they may be regarded as transparent in the NIR/VIS regions. On the contrary, the absorption on-set wavelength in semiconductors is considerably larger, so that the shape of the absorption bands has

to be taken into account when performing optical spectroscopy with semiconductors. Therefore, in many cases we will apply a terminology which is usually relevant in semiconductor optics with regard to transitions between the valence and the conduction bands.

Caused by the mentioned band structure, we cannot further work with discrete energy levels E_n as assumed in the Chaps. 10 and 11, but have to replace them by the functions $E_n(\boldsymbol{k}_n)$:

$$E_n \to E_n\left(\boldsymbol{k}_n\right).$$

The quantum number n is now to count the energy bands instead of the energy levels. Accordingly, the transition frequencies (10.19) have to be replaced according to:

$$\omega_{nl} = \frac{E_n - E_l}{\hbar} \to \frac{E_n\left(\boldsymbol{k}_n\right) - E_l\left(\boldsymbol{k}_l\right)}{\hbar}$$

Thus, absorption of light may cause an electron from the l-th energy band and an initial wavevector $\boldsymbol{k}_l$ to perform a quantum transition into the n-th energy band. In general, its wavevector may also change to $\boldsymbol{k}_n$ due to quasimomentum conservation. As long as $l \neq n$ holds, such transitions are called *interband transitions*. If $l = n$, we have an *intraband transition*, because the initial and final quantum states belong to the same energy band.

In the case of so-called *direct* transitions (no phonon creation or annihilation), we have

$$\boldsymbol{k}_l \approx \boldsymbol{k}_n \equiv \boldsymbol{k} \Rightarrow \frac{E_n\left(\boldsymbol{k}_n\right) - E_l\left(\boldsymbol{k}_l\right)}{\hbar} = \frac{E_n\left(\boldsymbol{k}\right) - E_l\left(\boldsymbol{k}\right)}{\hbar} \equiv \omega_{nl}\left(\boldsymbol{k}\right).$$

The reason is, that in optics, the light wavelength is much larger than the period of the crystalline lattice, so that the light wavevector is negligible when comparing with the dimension of the Brillouin zone. Direct transitions are vizualized in a zone diagram as vertical arrows, see Fig. 12.1. Please note that direct transitions may only occur as interband transitions. On the contrary, any intraband transition must be indirect ($\boldsymbol{k}_n \neq \boldsymbol{k}_l$), as it may easily be guessed from Fig, 12.1. In more general, the quasimomentum does not need to be conserved in the strong sense of a momentum, but may change for any integer multiple of the reciprocal lattice vector. The corresponding transitions are called Umklapp processes, but are not discussed in our treatment here.

Similar to the transition frequencies, the other values encountering into (11.31) also depend on the electron wavevector, although the dependence may be weak. Let us further recall, that the electron wavefunctions in a periodic potential are delocalized. Then, the moving electron may rather feel the medium electric field than the local one (compare Table 3.2). In such cases, a treatment in terms of the Lorentz-Lorenz formula does not make sense. We therefore assume:

$$\varepsilon = 1 + N\beta.$$

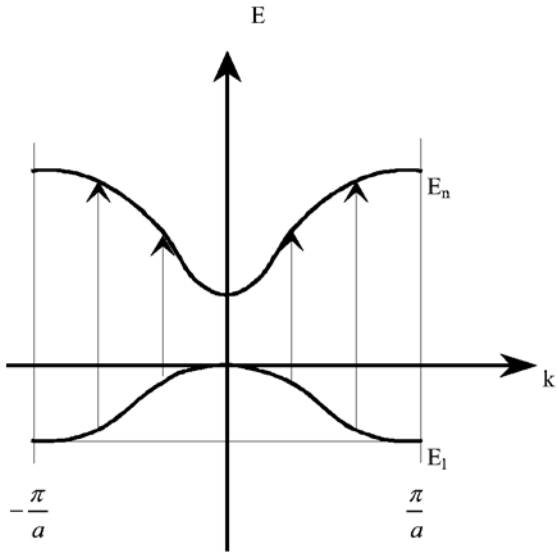

Fig. 12.1. Direct transitions between the l-th and the n-th energy band (*interband transitions*), a is the lattice period

However, due to Pauli's principle, any quantum state can only be occupied by a single electron. Summarizing over all occupied quantum states does therefore automatically sum up over the electrons. We find:

$$\begin{aligned}\varepsilon(\omega) &= 1 + \frac{2}{\varepsilon_0 \hbar} \sum_{\boldsymbol{k}} \sum_{l} \sum_{n>l} \frac{[\rho_{ll}^{(0)}(\boldsymbol{k}) - \rho_{nn}^{(0)}(\boldsymbol{k})]|p_{nl}(\boldsymbol{k})|^2 \omega_{nl}(\boldsymbol{k})}{\omega_{nl}^2(\boldsymbol{k}) + \Gamma_{nl}^2(\boldsymbol{k}) - \omega^2 - 2\mathrm{i}\omega\Gamma_{nl}(\boldsymbol{k})} \\ &= 1 + \frac{1}{4\pi^3 \varepsilon_0 \hbar} \int \mathrm{d}^3\boldsymbol{k} \sum_{l} \sum_{n>l} \frac{[\rho_{ll}^{(0)}(\boldsymbol{k}) - \rho_{nn}^{(0)}(\boldsymbol{k})]|p_{nl}(\boldsymbol{k})|^2 \omega_{nl}(\boldsymbol{k})}{\omega_{nl}^2(\boldsymbol{k}) + \Gamma_{nl}^2(\boldsymbol{k}) - \omega^2 - 2\mathrm{i}\omega\Gamma_{nl}(\boldsymbol{k})}\end{aligned} \tag{12.1}$$

In deriving (12.1), we made use of the transformation:

$$\sum_{\boldsymbol{k}} \to \frac{1}{(2\pi)^3} \int \mathrm{d}^3\boldsymbol{k}\,.$$

Equation (12.1) represents the expression for the electronic contribution to the dielectric function of a crystal, as long as only direct transitions are involved. Moreover, in the present form it is only valid in a single electron picture, without any effects caused by Coulomb's interaction between the electrons. In solids with well-filled electronic bands, of course, Fermi-Dirac statistics have to be applied for the calculation of the diagonal elements of the density matrix rather than Boltzmann's statistics.

In solid state physics, it is common to regard the transition matrix element of the electron momentum rather than of the dipole moment (as we do). In this case, (12.1) holds as well, but there is an additional pre-factor of $(e/m\omega)^2$.

Let us look at a few examples in order to get an idea on the shape of the dielectric function as described by (12.1). Let us regard the contribution of an interband transition between the l-th and the n-th band to the dielectric function. For simplicity, we will assume that the l-th band is nearly completely filled ($\rho_{ll}^{(0)} \approx 1\ \forall \boldsymbol{k}$; this might be the valence band of a semiconductor), while the n-th band is essentially empty ($\rho_{nn}^{(0)} = 0\ \forall \boldsymbol{k}$; this might be the conduction band). In a band scheme like sketched in Fig. 12.1 (and assuming isotropy), the resonance frequency might then be given by:

$$\omega_{nl}(\boldsymbol{k}) = \frac{1}{\hbar}\left[E_g + \frac{B}{2}\left(1 - \cos ka\right)\right] \tag{12.2}$$

where E_g marks the direct band gap, and B is a constant that characterizes the band width. Let us further neglect the $\boldsymbol{k}$-dependence of all the other values encountering into (12.1). Then, the contribution of the $l \to n$ transition to the dielectric function of the system may be directly calculated performing a numerical integration in (12.1). It is interesting to perform this calculation with different assumed values of the bandwidth B and the homogeneous linewidth Γ. Two examples are given in Fig. 12.2. In the case that $B \ll \Gamma$, the band structure does not give any effect, and the imaginary part of the dielectric function appears as a typical Lorentz-line as known from the classical picture or the quantum mechanical treatment of systems with discrete energy levels. On the contrary, when the homogeneous linewidth is negligible compared to the bandwidth B, the imaginary part of the dielectric function shows a sharp on-set at $\hbar\omega = E_g$ (the so-called absorption edge). Consequently, optical measurements may be used to determine the direct band gap in crystalline solids by means of the absorption behaviour. For $\hbar\omega > E_g$, the imaginary part of the dielectric function increases like

$$\mathrm{Im}\varepsilon \propto \sqrt{\hbar\omega - E_g}\,,$$

a behaviour that is typical for allowed electronic transitions in the vicinity of a direct band gap. In Sect. 12.2, this behaviour will be reproduced and explained in a less formal manner.

On the other hand, it is interesting to check the behaviour of the dielectric function depending on the dimensionality of the system. Figure 12.2 clearly corresponds to the three-dimensional case (3D), and has been calculated in spherical coordinates according to:

$$\text{3D}: \quad \mathrm{d}^3\boldsymbol{k} = \mathrm{d}k_x \mathrm{d}k_y \mathrm{d}k_z \to 4\pi k^2 \mathrm{d}k$$

The same calculation may be carried out for the 2D and 1D cases. We obtain:

$$\begin{aligned} \text{2D}: &\quad \mathrm{d}^2\boldsymbol{k} = \mathrm{d}k_x \mathrm{d}k_y \to 2\pi k \mathrm{d}k \\ \text{1D}: &\quad \mathrm{d}\boldsymbol{k} = \mathrm{d}k_x \to \mathrm{d}k \end{aligned}$$

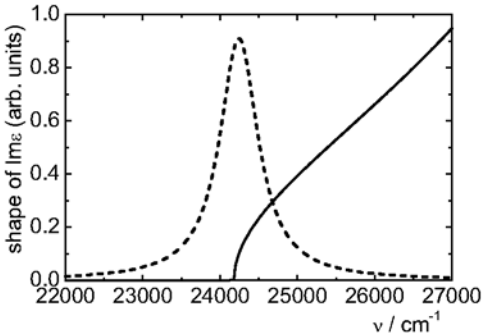

Fig. 12.2. Shape of the imaginary part of the dielectric function as calculated from (12.1) and (12.2), assuming $E_g = 3\,\mathrm{eV}$: *solid*: $B = 1\,\mathrm{eV}$, $\Gamma = 1\,\mathrm{cm}^{-1}$; *dash*: $B = 0.01\,\mathrm{eV}$, $\Gamma = 300\,\mathrm{cm}^{-1}$

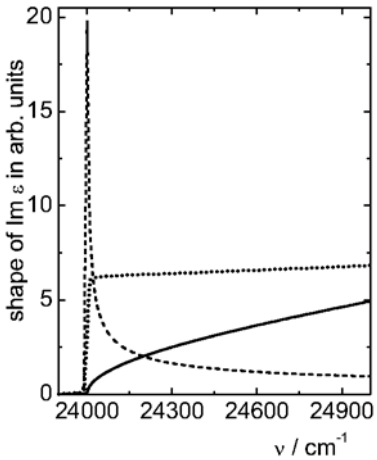

Fig. 12.3. Shape of the imaginary part of the dielectric function in the vicinity of the absorption edge, as calculated from (12.1) and (12.2). *Solid*: 3D; *dot*: 2D; *dash*: 1D. $E_g = 3\mathrm{e}\,\mathrm{V}$; $B = 1\,\mathrm{eV}$; $\Gamma = 1\,\mathrm{cm}^{-1}$

Figure 12.3 illustrates the shape of the imaginary part of the thus given dielectric function in the vicinity of the absorption edge.

It is obvious, that the dependence

$$\mathrm{Im}\varepsilon \propto \sqrt{\hbar\omega - E_g}$$

is only valid for the three-dimensional case. In the two-dimensional case, we rather find $\mathrm{Im}\varepsilon \propto$ const., while in the the one-dimensional case

$$\mathrm{Im}\varepsilon \propto \frac{1}{\sqrt{\hbar\omega - E_g}}$$

holds.

The mentioned behaviour of the dielectric function for a two-dimensional motion of an electron in a periodic potential is essential for the theory of so-called quantum well structures or superlattices. The one-dimensional case is practically relevant in so-called quantum wires. Particularly, the singularity at the direct gap is of practical importance to achieve a high oscillator strength necessary for luminescent devices.

12.2 Joint Density of States

Let us now come to a more qualitative understanding of the physics behind (12.1). It will be the purpose of this section to provide the reader with a simple and transparent derivation of the spectral shapes demonstrated in the Figs. 12.2 and 12.3. On a rather intuitive level, it is clear that the imaginary part of the dielectric function should be proportional to the square of the transition matrix element of the dipole operator, multiplied with the density of quantum states D which contribute to the transition at the given transition frequency. We write:

$$\mathrm{Im}\varepsilon \propto D\left(\omega_{nl}\right)\left|p_{nl}\right|^2 \tag{12.3}$$

Let us again concentrate on the case of direct transitions, so that the electron wavevector does not change as a result of the quantum transition. Then, ω_{nl} is given by the energy spacing between two bands at the same wavevector, and we obtain:

$$\mathrm{Im}\varepsilon \propto D[E_n\left(\boldsymbol{k}\right) - E_l\left(\boldsymbol{k}\right)]\left|p_{nl}\left(\boldsymbol{k}\right)\right|^2$$

As in Sect. 12.1, let us at the beginning assume that the transition matrix element is constant and different from zero. In this case, the behaviour of the dielectric function is determined by the *density of pairs of quantum states, which have the same wavevector and are separated from each other by a given suitable energy spacing*. We will call this density of states a *joint density of states*, because it depends on features of both energy bands that are participating in the quantum transition.

Let us have a closer look at the argument of the not yet quantitatively defined value D. For a band structure like shown in Fig. 12.1, the argument $E_n(\boldsymbol{k}) - E_l(\boldsymbol{k})$ looks like sketched in Fig. 12.4. In semiconductor physics, the minimal value of $E_n(\boldsymbol{k}) - E_l(\boldsymbol{k})$ is called the *direct gap* of the semiconductor, when the two mentioned bands are associated with the valence and conductive bands, respectively.

Let us imagine, that a system characterized by a $E_n(\boldsymbol{k}) - E_l(\boldsymbol{k})$ behaviour as given by Fig. 12.4 is illuminated with light of a sufficiently low frequency so that $\hbar\omega < E_g$ is fulfilled. There will clearly be no absorption of light. A rather sharp absorption onset is expected at $\hbar\omega = E_g$, which corresponds to

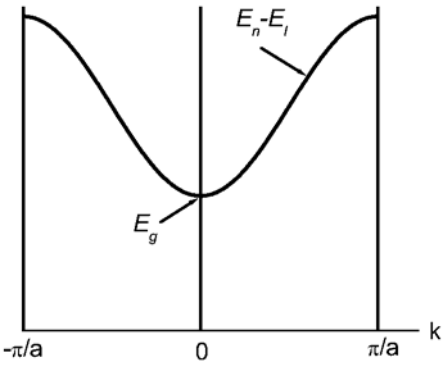

Fig. 12.4. Energy spacing between the bands from Fig. 12.1 as a function of the electron wavevector

transitions in the centre of the Brillouin zone ($k = 0$). The reason is, that at $k = 0$, the derivative

$$\frac{\mathrm{d}[E_n(\boldsymbol{k}) - E_l(\boldsymbol{k})]}{\mathrm{d}\boldsymbol{k}} = \mathbf{0}$$

as well. Consequently, a large amount of pairs of quantum states becomes involved into the optical transition, which usually leads to sharp features in the optical absorption spectrum. The same is valid for the k-values $\pm\pi/a$.

We come to the conclusion, that the main features in the imaginary part of the dielectric function are determined by the behaviour of the derivative

$$\frac{\mathrm{d}[E_n(\boldsymbol{k}) - E_l(\boldsymbol{k})]}{\mathrm{d}\boldsymbol{k}}.$$

Particularly, the points where this derivative is equal to zero, are called van-Hove-singularities.

Let us now derive a quantitative expression for the joint density of states. The number of quantum states in a given k-interval is given by:

$$\mathrm{d}Z = \frac{2V}{(2\pi)^3}\mathrm{d}k_x\mathrm{d}k_y\mathrm{d}k_z$$

The factor 2 has been introduced to account for the degeneracy of quantum states with respect to the electron spin. In spherical coordinates (which makes sense in optically isotropic materials) we find:

$$\begin{aligned} \text{3D}: \quad \mathrm{d}k_x\mathrm{d}k_y\mathrm{d}k_z = 4\pi k^2\mathrm{d}k \;\Rightarrow\; \mathrm{d}Z &= \frac{8\pi k^2 V}{(2\pi)^3}\mathrm{d}k \\ &= \frac{Vk^2}{\pi^2 \dfrac{\mathrm{d}\left[E_n(k) - E_l(k)\right]}{\mathrm{d}k}}\,\mathrm{d}[E_n(k) - E_l(k)] \end{aligned} \tag{12.4}$$

The density of states $D(k)$ is then given by:

$$\mathrm{d}Z \equiv D(k)\mathrm{d}k \Rightarrow D(k) = \frac{Vk^2}{\pi^2}$$

In full analogy, the joint density of states $D[E_n(k) - E_l(k)]$ will be defined as:

$$\mathrm{d}Z = D\big[E_n(k) - E_l(k)\big]\,\mathrm{d}\big[E_n(k) - E_l(k)\big] \tag{12.5}$$

Comparing finally (12.4) and (12.5), we find the expression:

$$D[E_n(k) - E_l(k)] = \frac{Vk^2}{\pi^2\,\dfrac{\mathrm{d}\big[E_n(k) - E_l(k)\big]}{\mathrm{d}k}} \tag{12.6}$$

Expression (12.6) is obviously only valid in the three-dimensional case. It really shows a singular behaviour at the *van-Hove* singularities. In general, it is determined by the particular band structure valid for the material under consideration.

Let us now regard the case of Fig. 12.4. At $k \to 0$, we obviously have:

$$E_n(k) - E_l(k) = E_g + \text{const.} \times k^2 = \hbar\omega$$

Consequently,

$$k \propto \sqrt{\hbar\omega - E_g}$$

and

$$\frac{\mathrm{d}\big[E_n(k) - E_l(k)\big]}{\mathrm{d}k} \propto k \propto \sqrt{\hbar\omega - E_g}$$

We obtain from (12.6):

$$D\big[E_n(k) - E_l(k)\big] \propto \sqrt{\hbar\omega - E_g};\ \ \hbar\omega > E_g \tag{12.7a}$$

which is valid in the three-dimensional case for light frequencies slightly above the absorption edge. According to (12.3), we have therefore to expect that the shape of the imaginary part of the dielectric function resembles the square root of $\hbar\omega - E_g$, which explains the behaviour of the solid lines in Figs. 12.2 and 12.3.

The same type of discussion may be performed for the 2D and 1D cases. That may be easily done by the reader himself. We find ($\hbar\omega > E_g$):

$$\begin{aligned}
&\text{3D: } \mathrm{d}^3\boldsymbol{k} = \mathrm{d}k_x\mathrm{d}k_y\mathrm{d}k_z \to 4\pi k^2\mathrm{d}k \Rightarrow D\big[E_n(k) - E_l(k)\big] \propto \sqrt{\hbar\omega - E_g}\\
&\text{2D: } \mathrm{d}^2\boldsymbol{k} = \mathrm{d}k_x\mathrm{d}k_y \to 2\pi k\mathrm{d}k \qquad \Rightarrow D\big[E_n(k) - E_l(k)\big] \propto \text{const.}\\
&\text{1D: } \mathrm{d}\boldsymbol{k} = \mathrm{d}k_x \to \mathrm{d}k \qquad\qquad\ \ \Rightarrow D\big[E_n(k) - E_l(k)\big] \propto \frac{1}{\sqrt{\hbar\omega - E_g}}
\end{aligned} \tag{12.7b}$$

We find, that the different curves in Fig, 12.3 resemble nothing else than the shape of the joint density of states in the relevant dimensionality.

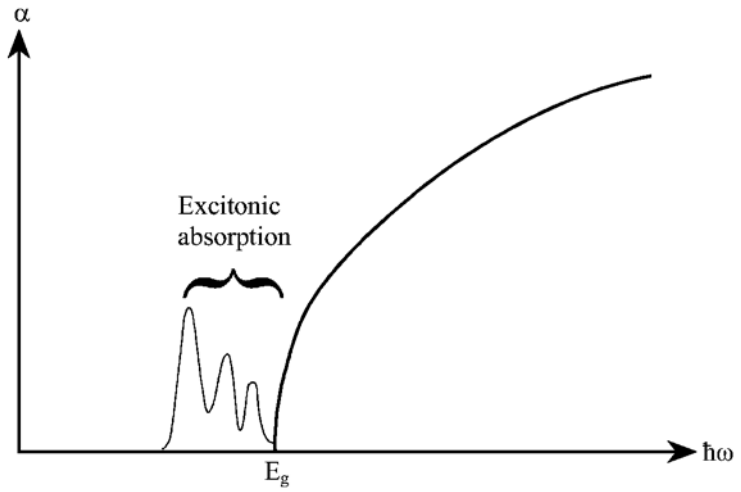

Fig. 12.5. Excitonic absorption in the region of the absorption edge of a direct semiconductor

As it has been mentioned at the beginning of this section, we assume that the transition matrix element as fixed in (12.3) is different from zero and does not strongly depend on the value of the electron wave vector k. This is true for the so-called *allowed* electronic transition.

We will now regard a somewhat different case with the confusing name of *forbidden* electronic transitions. In crystal optics, that means that the transition is forbidden in the centre of Brillouin's zone (at $k = 0$), but becomes allowed for k-values different from zero. Quite formally, the transition matrix element may be expanded into a power series according to:

$$p_{nl}(\boldsymbol{k}) = p_{nl}(0) + \frac{\partial p_{nl}(0)}{\partial k} k + \ldots$$

In the case of a forbidden transition, $p_{nl}(0) = 0$, and for $k \to 0$ it follows, that

$$p_{nl}(\boldsymbol{k}) \propto k.$$

In this case, (12.3) leads to a dependence:

$$\mathrm{Im}\varepsilon \propto D\big[E_n(\boldsymbol{k}) - E_l(\boldsymbol{k})\big] k^2 \propto \sqrt{\hbar\omega - E_g}^{\,3}; \quad \hbar\omega > E_g \qquad (12.7\mathrm{c})$$

which is again valid in the three-dimensional case.

So far, our discussion has only concerned the optical response of a single electron, moving in a periodic potential. We will not deal with a many-electron theory which allows to consider the effects caused by the Coulomb-interaction between the electrons. But our knowledge obtained so far is sufficient to account for one additional effect which is most important in semiconductor optics: Imagine the situation sketched in Fig. 12.1. An electron that is

excited from the l-th (the valence) band to the n-th (the conduction) band is well-known to leave a hole in the valence band. In their respective bands, both the created conduction electron and the hole are expected to move with a group velocity determined by the first derivative of the band energy with respect to the wavevector. In the general case, these velocities are different, so that the electron and the hole are immediately separated from each other. However, at the band edges, the group velocities are identical, so that the electron and the hole remain spatially close to each other and form a new quasiparticle, a *Wannier–Mott-exciton*. Similar to a hydrogen atom, such an exciton has Rydberg-like energy levels, which contribute to the optical absorption behaviour of the semiconductor. As a consequence, there appear sharp absorption lines in the region of the absorption edge, corresponding to the excitation of different excitonic energy levels, as exemplified in Fig. 12.5.

12.3 Indirect Transitions

So far, we have only considered direct transitions. In semiconductor practice, it appears that many of the semiconductors belong to the class of indirect semiconductors. In an indirect semiconductor, indirect interband transitions between the valence and the conduction bands may occur at photon energies which are lower than the direct gap defined in the previous section. In other words, a semiconductor is indirect, when the condition:

$$\begin{aligned} \text{indirect gap } E_{g,ind} &\equiv \min \left[E_n \left(\boldsymbol{k}_n \right) - E_l \left(\boldsymbol{k}_l \right) \right] |_{\boldsymbol{k}_n \neq \boldsymbol{k}_l} \\ &< \min \left[E_n \left(\boldsymbol{k}_n \right) - E_l \left(\boldsymbol{k}_l \right) \right] |_{\boldsymbol{k}_n = \boldsymbol{k}_l} \\ &\equiv \text{direct gap } E_g \end{aligned}$$

is fulfilled. Such a situation is shown in Fig. 12.6.

Let us now look how the absorption shape at an indirect gap looks like.

The main difference to the previously discussed case of direct transitions is in the violation of electron quasimomentum conservation. Indeed, when the absorption of light is accompanied by the generation or annihilation of one or several phonons, the electron wavevectors in the initial and final states may significantly differ from each other. Neglecting the light wavevector, the quasimomentum conservation yields:

$$\boldsymbol{k}_n - \boldsymbol{k}_l \approx \pm \sum \boldsymbol{k}_{\text{phonon}} \quad (\pm \text{ reciprocal lattice vector})$$

Additionally, energy conservation leads to:

$$E_n - E_l = \hbar\omega \pm \sum E_{\text{phonon}}$$

Here, the sign "+" corresponds to phonon annihilation, while "−" denotes phonon creation. Due to the violation of electron quasimomentum conservation, the joint density of states is no more significant for the quantitative

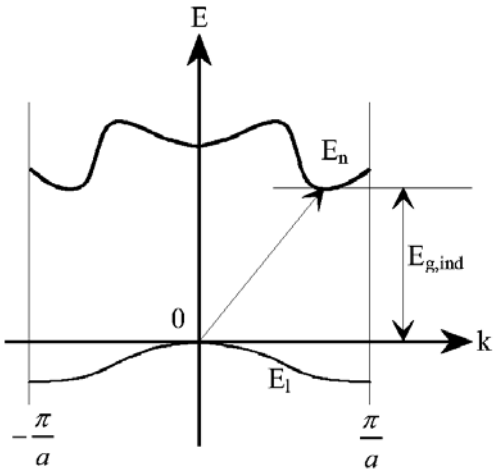

Fig. 12.6. Band structure of an indirect semiconductor

description of the absorption process. Instead, it is reasonable to consider the convolution of the densities of the initial and final quantum states, regardless on the quasimomentum. Hence, instead of (12.3), we make use of:

$$\mathrm{Im}\varepsilon \propto |p_{nl}|^2 \int_{-\infty}^{\infty} D_l(E) D_n\left(E + \hbar\omega \pm \sum E_{\mathrm{phonon}}\right) \mathrm{d}E \tag{12.8}$$

where D is the usual density of states in the corresponding band as indicated by index in (12.8). Again, near the extremal values of the $E(k)$ dependence as shown in Fig. 12.6, the energy behaves proportional to the square of the wavevector. In analogy to the treatment in the previous section, we suppose:

$$\begin{aligned}
\mathrm{d}Z &= D(k)\,\mathrm{d}k \\
&= \frac{D(k)}{\dfrac{\mathrm{d}E}{\mathrm{d}k}}\,\mathrm{d}E \equiv D(E)\,\mathrm{d}E \Rightarrow D(E) \;\propto k(E) \\
&\qquad\qquad \Rightarrow D_l(E) \propto \sqrt{-E}\,; \quad E < 0 \\
&\qquad\qquad\quad D_n(E) \propto \sqrt{E - E_{g,ind}}\,; \quad E > E_{g,ind}
\end{aligned}$$

Then, from (12.8) we obtain:

$$\mathrm{Im}\varepsilon \propto \int_{0}^{E_{g,ind} - \hbar\omega \mp \sum E_{\mathrm{phonon}}} \sqrt{-E}\sqrt{E - E_{g,ind} + \hbar\omega \pm \sum E_{\mathrm{phonon}}}\,\mathrm{d}E\,;$$

$$\hbar\omega > E_{g,ind} \mp \sum E_{\mathrm{phonon}}$$

We do not need to calculate this integral exactly. The only thing we need to know is the frequency dependence of the dielectric function. Performing the substitution:

$$-z = -E_{g,ind} + \hbar\omega \pm \sum E_{\text{phonon}}$$

we find

$$\text{Im}\varepsilon \propto \int_0^z \sqrt{Ez - E^2}\,\mathrm{d}E$$

The integrand itself represents half a circle with the diameter z, centred at $z/2$ on the abszissa. Hence it includes an area that is proportional to z^2. Consequently, the integral itself is proportional to z^2, and we find for the dielectric function:

$$\text{Im}\varepsilon\,(\omega) \propto \left(\hbar\omega - E_{g,ind} \pm \sum E_{\text{phonon}}\right)^2 ;\ \hbar\omega > E_{g,ind} \mp \sum E_{\text{phonon}} \quad (12.9)$$

We see, that (12.9) is different from the expressions (12.7a) and (12.7c), valid for the direct transitions.

In the next subsection, we will turn to another important class of solids, namely amorphous solids. Let us therefore shortly resume what we have learned about crystalline solids and their optics so far.

The main point is, that instead of atomic or molecular energy levels, in solid state physics we deal with energy bands. In crystal physics, these energy

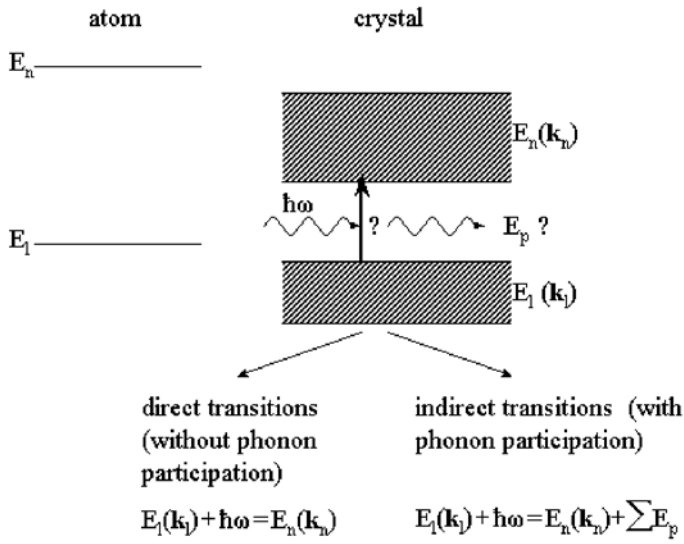

Fig. 12.7. Optical transitions in a crystal

bands are described by a $E(\boldsymbol{k})$ dependence. As in molecules, an electronic excitation may be accompanied by excitation of vibrational degrees of freedom, which gives rise to the division of optical transitions in a crystal into direct and indirect transitions. Both types of transitions differ from each other in their energy balance and the shape of the absorption structure near the absorption on-set. Figure 12.7 is to summarize these considerations.

12.4 Amorphous Solids

12.4.1 General Considerations

Let us now come to another kind of solids, namely amorphous solids. Generally speaking, amorphous solids lack long-range order in the atomic arrangement (which is characteristic for crystals), while short-range order is present. Optical glasses are prominent examples for the application of amorphous solids as optical materials.

It should be pointed out that amorphous solids shall not be confused with completely disordered matter. A snapshot of the atomic positions in a dilute gas will yield a picture that entirely corresponds to a disordered system, and clearly lacks even short-range order. On the other hand, in an amorphous solid, the mentioned short-range order is of great importance for its electrical and optical properties. Richard Zallen mentioned a very simple and helpful thought experiment to distinguish between an amorphous solid and a disordered system: Imagine a man with a bad memory (surely not a reader of this book), who removes exactly one atom from an amorphous structure and from a disordered system. Some days later he wants to reinsert the atoms into their correct positions. Clearly, he has forgotten from where the atoms have been taken. But no doubt, a glance on the positions of the remaining atoms in the amorphous structure will enable him to identify the former neighbours of the removed atom, so that he will reinsert the atom approximately at the right place. In a disordered system, however, the remaining atomic positions will give no clue about the missing one, and he will not be able to identify the former position.

In practice, amorphous solids may be identified from the radial distribution function (RDF) of their atoms, as experimentally determined for example by electron diffraction pattern. In a real crystal at finite temperature, the RDF shows well-defined peaks out to about a dozen of coordination shells. In a disordered system, peaks are generally absent, but the RDF shows a smooth parabolic increase with interatomic distance. In an amorphous solid, the RDF shows a few peaks, corresponding to the first, second, and maybe third neighbour distances, before it merges to the structureless behaviour of a disordered system like a dilute gas. An amorphous solid therefore resembles some of the properties of its crystalline counterpart (namely those which are determined by the short range order), while the properties basing on the long range order will not be found in amorphous solids.

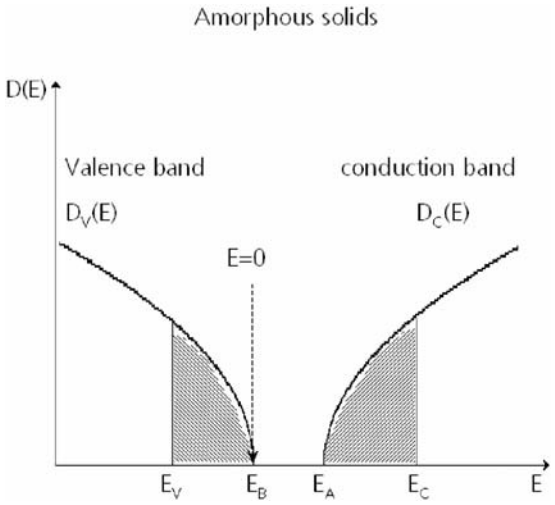

Fig. 12.8. Possible shape of the density of states in an amorphous semiconductor

These general considerations may serve here as an introduction into the specifics of the optical properties of amorphous solids. For the same reasons as in the previous chapters, we will concentrate on amorphous semiconductors, as important for example for solar cells.

In an amorphous solid, the interatomic distances are comparable to those in a crystal. Therefore, the spatial overlap of the atomic electronic wavefunctions gives rise to the formation of broad energy regions with allowed electron energy values, similar as in a crystal. On the other hand, the absence of translational invariance in the atomic arrangement does not allow one to use Bloch's theorem for the description of the electronic wavefunctions. This has several consequences:

- Although there are broad regions of allowed electron energy values, there is no $E(\boldsymbol{k})$-dependence as in crystals. Nevertheless it is common to speak of energy bands in amorphous semiconductor theory.
- Despite the delocalised electronic states characteristic for a periodic potential, there may be localized electronic quantum states as well (Anderson localization). They may deeply extend into the forbidden zone (not shown in Fig. 12.8).
- There is no quasimomentum conservation in optical transitions.
- There is no joint density of states.

Nevertheless, we may introduce a conventional density of states defining:

$$\mathrm{d}Z \equiv D\,(E)\,\mathrm{d}E$$

where dZ again is again the full number of quantum states in the given E-interval. In amorphous semiconductor physics, there exist several models

to describe the density of states in the region of the valence and conduction bands. Figure 12.8 shows an example for the density of states in the vicinity of the energy gap between the valence and conduction bands. It is characteristic, that in the vicinity of the band edges, the electronic states are spatially localized (grey regions in Fig. 12.8). An electron in such a quantum state has only a small mobility, so that E_C and E_V are called mobility edges, while the value of $E_C - E_V$ marks the so-called mobility gap. It is utmost important for the description of electrical properties of an amorphous semiconductor.

Concerning the optical properties (particularly the absorption of light), in a system like shown in Fig. 12.8, we have to distinguish two entirely different cases:

1. both the initial and final states of the quantum transition are spatially localized
2. at least one of the participating quantum states is delocalised.

For a non-vanishing transition matrix element, it is necessary that the wavefunctions of the initial and final quantum states are spatially overlapping (see formula 10.21). This is automatically fulfilled for quantum transitions which involve delocalised states. However, for localized quantum states, this requirement may make a quantum transition impossible even if the energy spacing between the states is suitable.

One should therefore expect that the transitions, which involve delocalised quantum states, give more intense contributions to the full absorption spectrum than the transitions between localized states.

The calculation of the imaginary part of the dielectric function follows the philosophy from Sect. 12.3 (indirect transition). The absence of electronic quasimomentum conservation makes (12.8) applicable:

$$\mathrm{Im}\varepsilon \propto |p_{nl}|^2 \int_0^{-\hbar\omega} D_V(E) D_C(E+\hbar\omega)\,\mathrm{d}E \tag{12.10}$$

Let us assume, that the light frequency is high enough to induce transitions from the valence band far into the conduction band. We will consequently regard a structureless conduction band, that may be described by a step-function according to:

$$D_C(E) \propto \theta(E - E_C) \Rightarrow D_C(E+\hbar\omega) \propto \theta(E - E_C + \hbar\omega)$$

From (12.10) we find:

$$\mathrm{Im}\varepsilon \propto |p_{nl}|^2 \int_0^{E_C-\hbar\omega} D_V(E)\mathrm{d}E \Rightarrow \frac{\mathrm{d}}{\mathrm{d}\omega}\left[\frac{\mathrm{Im}\varepsilon(\omega)}{|p_{nl}|^2}\right] \propto D_V(E_C - \hbar\omega)$$

Consequently, knowledge of the dielectric function and the behaviour of the transition matrix element allows to determine the shape of the valence band, as long as the conduction band may be regarded as structureless.

12.4.2 Tauc-Gap and Urbach-Tail

Let us now come to the more interesting case of not too high frequencies, so that quantum transitions are expected in the region of the band edges. This will give us an idea about the shape of the absorption edge in an amorphous semiconductor. Making (for simplicity) the assumption of parabolic band edges, in complete analogy to the indirect transitions in crystals, we obtain:

$$\begin{aligned} &D_V(E) \propto \sqrt{-E};\ D_C(E+\hbar\omega) \propto \sqrt{E+\hbar\omega-E_0} \Rightarrow \\ &\mathrm{Im}\varepsilon(\omega) \propto |p_{nl}|^2 (\hbar\omega - E_0)^2 \end{aligned} \tag{12.11}$$

where E_0 represents the so-called optical gap of the material (for the system shown in Fig. 12.8, it is always lower than the mobility gap). If the dispersion of the refractive index is negligible in the frequency range of interest, and if the matrix element of the dipole operator in (12.11) is also constant, from (12.11) and (2.18), (2.20a) we find for the absorption coefficient:

$$\sqrt{\frac{\alpha(\omega)}{\omega}} \propto (\hbar\omega - E_0) \tag{12.12}$$

This convenient expression relates the absorption coefficient to the optical gap, hence, the gap may be determined from experimental absorption coefficient data, fitting the data by means of (12.12). The thus determined optical gap is called the Cody gap. It is connected to the requirement of a constant transition matrix element of the dipole operator.

A somewhat modified dependence is obtained, when the matrix element of the momentum operator is supposed to be constant. We must then require, that $|p_{nl}|^2\omega^2 = \text{const.}$, and instead of (12.12) we find:

$$\sqrt{\alpha(\omega)\,\omega} \propto (\hbar\omega - E_0) \tag{12.13}$$

The thus defined optical gap is called the Tauc-gap. It is conveniently determined from the so-called Tauc-plot, where

$$\sqrt{\alpha(\omega)\,\omega}$$

is plotted against the photon energy. The Tauc gap is often applied in practise to characterize the optical properties of amorphous materials.

Nevertheless, all the optical gaps defined in this section are nothing else than fitting parameters in dependences like (12.12) and (12.13). This is a difference to the definition of forbidden zones as it is possible in the case of crystals, where the existence of energy gaps is a direct conclusion from the $E(\boldsymbol{k})$-dependence. On the contrary, in an amorphous solid it is quite possible that even in the 'forbidden' zone in Fig. 12.8, there exists a finite density of localized states. For that reason, the terminus "optical gap" is not well defined for amorphous semiconductors. On the other hand, the introduction

of the optical gap by a dependence like (12.13) gives at least a recipe for the unambiguous and convenient determination of a parameter that may be used to judge the quality of a prepared material with respect to certain optical applications. For that reason, these parameters are widely used in applied semiconductor research. In practice, (12.12) or (12.13) are used to fit experimentally determined absorption coefficients which are larger than approximately $10000\,\mathrm{cm}^{-1}$.

Keeping this in mind, it should be clear that an experimentally determined absorption coefficient that follows (12.13) should never be regarded as a proof for the validity of a density-of-states behaviour as postulated in (12.11). It may be shown quite easily, that the same type of absorption coefficient behaviour may be obtained assuming quite other shapes for $D(E)$. Let us for a moment return to Fig. 12.8. In contrast to the band shape presented there, let us assume, that in the band tail (they grey regions in Fig. 12.8, which symbolize the localized states), the density of states increases linearly with energy. For example, regarding the band tail of the valence band, we suppose:

$$D_V(E) \propto -E$$

On the other hand, we postulate that the conduction band is structureless:

$$D_C(E) \propto \theta(E - E_C)$$

Let us now consider transitions from the valence band tail into the conduction band. Neglecting transitions between localized states, we find:

$$\mathrm{Im}\varepsilon \propto |p_{nl}|^2 \int_0^{-\hbar\omega} D_V(E) D_C(E + \hbar\omega)\,\mathrm{d}E \propto |p_{nl}|^2 \int_0^{E_C - \hbar\omega} D_V(E)\mathrm{d}E \propto$$
$$|p_{nl}|^2 \int_0^{E_C - \hbar\omega} E\mathrm{d}E \propto |p_{nl}|^2 (\hbar\omega - E_C)^2 \tag{12.14}$$

This is exactly the same type of frequency dependence as in (12.11), although the latter was obtained assuming parabolic band edges. The physical sense of the optical gap in (12.14) is different from that in (12.11), it is now identical to $(E_C - E_B)$. We may include transitions from the valence band into the conduction band tail as well, and in this case, the observed optical gap will correspond to the lower value of $(E_C - E_B)$ and $(E_A - E_V)$.

We have already mentioned, that the so-called power-law dependences (12.11)–(12.14) are usually observed for absorption coefficients at the fundamental absorption edge which are larger than approximately $10000\,\mathrm{cm}^{-1}$. In the region of lower absorption, one usually observes an exponential increase of the absorption coefficient with frequency. This so-called Urbach tail is a general disorder-induced feature in solid state optics, it is also apparent at

the fundamental absorption edge of crystals due to the thermal motion of the lattice atoms. In the Urbach-tail region, the absorption coefficient is given by:

$$\alpha(\omega) = \alpha_{00} \mathrm{e}^{\frac{\omega}{\omega_{00}}} \tag{12.15}$$

where α_{00} and ω_{00} are constants. Up to now, the physical origin of the exponential behaviour of the absorption coefficient is not clear. According to (12.10), it may be caused by an exponentially increasing density of states at the band tails. It is also possible, that the exponential behaviour arises from the frequency-dependence of the matrix element. We will not discuss these theories, but rather look at an example how the mentioned absorption laws like (12.15) and (12.13) may work in practice.

As an example, let us regard an amorphous hydrogenated carbon (a-C:H) film, deposited onto a fused silica substrate. The carbon film has been produced by a plasma deposition technique and has a thickness of approximately 820 nm. The experimental transmission and reflection spectra of this sample are shown in Figs. 12.9a and 12.10a by the full and empty circles ("exp"). It has then been attempted to fit the spectra by means of a curve fitting procedure (compare Sect. 7.4.6). For the refractive index, according to (4.9) we assumed:

$$n^2 = A + B\nu^2$$

Neither of the dependences (12.13) or (12.15) fits the experimental spectra in the full spectral region. Therefore, the spectrum has been subdivided into two regions, which have been fitted separately assuming:

$$\nu < 13000\,\mathrm{cm}^{-1}: \ \alpha(\nu) = \alpha_{00} \mathrm{e}^{\frac{\nu}{\nu_{00}}} \quad \text{Urbach}$$

$$\nu > 13000\,\mathrm{cm}^{-1}: \ \sqrt{\alpha(\nu)\,\nu} = \text{const.} \times (h c \nu - E_0) \quad \text{Tauc}$$

In each of these dependencies, only two constant parameters have to be determined as a result of the fit. Figure 12.9a demonstrates the quality of the fit at low wavenumbers, assuming the exponentially increasing absorption coefficient according to Urbach's law. At higher wavenumbers, the theoretically calculated spectra ("theor") show significant deviations from the experimental data ("exp"). Figure 12.9b shows the corresponding optical constants, which are valid for wavenumbers below $13000\,\mathrm{cm}^{-1}$. Despite of the exponential increase of the absorption coefficient, we notice the normal dispersion of the refractive index. All fits have been performed minimizing (7.27) with respect to (7.25), (7.26) and (7.12)–(7.15).

Figure 12.10a, on the contrary, shows the fit at higher wavenumbers, assuming Tauc's law for the absorption coefficient. The fit is quite good at high wavenumbers, but insufficient at lower wavenumbers. The optical constants (Fig. 12.10b demonstrate the anomalous dispersion of the refractive index, as valid in the region of high absorption. The Tauc gap E_0 is determined this way to be equal to 1.14 eV.

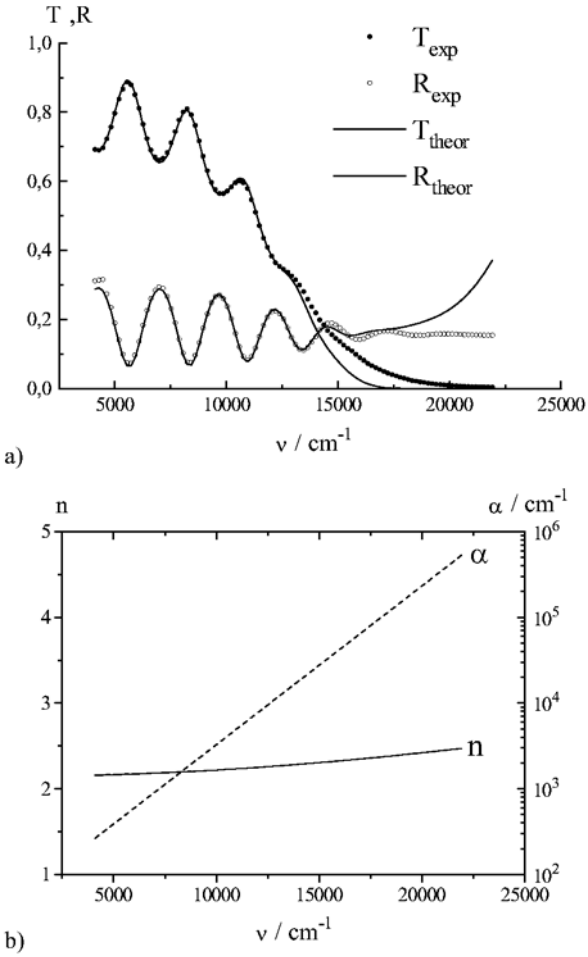

Fig. 12.9. a) Fit of the long wavelength section of the experimental spectra of an a-C:H film on fused silica assuming (12.15) for the absorption coefficient and (4.9) for the refractive index; **b**) optical constants corresponding to the theoretical spectra in **a**)

The final result is then obtained combining the results obtained from both fits. This is demonstrated in Fig. 12.11, where the relevant optical constants are plotted in the full wavelength range. Obviously, the optical constants obtained from both models are in good mutual accordance, as seen from the nearly continuous behaviour at the wavenumber of $13000\,\mathrm{cm}^{-1}$.

Figure 12.11 thus depicts a typical behaviour of the absorption coefficient in the vicinity of the fundamental absorption edge: an exponential Urbach tail at lower frequencies, followed by an absorption region where the absorption coefficient is described by a power-law.

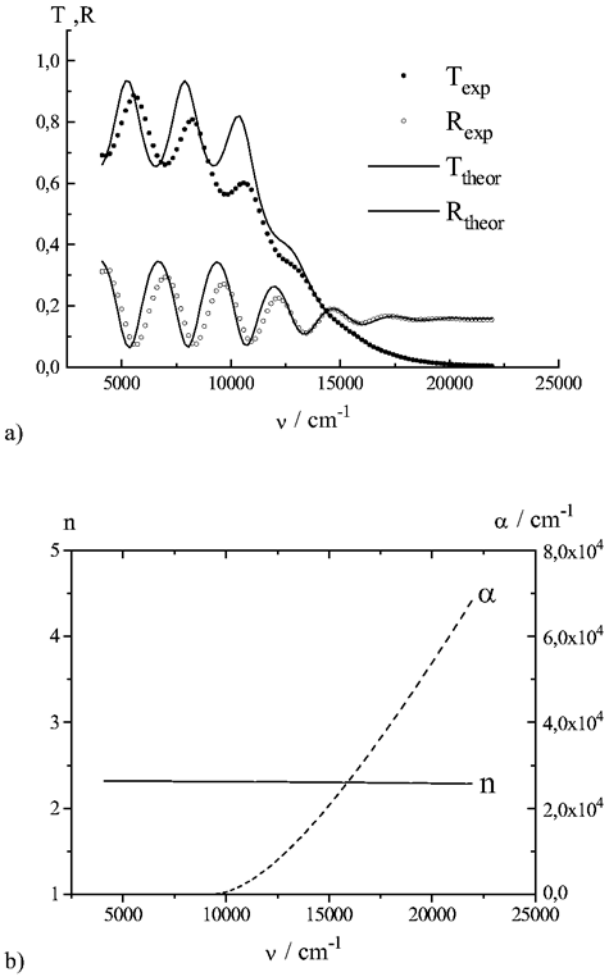

Fig. 12.10. **a**) Fit of the short wavelength section of the experimental spectra of an a-C:H film on fused silica assuming (12.13) for the absorption coefficient and (4.9) for the refractive index; **b**) optical constants corresponding to the theoretical spectra in **a**)

12.5 Resume from Chapters 10–12

12.5.1 Overview on Main Results

The Chaps. 10–12 have been devoted to the semiclassical theory of the optical constants of different kinds of matter. Let us shortly resume the main results of this third part of the present book.

– For a quantum system with discrete energy levels, the dielectric function appears to be determined by the resonance frequencies and intensities of

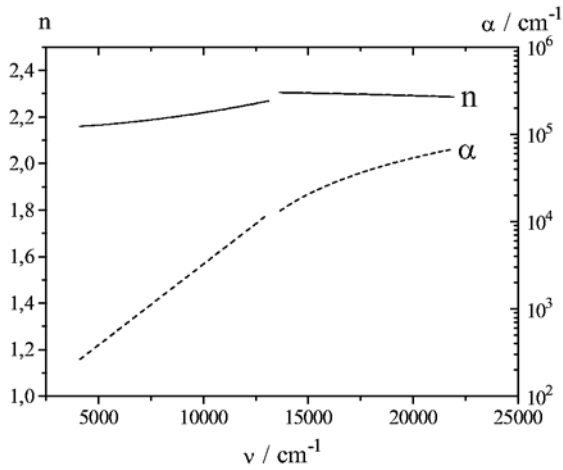

Fig. 12.11. Optical constants obtained from the merger of the two dependencies (12.15) and (12.13)

quantum transitions which are induced by the impinging light. It has the same general frequency dependence as that obtained from the classical multioscillator model. In the quantum mechanical description, the classical resonance frequencies appear to be replaced by transition frequencies between the single energy levels. The intensity of the quantum transition is determined by the transition matrix element of the perturbation operator and the populations of the participating energy levels.

– The transition matrix element of a given perturbation operator allows to classify a quantum transition as allowed or forbidden with respect to that type of perturbation. In optical spectroscopy, it is often sufficient to regard the electrical dipole interaction between the material system (an atom, or a molecule, or an elementary cell in a crystal) as the perturbation that induces the transition. If the dipole transition matrix element is zero, then the transition is dipole-forbidden. If the matrix element is different from zero, the corresponding transition is called to be allowed.
– If there occurs population inversion between two energy levels, the system does not absorb light at the corresponding resonance, but tends to amplify the incoming light by stimulated emission. This effect is in the fundament of the working principle of lasers.
– In a crystalline solid one finds energy bands instead of discrete energy levels. As it follows from quasimomentum conservation, in the case of direct transitions, it is the joint density of states that determines the main features of the shape of the dielectric function. This is especially evident in the case of broad energy bands, as it is typical for covalent materials. In molecular solids, the electronic wavefunction overlap between the molecules may be small, so that the bandwidth which might be formally

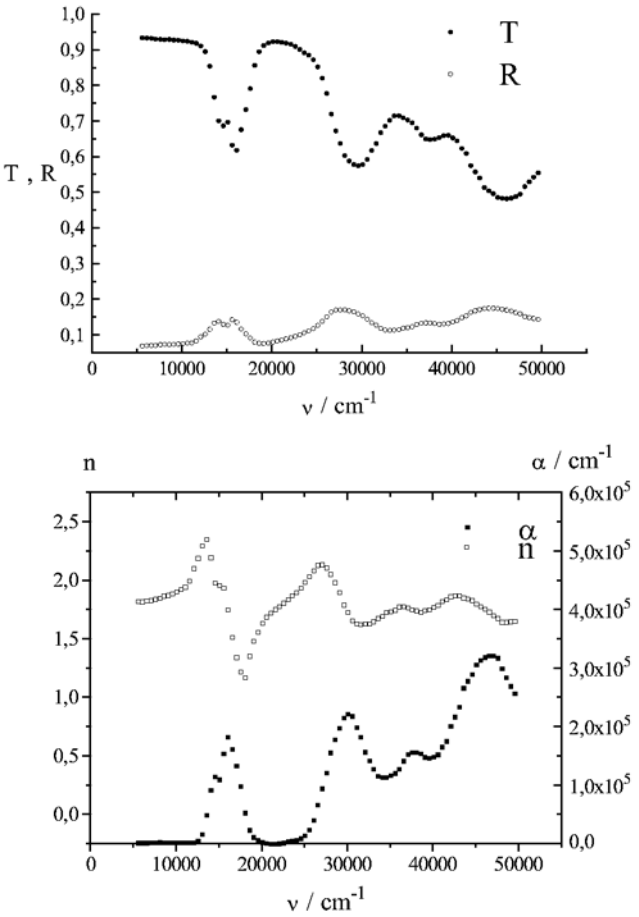

Fig. 12.12. Spectra of a CuPc-film (**top**) and corresponding optical constants (**bottom**)

calculated is also small. The optical behaviour of such a material is close to the behaviour of the molecules which build the solid. The spectra of such films may therefore be approximated by a few Lorentzian oscillators. As an example, Fig. 12.12 (top) shows the normal incidence transmission and reflection spectra of a 18 nm thick copperphthalocyanine (CuPc) film on fused silica. On bottom, one sees the corresponding optical constants, which are qualitatively close to the behaviour known from the multioscillator model (Fig, 4.2).

When the band structure $E(\boldsymbol{k})$ has to be considered, the resulting shape of the dielectric function becomes more complicated. Although it may formally be approximated by a continuous distribution of Lorentz-oscillators (for direct transitions this follows from (12.1)), it is dominated by the joint density of

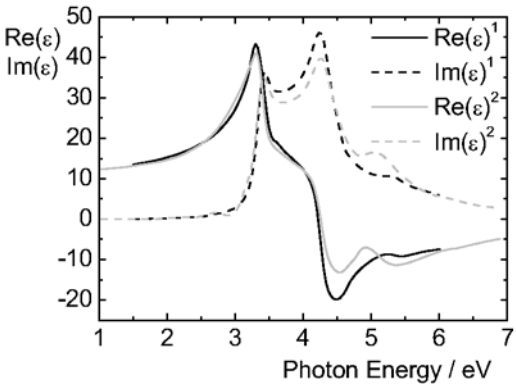

Fig. 12.13. Dielectric function of crystalline silicon versus photon energy as taken from two sources: [1] D.E. Aspnes, A.A. Studna: Dielectric functions and optical parameters of Si, Ge, GeAs, GeSb, InP, InAs and InSb from 1.5 to 6.0 eV, Phys. Rev. B **27**, No. 2, 985–1009 (1983); [2] В.В. Соболев, С.А. Алексеев, В.И. Донецких: Расчеты оптических функций полупроводников по соотношениям Крамерса–Кронига (Кишинев, Штиница 1976) (in Russian), (engl.: Calculation of semiconductor optical functions from Kramers–Kronig-relations)

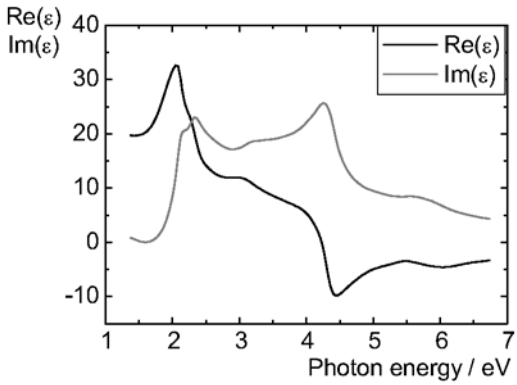

Fig. 12.14. Dielectric function of crystalline germanium versus photon energy; according to source 2 from Fig. 12.13

states and particularly the van-Hove-singularities. As examples, Figs. 12.13 and 12.14 show the dielectric function of crystalline silicon and germanium. In these experimental curves, all direct and indirect transitions that contribute to the dielectric functions in the given photon energy range are involved.

In amorphous solids, the absorption edge is usually dominated by a power-law region, followed by an Urbach tail at higher wavelength.

12.5.2 Problems

1. Estimate the lifetime of an excited discrete quantum state, which relaxes into the ground state by emission of a photon via dipole radiation. In (10.52), the charge should be equal to the elementary charge, while the transition matrix element of the coordinate could be 10^{-8} cm. The emission wavelength might be 500 nm.

   Answer: $\tau \approx 1.6 \times 10^{-8}$ s.

   Remark: This is a typical radiative lifetime for well-allowed dipole transitions (compare Sect. 4.1). For a lower absolute value of the transition matrix element, the lifetime will be larger.

2. Basing on Planck's formula, discuss the extraterrestrial spectrum of the sunlight and compare it with your experience as obtained under terrestrial conditions. The surface temperature of the sun is close to 6000 K.

3. Imagine an electron, which is allowed to perform a one-dimensional motion between two impermeable walls separated by the length L (a similar system is sketched in Fig. 10.3)! The electronic wavefunctions are given by:

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin\left(\frac{n\pi x}{L}\right); n = 1, 2, 3, \ldots ,$$

   and the eigenvalues by:

$$E_n = \frac{\hbar^2 \pi^2 n^2}{2mL^2}; n = 1, 2, 3, ...$$

   Calculate the transition matrix elements for electric dipole transitions between two arbitrary levels l and n!

   Result: $p_{nl}^2 = q^2 \frac{64L^2}{\pi^4} \frac{n^2 l^2}{(n-l)^4 (l+n)^4}$ $(l - n$ is odd)
   and $p_{nl} = 0$ for even values of $(l - n)$. q and m are the electron charge and mass, respectively.

   Remark: It appears that a large amount of the principally possible quantum transitions are dipole forbidden. This is a particular conclusion from a more general selection rule derived in quantum mechanics: In a centrosymmetric potential, the solutions of Schrödinger's equation in the coordinate picture are either even or odd functions of the coordinates with respect to the inversion centre. Those quantum states are called to have even or odd parity. Because the coordinate itself is an odd function, it appears that a dipole transition may only take place between quantum states that have different parity. Indeed, the regarded quantum well is centrosymmetric, and the allowed transitions correspond to transitions

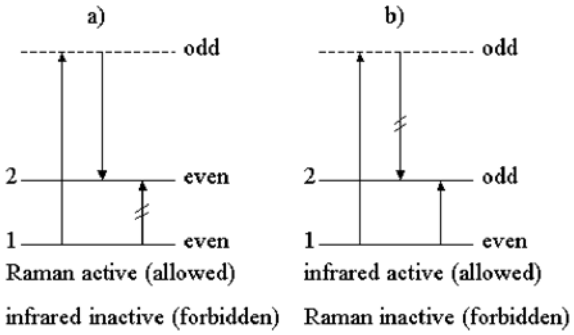

Fig. 12.15. Visualization of the alternative rule: **a**) the transition between 1 and 2 is dipole forbidden, but Raman-allowed; **b**) the transition between 1 and 2 is dipole allowed, but Raman-forbidden

from an even to an odd function or vice versa, but never from an even to another even function or an odd to another odd.
The parity selection rule is extremely important in the optical spectroscopy of any centrosymmetric system. A particular conclusion from this selection rule is the so-called alternative rule in vibrational spectroscopy: In a centrosymmetric system, an infrared-active transition (that means dipole-allowed transition in the IR, as typical for vibrational spectroscopy) cannot be Raman-active (allowed in Raman spectroscopy), and vice versa. The reason is illustrated in Fig. 12.15.
As shown in Fig. 12.15, an optical transition between two states of identical parity is possible by a Raman process via an intermediate virtual level of the other parity. On the other hand, a Raman-process cannot occur between states of different parity, because the intermediate virtual level must have the same parity as one of the levels 1 and 2. Consequently, infrared (T or R) spectroscopy and Raman spectroscopy are complementary methods, and may yield information about different quantum transitions. This is of importance in centrosymmetric molecules as well as special crystal classes. Concerning thin film technology, the diamond films (which should not be confused with diamond-like films) may be mentioned as a prominent example: Usually, it is the Raman-spectrum of a diamond film that serves as a criterion for the quality of the film. The reason is, that the zone-centre optical phonon in diamond is Raman-active, but not infrared-active.

4. The oscillator strength of a quantum transition of a single electron between the states l and n is defined as:

$$f_{nl} = \frac{2m|x_{nl}|^2\omega_{nl}}{\hbar}.$$

Basing on the correspondence principle, derive the sum rule for the oscillator strength:

$$\sum_{n \neq l} f_{nl} = 1!$$

Remark: Use expressions (5.10) and (11.31) together with (3.24)! Require further, that the thus obtained expressions for the dielectric function in classics and quantum mechanics have identical asymptotes for infinitively large photon energies ($\omega \to \infty$). From that, you should find:

$$\sum_{l} \sum_{n>l} \frac{|x_{nl}|^2}{\hbar} \cdot 2\omega_{nl} \cdot m \left(\rho_{ll}^{(0)} - \rho_{nn}^{(0)} \right) = 1$$

You will now need to perform some algebraic operations. This leads to:

$$\begin{aligned}
&\frac{2m}{\hbar} \left\{ \sum_{l} \sum_{n>l} |x_{nl}|^2 \omega_{nl} \cdot \rho_{ll}^{(0)} + \sum_{l} \sum_{n>l} |x_{nl}|^2 \omega_{ln} \cdot \rho_{nn}^{(0)} \right\} \\
&= \frac{2m}{\hbar} \left\{ \sum_{l} \sum_{n>l} |x_{nl}|^2 \omega_{nl} \cdot \rho_{ll}^{(0)} + \sum_{n} \sum_{l<n} |x_{nl}|^2 \omega_{ln} \cdot \rho_{nn}^{(0)} \right\} \\
&= \frac{2m}{\hbar} \left\{ \sum_{l} \sum_{n>l} |x_{nl}|^2 \omega_{nl} \cdot \rho_{ll}^{(0)} + \sum_{l} \sum_{n<l} |x_{nl}|^2 \omega_{nl} \cdot \rho_{ll}^{(0)} \right\} \\
&= \sum_{l} \left[\sum_{n \neq l} \frac{2m|x_{nl}|^2 \omega_{nl}}{\hbar} \right] \rho_{ll}^{(0)} = 1
\end{aligned}$$

This should be valid for any time-independent and stationary value $\rho_{ll}^{(0)}$. Therefore, the term in parentheses should not depend on l. On the other hand, we have:

$$\sum_{l} \rho_{ll}^{(0)} = 1$$

From that, it follows that

$$\sum_{n \neq l} \frac{2m|x_{nl}|^2 \omega_{nl}}{\hbar} = 1$$

must be fulfilled.

5. Starting from the sum rule for the oscillator strength, calculate the absolute values of the transition matrix elements of the coordinate of a

one-dimensional harmonic oscillator. You only need to assume the selection rule $n \to n \pm 1$, valid for the harmonic oscillator.

Answer: See (10.47).

Remark: These results have been obtained by Werner Heisenberg in terms of his matrix theory of quantum mechanics one year before Schrödinger's Equation has been formulated.

6. Basing on problem 3, calculate the oscillator strength for the allowed dipole transitions of an electron confined between two walls (one-dimensional case).

Result: $f_{nl} = \dfrac{64}{\pi^2} \dfrac{n^2 l^2}{(n-l)^3 (l+n)^3}$ for odd $n - l$

For some (n, l)-pairs, exemplify the validity of the sum rule for the oscillator strength. Keep in mind, that the oscillator strength may be negative.

7. Make sure, that the polarizability as calculated by (11.31) has the correct dimension (m^3). Repeat the same for the dielectric function according to (12.1), which should be dimensionless!

8. Estimate the Tauc gap immediately from the normal incidence transmission spectrum of an amorphous semiconductor film on a transparent substrate in the region of low transmittance!

Answer: $E_0 \approx \dfrac{hc}{\sqrt{\lambda_1 \lambda_2}} \dfrac{\sqrt{-\lambda_2 \ln T(\lambda_2)} - \sqrt{-\lambda_1 \ln T(\lambda_1)}}{\sqrt{-\lambda_1 \ln T(\lambda_2)} - \sqrt{-\lambda_2 \ln T(\lambda_1)}}$

Remark: For high absorption, from (7.13), (7.15) and (7.25) it follows that

$$T \approx f(n, n_{sub})\, \mathrm{e}^{-\alpha d},$$

where d is the film thickness. From that, and assuming the power law (12.13) for the absorption coefficient, we find:

$$\alpha d = \ln \frac{f}{T} = \frac{const. \times d}{\omega} (\hbar\omega - E_0)^2 .$$

Regarding f as a constant in a limited wavelength region with $1 > f \gg T$, and assuming two wavelength values λ_1 and λ_2, the product const.$\times d$ may be excluded from the equation. From that, the final result is obtained.
It is useful to demonstrate the relative accuracy by means of a simple example. Let us regard the transmission spectrum from Fig. 12.9a. We choose two data points from the spectrum: $\lambda_1 = 640\,\mathrm{nm}$ corresponding to $T(\lambda_1) = 0.11$ and $\lambda_2 = 586\,\mathrm{nm}$ corresponding to $T(\lambda_2) = 0.05$. From the derived equation, we obtain an estimated gap of 1.12 eV, quite close to the more accurate value of 1.14 eV as obtained from the spectra fit. The merit of gap estimation by this equation decreases with increasing refractive indices and refractive index dispersion.

9. As mentioned at the end of Sect. 10.7.2, it would be necessary to excite a large amount of mutually adjacent longitudinal resonator modes in order to cause a laser to supply us with short light pulses instead of a continuous light wave. In fact, it turns out to be sufficient to excite these laser modes with well-defined mutual phase relations (by means of a so-called *mode-locking* mechanism) in order to force a laser to produce a sequence of short light pulses. You shall now show theoretically, that the superposition of trains of travelling electromagnetic waves with different but equidistant frequency values and identical zero phase values is indeed identical to a travelling sequence of short light pulses. You may assume for simplicity, that all of the participating waves have the same amplitude E_0.

   Solution: Let us assume that a number of M adjacent modes is excited in a cavity of length L. Their angular frequency values are given by:

$$\omega, \omega - \frac{c\pi}{L}, \omega - \frac{2c\pi}{L}, \ldots, \omega - \frac{(M-1)\,c\pi}{L}.$$

   Assuming identical zero phase values, the full electric field strength may be calculated according to:

$$\begin{aligned} E &= E_0\,\mathrm{e}^{-\mathrm{i}\omega t} + E_0\,\mathrm{e}^{-\mathrm{i}\omega t}\,\mathrm{e}^{\mathrm{i}\frac{c\pi}{L}t} + \ldots \\ &= E_0\,\mathrm{e}^{-\mathrm{i}\omega t}\sum_{j=1}^{M}\left[\mathrm{e}^{\mathrm{i}\frac{c\pi}{L}t}\right]^{j-1} = E_0\,\mathrm{e}^{-\mathrm{i}\omega t}\frac{\mathrm{e}^{\mathrm{i}\frac{Mc\pi}{L}t}-1}{\mathrm{e}^{\mathrm{i}\frac{c\pi}{L}t}-1}. \end{aligned}$$

   The intensity of the light is proportional to the square of the modulus of the electric field amplitude. Hence we have to discuss the function:

$$\begin{aligned} |E|^2 &= |E_0|^2\frac{\left(\cos\frac{Mc\pi}{L}t-1\right)^2+\sin^2\frac{Mc\pi}{L}t}{\left(\cos\frac{c\pi}{L}t-1\right)^2+\sin^2\frac{c\pi}{L}t} = |E_0|^2\frac{\left(1-\cos\frac{Mc\pi}{L}t\right)}{\left(1-\cos\frac{c\pi}{L}t\right)} \\ &= |E_0|^2\frac{\sin^2\frac{Mc\pi}{2L}t}{\sin^2\frac{c\pi}{2L}t} \end{aligned}$$

   From here we see, that the light intensity is time dependent. It is moreover periodic with a repetition time of $2L/c$. This is exactly the time necessary for one loop of the light in the resonator. Therefore, the laser generates a periodic intensity profile, which is shown in Fig. 12.16 for different values of M. At the moments:

$$t_m = \frac{2L}{c}m \equiv T_0 m;\quad m = 0, 1, 2, \ldots$$

   this function reaches its maximal values according to:

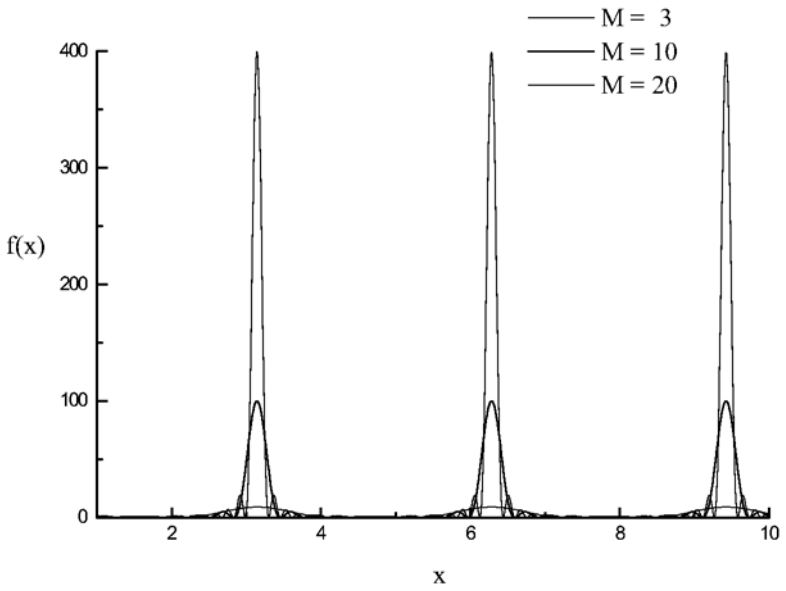

Fig. 12.16. Shape of the function $f(x) = \sin^2(Mx)/\sin^2 x$

$$|E|^2_{\max} = |E_0|^2 M^2.$$

The more modes are involved into the process, the more intense the peak is. Please note that this result is tightly connected to the assumed identical zero phase values of the modes: In that (coherent) case, the field strength values add up to the full field strength, which is proportional to M. Therefore, the intensity turns out to be proportional to the square of M. If the phases would be distributed in a stochastic manner (incoherent superposition), one would have to superimpose the intensities of the individual wave trains, so that the resulting intensity would be proportional to M. The background physics is the same as in the discussion from Sect. 7.2.

We come to the result, that in mode-locking conditions, the laser produces a sequence of short and intense light pulses with a repetition rate T_0^{-1}. The pulse duration is $2T_0/M$. Again, the more modes are involved into the process, the shorter the pulses are. This is consistent with the uncertainty principle, because a larger number of longitudinal modes is connected with a broader spectral width of the laser light, necessary for a shorter pulse duration. Practically, by means of mode-locking, subpicosecond light pulses may be generated.

Remark: This is a nice example to demonstrate, that laser light may reach extraordinarily high electric field strength and intensity values. As we mentioned in Sect. 2.2, the linear material (2.4) corresponds to the

linear term in an expansion of the polarization into a Taylor's power series. Clearly, at very high field strength values, it is insufficient to keep only the linear term, instead, we have to regard nonlinear polarization terms as well. That means, that a correct description of the interaction of intense laser light with matter requires a more general description, which is subject of the field of nonlinear optics. The last part of this book is therefore dedicated to basic effects of nonlinear optics, relevant at very high electric field strength values.

Part IV

Basics of Nonlinear Optics

13 Some Basic Effects of Nonlinear Optics

13.1 Nonlinear Susceptibilities: Phenomenological Approach

13.1.1 General Idea

In Chap. 2, we formulated the linear material equation (2.4) as a special case of the more general formulation (2.3). All optical effects, which have been described so far basing on (2.4), belong to the field of linear optics (LO).

On the other hand it is clear, that in general, the validity of (2.3) may require to consider higher order terms in the expansion of the polarisation into a power series of the field strength. The natural generalization of (2.9) would then read as:

$$\begin{aligned} \boldsymbol{D} &= \varepsilon_0 \boldsymbol{E} + \boldsymbol{P} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}^{(1)} + \boldsymbol{P}^{(2)} + \boldsymbol{P}^{(3)} + \dots \\ &= \varepsilon_0 \Big\{ \boldsymbol{E} + \chi^{(1)} \boldsymbol{E} + \chi^{(2)} : \boldsymbol{EE} + \chi^{(3)} : \boldsymbol{EEE} + \dots \Big\} \end{aligned} \tag{13.1}$$

where the superscripts (1)–(3) in the polarization indicate polarization contributions that increase in a linear (1), quadratic (2) or cubic (3) manner with field strength. Even higher order polarization terms are possible, however, we will restrict our discussion to the mentioned terms. $P^{(1)}$ indicates the so-called linear polarization, while all higher order polarization terms form the nonlinear polarization. They are responsible for the effects of nonlinear optics (NLO), which are entirely different from those found in linear optics.

In (13.1), the value $\chi^{(1)}$ is nothing else than the familiar linear susceptibility as defined earlier by (2.7). The proportionality coefficients $\chi^{(2)}$, $\chi^{(3)}$ and so on represent the quadratic, cubic, and higher order susceptibilities. Similar to linear optics, they carry the information about the specifics of the material interacting with the light wave.

Regardless on the concrete values of the linear and nonlinear susceptibilities, it is clear that (13.1) converges to the linear equation (2.9) when the electric field strength becomes sufficiently small. Figure 13.1 demonstrates the principal dependence of the individual contributions to the polarization on the field strength. At weak fields, the linear contribution is dominating, and in this case the previously discussed effects of linear optics (LO) are sufficient to describe the optical properties of a material. At higher field strength,

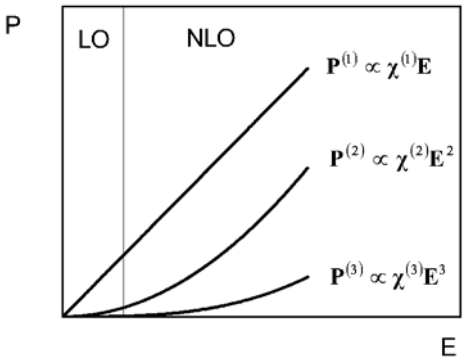

Fig. 13.1. Linear and nonlinear contributions to the polarization of a nonlinear medium

the nonlinear polarization becomes significant, so that we enter the field of nonlinear optics. Here, both linear and nonlinear contributions to the full polarization have to be taken into account. Practically, this is usually of significance when we deal with laser light. As a kind of slang, a medium which shows nonlinear optical properties is sometimes called a nonlinear medium.

Equation (13.1) is written in a somewhat symbolic manner. In fact, the products at the right hand of (13.1) have to be understood as tensor products, the susceptibilities themselves represent tensors of different orders. We will now present some alternative versions of writing material equations like (13.1).

Let P_i be the i-th Cartesian component of the polarization with $i = x$, y, z, we may write:

$$\begin{aligned}
\boldsymbol{P}^{(1)} &= \varepsilon_0 \chi^{(1)} \boldsymbol{E} \Leftrightarrow P_i^{(1)} = \varepsilon_0 \sum_{j=x,y,z} \chi_{ij}^{(1)} E_j \,; \\
\boldsymbol{P}^{(2)} &= \varepsilon_0 \chi^{(2)} \boldsymbol{E}\boldsymbol{E} \Leftrightarrow P_i^{(2)} = \varepsilon_0 \sum_{j=x,y,z} \sum_{k=x,y,z} \chi_{ijk}^{(2)} E_j E_k ; \\
\boldsymbol{P}^{(3)} &= \varepsilon_0 \chi^{(3)} \boldsymbol{E}\boldsymbol{E}\boldsymbol{E} \Leftrightarrow P_i^{(3)} = \varepsilon_0 \sum_{j=x,y,z} \sum_{k=x,y,z} \sum_{l=x,y,z} \chi_{ijkl}^{(3)} E_j E_k E_l
\end{aligned} \tag{13.2}$$

The first equation in (13.2) is nothing else than the general (anisotropic) version of the linear material equation, as already used in Sect. 6.5. According to Sect. 6.5 the linear susceptibility may be regarded as a 3×3 quadratic matrix. So that another form of writing the linear material equation is:

$$\boldsymbol{P}^{(1)} = \begin{pmatrix} P_x^{(1)} \\ P_y^{(1)} \\ P_z^{(1)} \end{pmatrix} = \varepsilon_0 \begin{pmatrix} \chi_{xx}^{(1)} & \chi_{xy}^{(1)} & \chi_{xz}^{(1)} \\ \chi_{yx}^{(1)} & \chi_{yy}^{(1)} & \chi_{yz}^{(1)} \\ \chi_{zx}^{(1)} & \chi_{zy}^{(1)} & \chi_{zz}^{(1)} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix}$$

The second order susceptibility, $\chi^{(2)}$, represents a $3 \times 3 \times 3$ tensor with 27 components, while $\chi^{(3)}$ has 81 components.

13.1.2 Formal Treatment and Simple Second Order Nonlinear Optical Effects

In order to get some idea on the frequency arguments relevant for the nonlinear susceptibilities introduced previously, we will have to proceed the same way as we did in the case of the linear material (2.4). So when starting from (2.3), and assuming a nonlinear material equation while keeping the requirements of causality and homogeneity in time, we may write:

$$\begin{aligned} P(t) &= P^{(0)} + P^{(1)} + P^{(2)} + P^{(3)} + \ldots = \\ &= \varepsilon_0 \,\text{const.} + \varepsilon_0 \int_0^\infty \boldsymbol{\kappa}^{(1)}(\xi)\, E(t-\xi)\, \mathrm{d}\xi + \\ &+ \varepsilon_0 \int_0^\infty \int_0^\infty \boldsymbol{\kappa}^{(2)}(\xi_1, \xi_2)\, E(t-\xi_1)\, E(t-\xi_1-\xi_2)\, \mathrm{d}\xi_1 \mathrm{d}\xi_2 + \\ &+ \varepsilon_0 \int_0^\infty \int_0^\infty \int_0^\infty \boldsymbol{\kappa}^{(3)}(\xi_1, \xi_2, \xi_3)\, E(t-\xi_1)\, E(t-\xi_1-\xi_2) \times \\ &\quad \times E(t-\xi_1-\xi_2-\xi_3)\, \mathrm{d}\xi_1 \mathrm{d}\xi_2 \mathrm{d}\xi_3 + \ldots \end{aligned} \tag{13.3}$$

Equation (13.3) is nothing else than a nonlinear generalization of (2.5). We do not regard ferroelectrics, so that the constant, field-independent contribution $P^{(0)}$ is assumed to be zero. In (13.3), the response functions $\boldsymbol{\kappa}^{(i)}$ again represent tensors of a range according to (13.2).

When comparing (2.5) and (13.3), we notice a further complication of the mathematical treatment of nonlinear optical processes. Because both the electric field strength and the polarization are real physical values, the response functions $\boldsymbol{\kappa}^{(i)}$ must be real as well. Nevertheless, in linear optics, we were used to work with complex fields and polarizations. The reason is, that in the linear equation (2.5), we may make use of the superposition principle. Regarding the identity:

$$|E_0| \cos(\omega t + \varphi) = \frac{1}{2}\left[|E_0|\, \mathrm{e}^{\mathrm{i}\omega t}\mathrm{e}^{\mathrm{i}\varphi} + |E_0|\, \mathrm{e}^{-\mathrm{i}\omega t}\mathrm{e}^{-\mathrm{i}\varphi}\right] = \frac{1}{2}\left[E_0^* \mathrm{e}^{\mathrm{i}\omega t} + E_0 \mathrm{e}^{-\mathrm{i}\omega t}\right]$$

with

$$E_0 \equiv |E_0|\, \mathrm{e}^{-\mathrm{i}\varphi}$$

for the linear polarization it is found (2.5):

$$P(t) = \frac{\varepsilon_0}{2}\left[E_0 \mathrm{e}^{-\mathrm{i}\omega t}\int_0^\infty \kappa(\xi)\,\mathrm{e}^{\mathrm{i}\omega\xi}\mathrm{d}\xi + E_0^* \mathrm{e}^{\mathrm{i}\omega t}\int_0^\infty \kappa(\xi)\,\mathrm{e}^{-\mathrm{i}\omega\xi}\mathrm{d}\xi\right]$$

This relation defines a real polarization, because the first term (which is the analogue to (2.6)) is conjugated complex to the second one. The second term does therefore not contain any new physical information, and all derivations in linear optics may be performed basing on the first term only, operating with a complex polarization. When in the final result the real polarization is required, one simply has to add the conjugated complex term, and all will be fine.

This treatment is impossible in nonlinear optics, because such a treatment would result in a loss of polarization terms. This may be demonstrated by a simple example. Let us regard the quadratic nonlinearity according to the simplified equation:

$$P^{(2)} = \varepsilon_0 \chi^{(2)} E^2$$

The assumption

$$E = \frac{E_0}{2}\mathrm{e}^{-\mathrm{i}\omega t}$$

leads us to

$$P^{(2)} = \varepsilon_0 \chi^{(2)} \frac{E_0^2}{4}\,\mathrm{e}^{-2\mathrm{i}\omega t} \tag{13.4a}$$

For simplicity, the amplitude of the electric field strength should be real throughout this discussion. We find, that the assumed time dependence of the electric field leads to a polarization in the medium which oscillates with twice the frequency of the incoming field. Of course, such an oscillating polarization gives rise to the generation of an electromagnetic wave at the angular frequency 2ω. This means, that at least a part of the energy of the ingoing wave is transferred to a new wave with the doubled frequency, an effect which is called *Second Harmonic Generation* (SHG). SHG is the most prominent effect of nonlinear optics in media with a quadratic nonlinearity.

Let us now regard another case. We assume:

$$E = \frac{E_0}{2}\,\mathrm{e}^{+\mathrm{i}\omega t}$$

The resulting second order polarization is:

$$P^{(2)} = \varepsilon_0 \chi^{(2)} \frac{E_0^2}{4}\,\mathrm{e}^{2\mathrm{i}\omega t} \tag{13.4b}$$

(13.4c)

Both equations (13.4a) and (13.4b) describe a polarization that oscillates at a frequency 2ω. Adding up (13.4a) and (13.4b) will lead to a real value of the polarization, but it will not give any new physical effects.

Let us now regard a real field strength given by the algebraic sum of the versions discussed so far. We assume:

$$E = \frac{E_0}{2}\left(\mathrm{e}^{+\mathrm{i}\omega t} + \mathrm{e}^{-\mathrm{i}\omega t}\right) = E_0 \cos\omega t$$

The corresponding second order polarization becomes:

$$P^{(2)} = \varepsilon_0 \chi^{(2)} \frac{E_0^2}{4}\left(\mathrm{e}^{2\mathrm{i}\omega t} + \mathrm{e}^{-2\mathrm{i}\omega t} + 2\right) \tag{13.4d}$$

(13.4e)

This is more than what has been predicted by the algebraic sum from (13.4a) and (13.4b). Although the SHG terms are present as expected, there is a further term, which corresponds to a time-independent (static) polarization, created by the quadratic nonlinearity as the result of a nonlinear effect called *optical rectification*. Equation (13.4c) states, that as the result of the nonlinear interaction of a monochromatic wave with matter, second order polarization terms occur that are constant or oscillating with twice the frequency of the incoming wave.

We see, that the application of the simplified complex electric fields will lead to a serious loss of information, when we deal with nonlinear optics. Therefore, in the forthcoming, we will always regard a real expression for the electric fields according to:

$$E(t) = \frac{1}{2}\sum_j E_{0j}\mathrm{e}^{-\mathrm{i}\omega_j t} + \text{c.c.} \tag{13.5}$$

with c.c. – conjugate complex value.

Being equipped with a grammar expression for the electric field like (13.5), we may discuss the complete output of (13.3) for the quadratic polarization. For simplicity, let us use a scalar version of (13.3), with scalar electric field amplitudes in (13.5). We obtain:

$$\begin{aligned} P^{(2)}(t) &= \varepsilon_0 \int_0^\infty \int_0^\infty \kappa^{(2)}(\xi_1, \xi_2) \cdot \left[\frac{1}{2}\sum_j E_{0j}\mathrm{e}^{-\mathrm{i}\omega_j t}\mathrm{e}^{\mathrm{i}\omega_j \xi_1} + \text{c.c.}\right] \\ &\times \left[\frac{1}{2}\sum_l E_{0l}\mathrm{e}^{-\mathrm{i}\omega_l t}\mathrm{e}^{\mathrm{i}\omega_l \xi_1}\mathrm{e}^{\mathrm{i}\omega_l \xi_2} + c.c.\right]\mathrm{d}\xi_1\mathrm{d}\xi_2 \\ &= \frac{1}{4}\varepsilon_0 \int_0^\infty \int_0^\infty \kappa^{(2)}(\xi_1, \xi_2)\, f(t, \xi_1, \xi_2)\, \mathrm{d}\xi_1\mathrm{d}\xi_2 \end{aligned}$$

$$\begin{aligned} f &= \sum_j \left(E_{0j}\mathrm{e}^{-\mathrm{i}\omega_j t}\mathrm{e}^{\mathrm{i}\omega_j \xi_1} + E_{0j}^*\mathrm{e}^{\mathrm{i}\omega_j t}\mathrm{e}^{-\mathrm{i}\omega_j \xi_1}\right) \\ &\times \sum_l \left(E_{0l}\mathrm{e}^{-\mathrm{i}\omega_l t}\mathrm{e}^{\mathrm{i}\omega_l \xi_1}\mathrm{e}^{\mathrm{i}\omega_l \xi_2} + E_{0l}^*\mathrm{e}^{\mathrm{i}\omega_l t}\mathrm{e}^{-\mathrm{i}\omega_l \xi_1}\mathrm{e}^{-\mathrm{i}\omega_l \xi_2}\right) \end{aligned}$$

$$
\begin{aligned}
&= \sum_j E_{0j} \sum_l \left[E_{0l} \mathrm{e}^{-\mathrm{i}(\omega_j+\omega_l)t} \mathrm{e}^{\mathrm{i}(\omega_j+\omega_l)\xi_1} \mathrm{e}^{\mathrm{i}\omega_l\xi_2} \right] \\
&+ \sum_j E_{0j}^* \sum_l \left[E_{0l} \mathrm{e}^{-\mathrm{i}(\omega_l-\omega_j)t} \mathrm{e}^{\mathrm{i}(\omega_l-\omega_j)\xi_1} \mathrm{e}^{\mathrm{i}\omega_l\xi_2} \right] \\
&+ \sum_j E_{0j} \sum_l \left[E_{0l}^* \mathrm{e}^{-\mathrm{i}(\omega_j-\omega_l)t} \mathrm{e}^{\mathrm{i}(\omega_j-\omega_l)\xi_1} \mathrm{e}^{-\mathrm{i}\omega_l\xi_2} \right] \\
&+ \sum_j E_{0j}^* \sum_l \left[E_{0l}^* \mathrm{e}^{+\mathrm{i}(\omega_j+\omega_l)t} \mathrm{e}^{-\mathrm{i}(\omega_j+\omega_l)\xi_1} \mathrm{e}^{-\mathrm{i}\omega_l\xi_2} \right] \\
P^{(2)}(t) &= \frac{1}{4}\varepsilon_0 \sum_j \sum_l E_{0j} E_{0l} \mathrm{e}^{-\mathrm{i}(\omega_j+\omega_l)t} \times \\
&\times \int_0^\infty \int_0^\infty \kappa^{(2)}(\xi_1,\xi_2)\, \mathrm{e}^{\mathrm{i}(\omega_j+\omega_l)\xi_1} \mathrm{e}^{\mathrm{i}\omega_l\xi_2} \mathrm{d}\xi_1 \mathrm{d}\xi_2 \\
&+ \frac{1}{4}\varepsilon_0 \sum_j \sum_l E_{0j}^* E_{0l} \mathrm{e}^{-\mathrm{i}(\omega_l-\omega_j)t} \times \\
&\times \int_0^\infty \int_0^\infty \kappa^{(2)}(\xi_1,\xi_2)\, \mathrm{e}^{\mathrm{i}(\omega_l-\omega_j)\xi_1} \mathrm{e}^{\mathrm{i}\omega_l\xi_2} \mathrm{d}\xi_1 \mathrm{d}\xi_2 \\
&+ \frac{1}{4}\varepsilon_0 \sum_j \sum_l E_{0j} E_{0l}^* \mathrm{e}^{-\mathrm{i}(\omega_j-\omega_l)t} \times \\
&\times \int_0^\infty \int_0^\infty \kappa^{(2)}(\xi_1,\xi_2)\, \mathrm{e}^{\mathrm{i}(\omega_j-\omega_l)\xi_1} \mathrm{e}^{-\mathrm{i}\omega_l\xi_2} \mathrm{d}\xi_1 \mathrm{d}\xi_2 \\
&+ \frac{1}{4}\varepsilon_0 \sum_j \sum_l E_{0j}^* E_{0l}^* \mathrm{e}^{\mathrm{i}(\omega_j+\omega_l)t} \times \\
&\times \int_0^\infty \int_0^\infty \kappa^{(2)}(\xi_1,\xi_2)\, \mathrm{e}^{-\mathrm{i}(\omega_j+\omega_l)\xi_1} \mathrm{e}^{-\mathrm{i}\omega_l\xi_2} \mathrm{d}\xi_1 \mathrm{d}\xi_2
\end{aligned}
\tag{13.6}
$$

The general conclusion from (13.6) is, that the nonlinear second order polarization oscillates with all sum and difference frequencies resulting from the primary frequencies of the incoming field (13.5). Correspondingly, one speaks on *Sum Frequency Generation* (SFG) and *Difference Frequency Generation* (DFG) in nonlinear optics. The previously regarded effects of Second

Harmonic Generation and Optical Rectification appear as particular cases in (13.6), namely when $\omega_j = \omega_l$.

In complete analogy to the treatment in the linear case, the integral terms in (13.6) form the second order susceptibilities. They are expected to show a somewhat more complicated dispersion behaviour than the linear susceptibility, because they are dependent on two incident frequencies and the particular way in which they combine to the frequency of the polarization. Formally, (13.6) may be rewritten as:

$$\begin{aligned} P^{(2)}(t) = \frac{1}{4}\varepsilon_0 \sum_j \sum_l E_{0j} E_{0l} \mathrm{e}^{-\mathrm{i}(\omega_j+\omega_l)t} \chi^{(2)}(\omega = \omega_j + \omega_l) + \text{c.c} \\ + \frac{1}{4}\varepsilon_0 \sum_j \sum_l E_{0j}^* E_{0l} \mathrm{e}^{-\mathrm{i}(\omega_l-\omega_j)t} \chi^{(2)}(\omega = \omega_l - \omega_j) + \text{c.c} \end{aligned} \tag{13.7}$$

The concrete expressions for $\chi^{(2)}$ follow from the comparison with (13.6).

The frequency arguments in (13.7) have to be understood in the following way: The first frequency indicates the resulting frequency of the second order polarization. The following frequencies indicate the frequencies of the electric fields forming the polarization and the particular way of their combination. Thus, the effects described so far correspond to the following susceptibilities:

$$\chi^{(2)}(\omega = \omega_l + \omega_j) \Leftrightarrow \text{SFG}$$

$$\chi^{(2)}(\omega = \omega_l - \omega_j) \Leftrightarrow \text{DFG}$$

$$\chi^{(2)}(2\omega = \omega + \omega) \Leftrightarrow \text{SHG}$$

$$\chi^{(2)}(0 = \omega - \omega) \Leftrightarrow \text{optical rectification}$$

We will not perform a similarly detailed discussion of the properties of nonlinear susceptibilities as we did in the case of the linear susceptibility. We only remark that there exist several symmetry relations which may reduce the quantity of nonzero and independent components of the susceptibility. Our task will rather be to define nonlinear optical effects which may be of significance for our particular subject, namely the optics of thin films and optical effects at surfaces.

We will proceed with an extremely important selection rule: In a medium with an inversion centre (or so-called centrosymmetric materials), all components of any even-order nonlinear susceptibility are zero in the dipole approximation. For the particular case of the second order susceptibility, this is a direct conclusion from (13.7). Imagine an electric field like (13.5), which causes a certain second order polarization. From the supposed inversion symmetry it is clear, that an inversion of the electric field strength should be accompanied by an inversion of the polarization:

$$E \to -E \Rightarrow P \to -P \Rightarrow P^{(2)}(-E) = -P^{(2)}(+E)$$

On the other hand, it follows from (13.7), that

$$\begin{aligned} P^{(2)}(-E) &= \frac{1}{4}\varepsilon_0 \sum_j \sum_l (-E_{0j})(-E_{0l})\, \mathrm{e}^{-\mathrm{i}(\omega_j+\omega_l)t} \chi^{(2)}(\omega = \omega_j + \omega_l) + \text{c.c} \\ &+ \frac{1}{4}\varepsilon_0 \sum_j \sum_l \left(-E_{0j}^*\right)(-E_{0l})\, \mathrm{e}^{-\mathrm{i}(\omega_l-\omega_j)t} \chi^{(2)}(\omega = \omega_l - \omega_j) + \text{c.c} \\ &= P^{(2)}(+E) \end{aligned}$$

Both conditions together may only be fulfilled, when the second order polarisation is zero for any assumed electric field configuration. This means, that the second order susceptibility must be zero. This kind of discussion may be performed for any even order susceptibility, but obviously not for the odd order susceptibilities.

As it will be seen later in the quantum mechanical treatment, the vanishing of even order susceptibilities in media with inversion symmetry is a direct conclusion from the parity selection rule. For us it is important to notice, that second (and other even) order nonlinear processes are only allowed in media, which lack inversion symmetry. On the other hand, odd order processes are in principle allowed in any medium. For that reason, even order processes are rarely used in optical spectroscopy, because they cannot be applied to every bulk material. Nonlinear optical spectroscopy usually bases on odd order (basically third order) optical effects. On the other hand, second order processes are often applied for frequency conversion processes such as SHG, SFG or DFG, utilizing a couple of selected nonlinear materials which have the necessary nonzero components of the second order susceptibility tensor. Prominent examples are potassium dihydrogen phosphate KH_2PO_4 (KDP), ammonium dihydrogen phosphate $NH_4H_2PO_4$ (ADP), or lithium niobate $LiNbO_3$.

There is an important exclusion from this rule: At the interface between two materials, inversion symmetry is always destroyed, although both of the particular materials may be centrosymmetric. Therefore, second order processes may be used for interface spectroscopy. Being applied to the surface or interfaces between centrosymmetric materials, their advantage is to supply a background-free second order optical response of the interface region. Combined with local electric field strength enhancement mechanisms (for example in propagating surface plasmon arrangements at the metal-dielectric interface, compare Sect. 6.4.2), second order processes supply highly interface sensitive spectroscopic tools for interface and ultrathin adsorbate layer spectroscopy.

In finishing this Section, let us shortly discuss a further second order optical effect, which is of practical importance for light modulation purposes. It is the so-called linear electrooptic effect or Pockel's effect. Imagine a second order material, externally excited by a monochromatic field and a static one (E_s). The field is thus given by:

$$E = \frac{E_0}{2}\,\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.} + E_s$$

When calculating now the second order polarization, it will contain a term that oscillates with the frequency ω. It is given by:

$$P^{(2)}\Big|_{\omega} = \varepsilon_0 E_0 E_s \chi^{(2)}\left(\omega = \omega + 0\right)\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.}$$

Of course, the linear polarization will also contain terms that are oscillating with ω according to:

$$P^{(1)}\Big|_{\omega} = \frac{1}{2}\varepsilon_0 E_0 \chi^{(1)}\left(\omega\right)\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.}$$

So that the full polarization at ω is given by (neglecting higher order polarization terms):

$$P|_{\omega} = P^{(1)}\Big|_{\omega} + P^{(2)}\Big|_{\omega} = \varepsilon_0\left[\chi^{(1)}\left(\omega\right) + 2\chi^{(2)}\left(\omega = \omega + 0\right)E_s\right]\frac{E_0}{2}\,\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.} \tag{13.8}$$

Equation (13.8) is completely analogous to a linear material equation, when regarding the term in parentheses as an effective susceptibility, which depends on the strength of the static field as a parameter. Hence, we may define:

$$\chi^{(eff)}\left(\omega\right) \equiv \chi^{(1)}\left(\omega\right) + 2\chi^{(2)}\left(\omega = \omega + 0\right)E_s \tag{13.9}$$

An electromagnetic wave with the frequency ω will propagate in such a nonlinear medium in an identical manner as it would propagate in a linear one, when the linear susceptibility has the value as prescribed by (13.9). We may therefore define an effective refractive index analogously to the treatment in linear optics:

$$\begin{aligned} n^{(\mathrm{eff})}\left(\omega\right) &= \sqrt{\varepsilon^{(eff)}\left(\omega\right)} \equiv \sqrt{1 + \chi^{(eff)}\left(\omega\right)} \\ &= \sqrt{1 + \chi^{(1)}\left(\omega\right) + 2\chi^{(2)}\left(\omega = \omega + 0\right)E_s} = n^{(eff)}\left(\omega, E_s\right) \end{aligned} \tag{13.10}$$

According to (13.10), the value of the effective refractive index may be controlled by the strength of the static electric field. Hence, the propagation properties of the wave with frequency ω may be controlled by the static field. The name 'linear electrooptical effect' arises from the linear dependence of the effective susceptibility on the field strength. In real world, Pockel's effect leads to the appearance of a field-induced birefringence, or alters the already existent birefringence.

The mentioned effect is practically applied in so-called Pockel's cells. They may be used to modulate the resonator properties of lasers in order to produce short laser pulses. By the way, they offer a way to accomplish the mode

locking process discussed in problem 9 of Chap. 12: When modulating the cavity losses of a laser with the repetition time $2L/c$, we generate sidebands to the existing laser modes that exactly correspond to longitudinal resonator modes. They are in rigid mutual phase relations because they have been created by the same modulation mechanism. So we obtain the situation which has been discussed in problem 9 of Chap. 12: The laser starts to produce short laser pulses. This modulation of the cavity losses may be accomplished by the mentioned Pockel's cells. The corresponding mode locking mechanism is called active mode locking.

13.1.3 Some Third Order Effects

In complete analogy to the second order polarization, one may discuss the third order term in (13.3) in order to investigate optical effects that arise as a result of third order nonlinearity in different media. It is particularly important for any spectroscopist to become familiar with third order nonlinear effects, because the latter may be observed in any medium regardless on the concrete symmetry that is subject to the given sample. In the context of our discussion of thin film optical properties, we will concentrate on a few effects which might be helpful to understand the behaviour of optical materials when illuminated with laser light of high intensities.

Assuming an oscillating electric field according to:

$$E = \frac{E_0}{2}\,\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.}$$

and proceeding the same way as in the previous section, we obtain a third order polarization term, which oscillates with the frequency 3ω:

$$P^{(3)} = \ldots + \varepsilon_0 \chi^{(3)}\,(3\omega = \omega + \omega + \omega)\,\frac{E_0^3}{8}\,\mathrm{e}^{-3\mathrm{i}\omega t} + \text{c.c.} + \ldots \tag{13.11}$$

Consequently, in a third-order nonlinear medium, an electromagnetic wave will be generated that has the frequency 3ω. The corresponding process is called *Third Harmonic Generation* THG. Nevertheless, in practice third harmonic generation is usually not accomplished by means of a third order nonlinear frequency conversion. Instead, it turns out to be more efficient to use a cascade of second order processes to generate higher order harmonics of a given ground frequency. Thus, the third harmonic may be generated by SHG of the ground frequency, followed by an SFG process between the second harmonic and the ground frequency.

But let us return to (13.11), and in particular to terms which do not oscillate at the THG frequency. Indeed, as a result of the third order frequency mixing, we obtain a third order polarization that oscillates at the ground frequency ω. It is given by:

$$P^{(3)}\Big|_{\omega} = \frac{3}{8}\varepsilon_0 \chi^{(3)}\,(\omega = \omega + \omega - \omega)\,E_0^2 E_0^*\,\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.}$$

Of course, the linear polarization also yields a contribution at the ground frequency. So that the full polarization at the ground frequency will be given by:

$$\begin{aligned} P|_{\omega} &= P^{(1)}\Big|_{\omega} + P^{(3)}\Big|_{\omega} = \\ &= \frac{1}{2}\varepsilon_0 \left(\chi^{(1)}(\omega) + \frac{3}{4}\chi^{(3)}(\omega = \omega + \omega - \omega)\,|E_0|^2\right) E_0\,\mathrm{e}^{-\mathrm{i}\omega t} + \text{c.c.} \end{aligned} \tag{13.12}$$

We are now in a similar situation as in the previous section, when we discussed the linear electrooptical effect ((13.8)). Equation (13.12) is again equivalent to a linear material equation with an effective susceptibility, the latter depending on the intensity of the incoming light. The effective susceptibility is now given by:

$$\chi^{(\text{eff})}(\omega) = \chi^{(1)}(\omega) + \frac{3}{4}\chi^{(3)}(\omega = \omega + \omega - \omega)\,|E_0|^2 \tag{13.13}$$

It depends on the square of the field strength and represents a special version of the optical Kerr-effect. The square root of the effective dielectric function becomes:

$$\begin{aligned} \sqrt{\varepsilon^{(\text{eff})}(\omega)} &= \sqrt{1 + \chi^{(1)}(\omega) + \frac{3}{4}\chi^{(3)}(\omega = \omega + \omega - \omega)\,|E_0|^2} \\ &= \sqrt{\varepsilon}\sqrt{1 + \frac{3}{4}\frac{\chi^{(3)}(\omega = \omega + \omega - \omega)}{\varepsilon}|E_0|^2} \end{aligned}$$

where, as usual, $\varepsilon = 1 + \chi^{(1)}$.

Let us now assume, that the linear dielectric function is purely real at the given frequency. In linear optics, the medium would then be free of absorption, and the usual refractive index would be given by the square root of the dielectric function. The effective index of refraction is now given by:

$$\hat{n}^{(\text{eff})}(\omega) = n\sqrt{1 + \frac{3}{4}\frac{\chi^{(3)}(\omega = \omega + \omega - \omega)}{n^2}|E_0|^2}$$

If the nonlinear contribution is small compared to the linear one, this relationship may be rewritten as:

$$\hat{n}^{(\text{eff})}(\omega) \approx n + \frac{3}{8}\frac{\chi^{(3)}(\omega = \omega + \omega - \omega)}{n}|E_0|^2 \tag{13.14}$$

Although the linear dielectric function has been assumed to be real, the effective index of refraction may be complex, depending on the properties of the third order susceptibility at the given frequency. In any case, we obtain an intensity-dependent effective refractive index according to:

$$n^{(\text{eff})}(\omega) \approx n + \frac{3}{8}\frac{\mathrm{Re}\chi^{(3)}(\omega = \omega + \omega - \omega)}{n}|E_0|^2 \equiv n + n_2|E_0|^2 \tag{13.15}$$

It turns out to be dependent on the intensity of the electromagnetic wave. The value n_2 is called the *nonlinear refractive index* of the medium. The intensity-dependence of the refractive index in a third order (or even higher odd order) nonlinear optical medium is responsible for different self-interaction processes of highly intense light beams, such as self-focusing of laser beams, or self-phase-modulation processes in ultrashort light pulses, essential for white light continuum generation.

Let us now look at the imaginary part. From (13.14), it follows immediately that there is a nonlinear absorption coefficient given by:

$$\alpha_{nl}(\omega) \approx \frac{3}{4}\frac{\omega}{cn}\mathrm{Im}\chi^{(3)}(\omega = \omega + \omega - \omega)\,|E_0|^2 \tag{13.16}$$

With increasing intensity of the light, the medium may become absorbing. As it will become clear from the semiclassical expressions for the nonlinear susceptibilities, the nonlinear absorption as described by (13.16) results from two-photon absorption processes.

The mentioned effects give a short survey to the reader which kind of new optical effects may be expected in the field of nonlinear optics. We finish the formal treatment of nonlinear optical effects at this stage and turn to the semiclassical calculation of susceptibilities. The primary purpose of the following section is not to develop high mathematics for susceptibility calculations, but to reveal a picture on the quantum mechanical processes behind the scene of the formally introduced susceptibilities.

13.2 Calculation Scheme for Nonlinear Optical Susceptibilities

13.2.1 Macroscopic Susceptibilities and Microscopic Hyperpolarizabilities

Before starting with quantum mechanical calculations of the nonlinear optical response of matter, we have to perform the same formal work as we performed in the case of linear optics. The problem is, that quantum mechanical calculations as presented in Chap. 11 often concern the calculation of the microscopic dipole moment of an accessible quantum system, which might be a molecule or an atom. On the other hand, the susceptibilities have been introduced through the macroscopic polarization vector by (13.1):

$$\begin{aligned}\boldsymbol{D} &= \varepsilon_0\boldsymbol{E} + \boldsymbol{P} = \varepsilon_0\boldsymbol{E} + \boldsymbol{P}^{(1)} + \boldsymbol{P}^{(2)} + \boldsymbol{P}^{(3)} + \ldots \\ &= \varepsilon_0\Big\{\boldsymbol{E} + \chi^{(1)}\boldsymbol{E} + \chi^{(2)} :\ \boldsymbol{EE} + \chi^{(3)} :\ \boldsymbol{EEE} + \ldots\Big\}\end{aligned}$$

In complete analogy to linear optics, the corresponding microscopic material equation may be formulated as:

$$\begin{aligned}\boldsymbol{p} &= \boldsymbol{p}^{(1)} + \boldsymbol{p}^{(2)} + \boldsymbol{p}^{(3)} + \dots \\ &= \varepsilon_0 \Big\{ \beta^{(1)} \boldsymbol{E}_{\mathrm{micr}} + \beta^{(2)} : \boldsymbol{E}_{\mathrm{micr}} \boldsymbol{E}_{\mathrm{micr}} + \beta^{(3)} : \boldsymbol{E}_{\mathrm{micr}} \boldsymbol{E}_{\mathrm{micr}} \boldsymbol{E}_{\mathrm{micr}} + \dots \Big\}\end{aligned} \tag{13.17}$$

Equation (13.17) describes the microscopic dipole moment in terms of an expansion into a power series of the local or microscopic electric field. The expectation values for the linear and nonlinear contributions to the dipole moment may be calculated in analogy to what has been done in Chap. 11. The proportionality coefficient $\beta^{(1)}$ is nothing else than the usual linear polarizability as introduced in Chap. 3. The values $\beta^{(j>1)}$ are the so-called higher order polarizabilities (or nonlinear polarizabilities, or hyperpolarizabilities) of the microscopic oscillator. They may be determined immediately, when the dipole moment has been calculated. On the other hand, in order to determine the susceptibility, it will be necessary to establish a theoretical relationship between the hyperpolarizabilities and the macroscopic nonlinear susceptibilities which enter into (13.1). In linear optics, this relationship has been supplied by the Clausius–Mossotti-Equation. So what we need now is a nonlinear version of this equation.

We will present here only the final result, valid for spherical symmetry as it was in the case of the Clausius–Mossotti-Equation. The derivation may be performed by the reader himself (compare Sect. 13.3.2). As the result, we obtain:

$$\chi^{(j)} \left(\omega_{j+1} = \sum_{l=1}^{j} \omega_l \right) = N \prod_{l=1}^{j+1} \left[\frac{\varepsilon(\omega_l) + 2}{3} \right] \beta^{(j)} \left(\omega_{j+1} = \sum_{l=1}^{j} \omega_l \right) \tag{13.18}$$

where N is again the concentration of the microscopic dipoles. The local field correction (13.18) has to be applied with the same caution as in linear optics, in particular, the recommendations from Table 3.2 apply as well.

13.2.2 Density Matrix Approach for Calculating Optical Hyperpolarizabilities

In analogy to Sect. 11.2.3, let us write down the equation for the non-diagonal elements of the density matrix in the interaction picture according to (electric dipole interaction only):

$$\frac{\partial}{\partial t} \rho_{mn} + \frac{\rho_{mn}}{T_{2,mn}} = \frac{\mathrm{i}}{\hbar} [\boldsymbol{p}, \boldsymbol{\rho}]_{mn} \boldsymbol{E} \tag{13.19}$$

The non-diagonal elements of the density matrix are essential for the calculation of the dipole moment because of (11.22):

$$\langle \boldsymbol{p} \rangle = Tr(\boldsymbol{p}\boldsymbol{\rho}) = \sum_n \sum_m \boldsymbol{p}_{nm} \rho_{mn} \tag{13.20}$$

In (13.20) and in the following, the bold writing of $\boldsymbol{p}$ only indicates its vector character. Again, we assume

$$\boldsymbol{p}_{nn} = \mathbf{0} \; \forall \, n \, .$$

Equation (13.19) may be rewritten as:

$$\frac{\partial}{\partial t}\rho_{mn} + \frac{\rho_{mn}}{T_{2,mn}} = \frac{\mathrm{i}}{\hbar}\sum_{l} (\boldsymbol{p}_{ml}\rho_{ln} - \rho_{ml}\boldsymbol{p}_{ln})\boldsymbol{E} \tag{13.21}$$

In the absence of any electric field, the stationary solutions of (13.21) are identical to zero. We mark these solutions by a superscript (0). Hence, we have:

$$\rho_{mn}^{(0)} = 0$$

In the absence of any perturbing field, the diagonal elements of the density matrix are expected to approach their equilibrium values. According to (11.21), we suppose:

$$\rho_{nn}^{(0)} = \frac{\mathrm{e}^{-\frac{E_n}{k_\mathrm{B}T}}}{\sum\limits_{n} \mathrm{e}^{-\frac{E_n}{k_\mathrm{B}T}}}$$

This expression is at least applicable for gases, or molecular liquids and solids. Our subsequent derivation will deal with this case. For solids whose electronic properties are dominated by the band structure, the equilibrium values of the density matrix will be given by the Fermi-Dirac distribution.

The appearance of the electric field will obviously alter the elements of the density matrix. We therefore assume:

$$\rho = \rho\,(\boldsymbol{E}) = \rho^{(0)} + \rho^{(1)} + \rho^{(2)} + \rho^{(3)} + ... \tag{13.22}$$

with

$$\rho^{(0)} \propto \boldsymbol{E}^0 \, ; \; \rho^{(1)} \propto \boldsymbol{E}^1 \, ; \; \rho^{(2)} \propto \boldsymbol{E}^2 \, ; \; \rho^{(3)} \propto \boldsymbol{E}^3 \tag{13.23}$$

and so on.

Equations (13.22) and (13.23) in fact represent a perturbation theory approach and require fast convergence of the series. Some criteria will be given later. In order to simplify our treatment, let us regard conditions, which are either nonresonant or, if resonance occurs, do not lead to remarkable changes in the population of the individual energy levels due to sufficiently weak excitation. In this case, we can make the simplifying assumption on the diagonal elements of the density matrix, that

$$\rho_{nn}^{(j>0)} = 0 \; \forall \, n \; \Rightarrow \; \rho_{nn} = \rho_{nn}^{(0)} \tag{13.24}$$

Substituting ρ_{mn} in (13.21) by (13.22), and collecting the terms which belong to the same power in the field strength, we obtain the system of equations:

$$\begin{aligned}
&\frac{\partial}{\partial t}\rho_{mn}^{(j+1)} + \frac{\rho_{mn}^{(j+1)}}{T_{2,mn}} = \frac{\mathrm{i}}{\hbar}\sum_{l}\left(\boldsymbol{p}_{ml}\rho_{ln}^{(j)} - \rho_{ml}^{(j)}\boldsymbol{p}_{ln}\right)\boldsymbol{E}; \quad j = 0, 1, 2, \ldots \\
&\Rightarrow \frac{\partial}{\partial t}\rho_{mn}^{(j+1)} + \frac{\rho_{mn}^{(j+1)}}{T_{2,mn}} = \\
&= \frac{\mathrm{i}}{\hbar}\left[\boldsymbol{p}_{mn}\left(\rho_{nn}^{(j)} - \rho_{mm}^{(j)}\right) + \sum_{l\neq n}\left(\boldsymbol{p}_{ml}\rho_{ln}^{(j)}\right) - \sum_{l\neq m}\left(\rho_{ml}^{(j)}\boldsymbol{p}_{ln}\right)\right]\boldsymbol{E}
\end{aligned} \tag{13.25}$$

From (13.25) and (13.24), we immediately find the different equations for the first and higher order contributions to the density matrix:

$$\begin{aligned}
&\text{linear case:} \quad \frac{\partial}{\partial t}\rho_{mn}^{(1)} + \frac{\rho_{mn}^{(1)}}{T_{2,mn}} = \frac{\mathrm{i}}{\hbar}\left(\rho_{nn}^{(0)} - \rho_{mm}^{(0)}\right)\boldsymbol{p}_{mn}\boldsymbol{E}\,; \\
&\text{nonlinear case:} \quad \frac{\partial}{\partial t}\rho_{mn}^{(j+1)} + \frac{\rho_{mn}^{(j+1)}}{T_{2,mn}} = \frac{\mathrm{i}}{\hbar}\left[\sum_{l\neq n}\left(\boldsymbol{p}_{ml}\rho_{ln}^{(j)}\right) - \sum_{l\neq m}\left(\rho_{ml}^{(j)}\boldsymbol{p}_{ln}\right)\right]\boldsymbol{E}
\end{aligned} \tag{13.26}$$

According to (13.5), we write the microscopic electric field as:

$$\boldsymbol{E} = \frac{1}{2}\sum_{q}\boldsymbol{E}_q \mathrm{e}^{-\mathrm{i}\omega_q t} + \text{c.c.} \tag{13.27}$$

Let us solve equations (13.26) for the first and higher order perturbation cases. The first-order equation becomes:

$$\frac{\partial}{\partial t}\rho_{mn}^{(1)} + \frac{\rho_{mn}^{(1)}}{T_{2,mn}} = \frac{\mathrm{i}}{2\hbar}\left(\rho_{nn}^{(0)} - \rho_{mm}^{(0)}\right)\boldsymbol{p}_{mn}\left(\sum_{q}\boldsymbol{E}_q \mathrm{e}^{-\mathrm{i}\omega_q t} + \text{c.c.}\right) \tag{13.28}$$

We must now remember, that in the interaction picture, the matrix elements of the dipole operator $\boldsymbol{p}_{mn}$ carry a time-dependence according to (11.28). Therefore, we use the same kind of approach as in Sect. 11.2.3:

$$\begin{aligned}
\rho_{mn}^{(1)} &= \sum_{q}\left(\rho_{mn,q}^{-} + \rho_{mn,q}^{+}\right) \\
\rho_{mn,q}^{(-)} &= P_{mn,q}^{(-)}\mathrm{e}^{\mathrm{i}(\omega_{mn}-\omega_q)\,t} \\
\rho_{mn,q}^{(+)} &= P_{mn,q}^{(+)}\mathrm{e}^{\mathrm{i}(\omega_{mn}+\omega_q)\,t}
\end{aligned} \tag{13.29}$$

where the P-values shall not depend on time. (13.28) and (13.29) together yield:

$$\rho_{mn}^{(1)} = \frac{1}{2\hbar}\left(\rho_{nn}^{(0)} - \rho_{mm}^{(0)}\right)\boldsymbol{p}_{mn}\sum_q\left(\frac{\boldsymbol{E}_q \mathrm{e}^{-\mathrm{i}\omega_q t}}{\omega_{mn} - \omega_q - \mathrm{i}\Gamma_{mn}} + \frac{\boldsymbol{E}_q^* \mathrm{e}^{\mathrm{i}\omega_q t}}{\omega_{mn} + \omega_q - \mathrm{i}\Gamma_{mn}}\right) \tag{13.30}$$

From (13.30) and (13.20), one could now calculate the first order dipole moment. That would lead to an expression for the components of the first order polarizability tensor. In the case of optical isotropy, (11.31) would follow from there as a special case.

However, our purpose is to find expressions for the nonlinear polarizabilities. This may be done combining (13.20) and the equation for the higher order contributions to the density matrix (13.26). For the second order, we find:

$$\frac{\partial}{\partial t}\rho_{mn}^{(2)} + \frac{\rho_{mn}^{(2)}}{T_{2,mn}} = \frac{\mathrm{i}}{\hbar}\left[\sum_{l\neq n}\left(\boldsymbol{p}_{ml}\rho_{ln}^{(1)}\right) - \sum_{l\neq m}\left(\rho_{ml}^{(1)}\boldsymbol{p}_{ln}\right)\right]\boldsymbol{E} \tag{13.31}$$

The first order density matrix terms, which enter into (13.31), are now given by (13.30). For the electric field, we have (13.27). It is obvious that we become confronted with an extensive and tedious derivation, although the general strategy of the calculation is simple. Let us therefore concentrate on the general structure of the expressions which will be obtained.

Substituting the first order expressions in (13.31) by (13.30), and $\boldsymbol{E}$ by (13.27), we find:

$$\begin{aligned}
&\frac{\partial}{\partial t}\rho_{mn}^{(2)} + \rho_{mn}^{(2)}\Gamma_{mn} = \frac{\mathrm{i}}{2\hbar^2} \\
&\times\left\{\sum_{q'}\sum_{l}\left[\frac{\boldsymbol{p}_{ml}\left(\boldsymbol{p}_{ln}\boldsymbol{E}_{q'}\right)\mathrm{e}^{-\mathrm{i}\omega_{q'}t}}{\omega_{ln} - \omega_{q'} - \mathrm{i}\Gamma_{ln}}\left(\rho_{nn}^{(0)} - \rho_{ll}^{(0)}\right) + 3 \text{ more terms}\right]\right\} \\
&\times\left(\frac{1}{2}\sum_q \boldsymbol{E}_q \mathrm{e}^{-\mathrm{i}\omega_q t} + \text{c.c.}\right)
\end{aligned} \tag{13.32}$$

We will not explicitly write out all terms, they follow immediately from algebra. Again, the matrix elements of the dipole operator carry the relevant time dependence. It is given by:

$$p_{ml}p_{ln} \propto \mathrm{e}^{\mathrm{i}(\omega_{ml}+\omega_{ln})\,t} = \mathrm{e}^{\frac{\mathrm{i}}{\hbar}(E_m - E_l + E_l - E_n)t} = \mathrm{e}^{\mathrm{i}\omega_{mn}t}$$

Therefore, for the density matrix, we assume the time dependence:

$$\rho_{mn}^{(2)} \propto \mathrm{e}^{\mathrm{i}\left(\omega_{mn} \mp \omega_{q'} \mp \omega_q\right)t}$$

Substituting this into (13.32), we find the density matrix to be proportional to:

$$\rho_{mn}^{(2)} \propto \frac{1}{\hbar^2} \times \tag{13.33}$$
$$\times \sum_l \sum_{q'} \sum_q \left[\frac{(\boldsymbol{p}_{ml}\boldsymbol{E}_q)(\boldsymbol{p}_{ln}\boldsymbol{E}_{q'}) \left(\rho_{nn}^{(0)} - \rho_{ll}^{(0)}\right) \mathrm{e}^{-\mathrm{i}\left(\omega_{q'}+\omega_q\right)t}}{(\omega_{mn} - \omega_q - \omega_{q'} - \mathrm{i}\Gamma_{mn})(\omega_{ln} - \omega_{q'} - \mathrm{i}\Gamma_{ln})} + 7 \text{ more terms} \right]$$

Utilizing

$$\left\langle \boldsymbol{p}^{(2)} \right\rangle = \sum_n \sum_m \boldsymbol{p}_{nm} \rho_{mn}^{(2)}$$

we find the second order polarization (for example at the sum frequency):

$$\left\langle \boldsymbol{p}^{(2)}(\omega_q + \omega_{q'}) \right\rangle \tag{13.34}$$
$$\propto \frac{1}{\hbar^2} \sum_n \sum_m \sum_l \left[\frac{\boldsymbol{p}_{nm}(\boldsymbol{p}_{ml}\boldsymbol{E}_q)(\boldsymbol{p}_{ln}\boldsymbol{E}_{q'}) \left(\rho_{nn}^{(0)} - \rho_{ll}^{(0)}\right) \mathrm{e}^{-\mathrm{i}\left(\omega_{q'}+\omega_q\right)t}}{(\omega_{mn} - \omega_q - \omega_{q'} - \mathrm{i}\Gamma_{mn})(\omega_{ln} - \omega_{q'} - \mathrm{i}\Gamma_{ln})} + \dots \right]$$

From (13.34) and (13.17) it becomes clear, that the expression for the second order polarizability has the mathematical structure like:

$$\beta_{abc}^{(2)}(\omega = \omega_q + \omega_{q'}) \tag{13.35}$$
$$\propto \frac{1}{\varepsilon_0 \hbar^2} \sum_n \sum_m \sum_l \left[\frac{p_{nm,a} p_{ml,b} p_{ln,c} \left(\rho_{nn}^{(0)} - \rho_{ll}^{(0)}\right)}{(\omega_{mn} - \omega_q - \omega_{q'} - \mathrm{i}\Gamma_{mn})(\omega_{ln} - \omega_{q'} - \mathrm{i}\Gamma_{ln})} + \dots \right]$$

where a, b and c stand for the relevant Cartesian coordinates.

Equation (13.35) reveals the general structure of the expression for the second order hyperpolarizability. Practically, it differs from the linear expression by an additional pre-factor of the type:

$$\frac{p_{nm}}{\hbar(\omega_{mn} - \omega_q - \omega_{q'} - \mathrm{i}\Gamma_{mn})} \tag{13.36}$$

We will discuss this expression a bit later, namely in the next section. Let us now shortly look at the third order polarizability.

The philosophy of calculating is again the same. We have to start from the calculation of the third order contribution to the density matrix. The recurrent equation (13.26) yields:

$$\frac{\partial}{\partial t}\rho_{mn}^{(3)} + \frac{\rho_{mn}^{(3)}}{T_{2,mn}} = \frac{\mathrm{i}}{\hbar} \left[\sum_{l \neq n} \left(\boldsymbol{p}_{ml}\rho_{ln}^{(2)}\right) - \sum_{l \neq m} \left(\rho_{ml}^{(2)}\boldsymbol{p}_{ln}\right) \right] \boldsymbol{E}$$

Having calculated the second order contribution to the density matrix (13.33), the third order density matrix contribution may be calculated in complete

analogy. Then, we calculate the third order dipole moment in the usual manner. From there, comparing with (13.17), we obtain the third order hyperpolarizability.

Again, we will only present the general mathematical structure. In analogy to the previous calculations, in the general structure of the expression there appears an additional pre-factor of the type (13.36). So it will finally be obtained:

$$\beta^{(3)}_{abcd}\left(\omega = \omega_q + \omega_{q'} + \omega_{q''}\right) \propto \frac{1}{\varepsilon_0 \hbar^3} \sum_n \sum_m \sum_l \sum_k \times \tag{13.37}$$
$$\times \left[\frac{p_{nm,a} p_{ml,b} p_{lk,c} p_{kn,d}}{\left(\omega_{mn} - \omega_q - \omega_{q'} - \omega_{q''} - \mathrm{i}\Gamma_{mn}\right)\left(\omega_{ln} - \omega_{q'} - \omega_{q''} - \mathrm{i}\Gamma_{ln}\right)\left(\omega_{kn} - \omega_{q''} - \mathrm{i}\Gamma_{kn}\right)} + \ldots \right]$$

Here, a, b, c and d stand for the relevant Cartesian coordinates. In the same way, higher order hyperpolarizabilities may be calculated.

13.2.3 Discussion

Convergence

Let us start our discussion on the general properties of the nonlinear polarizabilities with some considerations on the convergence of the series (13.17). According to (13.36), fast convergence is achieved when the condition:

$$\left| \frac{|p_{nm} E_q|}{\hbar \left(\omega_{mn} - \sum\limits_q \omega_q - \mathrm{i}\Gamma_{mn} \right)} \right| \ll 1 \tag{13.38}$$

is fulfilled. Let us firstly regard a non-resonant case, when $\omega_{mn} \gg \sum\limits_q \omega_q$ holds. In this case, condition (13.38) leads to

$$\left| \frac{p_{nm} E_q}{\hbar} \right| << \omega_{mn} \tag{13.39}$$

In the non-resonant case, condition (13.39) guarantees that non-linear contributions to the polarization of the medium are small, and higher-order nonlinearities may surely be neglected. It is obvious, that the term

$$\left| \frac{p_{nm} E_q}{\hbar} \right|$$

must have the dimension of a frequency. In fact, it is nothing else than the so-called Rabi-frequency $\Omega (\Omega \equiv |p_{nm} E_q / \hbar|)$, which plays an important role in the theory of coherent optical spectroscopy.

The relation (13.39) is equivalent to the requirement, that the strength of the impinging electric field is small compared to the atomic electric fields.

In resonant conditions, from (13.38) one finds:

$$\left|\frac{p_{nm}E_q}{\hbar}\right| \ll \Gamma_{mn} \equiv T_{2,mn}^{-1} \tag{13.40}$$

This equation will be fulfilled, when relaxation processes in the medium are sufficiently fast to instantaneously destroy the polarization induced by the resonant electric field.

Selection Rules

Let us now come to the selection rules. According to the derivations from Sect. 13.2.2, the quantum mechanical selection rules for the different-order processes are determined by the following requirements:

$$\begin{aligned}
\beta_{ab}^{(1)} \neq 0 &\Leftrightarrow p_{nm,a}p_{mn,b} \neq 0 \\
\beta_{abc}^{(2)} \neq 0 &\Leftrightarrow p_{nm,a}p_{ml,b}p_{ln,c} \neq 0 \\
\beta_{abcd}^{(3)} \neq 0 &\Leftrightarrow p_{nm,a}p_{ml,b}p_{lk,c}p_{kn,d} \neq 0
\end{aligned} \tag{13.41}$$

and so on. This looks like a simple generalization of the familiar dipole interaction selection rules formulated, for example, in (10.22) or (11.24). In particular, (13.41) leads us to the conclusion, that in a quantum system with inversion symmetry, all even-order optical polarizabilities are zero in the dipole approximation: Indeed, let us look at the second order polarizability. Supposing that state $|n\rangle$ is of even parity, we must require that $|m\rangle$ is odd, otherwise p_{nm} will be zero. For the same reason, $|l\rangle$ must be even. But if so, p_{ln} will be zero. So that there is no way to arrange the quantum states in a manner that the product $p_{nm}p_{ml}p_{ln}$ becomes non-vanishing. This argumentation applies to all even-order polarizabilities, but it is not applicable to the odd-order polarizabilities. This is exemplified in Fig. 13.2 for the particular cases of linear polarization (a), SHG (b), THG (c), and Fourth-Harmonic-Generation (d). In application to macroscopic systems, it leads to the already mentioned vanishing even-order susceptibilities in systems with an inversion centre.

Resonance Behaviour

A glance at (13.37) shows us, that in nonlinear optics, principally new types of resonances may occur. Thus, the third-order susceptibility shows a resonant behaviour, when one of the frequencies of the impinging light falls close to a transition frequency of the medium. This type of resonance behaviour is well-known from linear optics. But in contrast to linear optics, there are

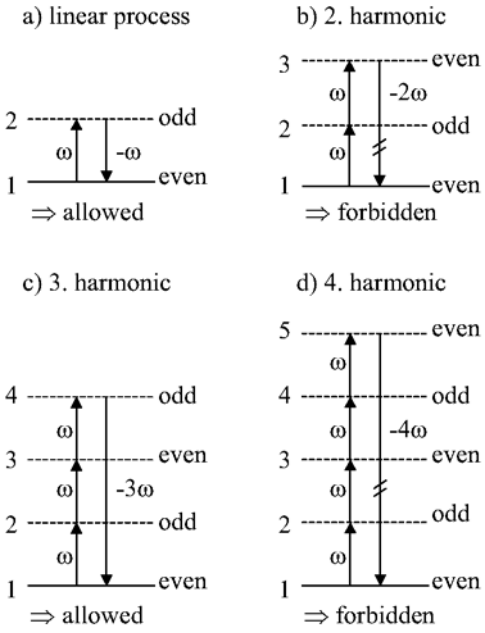

Fig. 13.2. Visualization of the effect of the parity selection rule on nonlinear optical polarizabilities: even-order processes are forbidden (right side, **b**) and **d**)), while odd-order processes are, in principle, allowed (left side, **a**) and **c**))

additional resonances, when the combination of several impinging frequencies corresponds to a transition frequency. These so-called multiphoton resonances are a characteristic feature of nonlinear optics.

In resonance conditions, the third order polarizability (and consequently the corresponding susceptibility) have a significant imaginary part. As we know from Sect. 13.1.3, this may lead to absorption processes with a nonlinear absorption coefficient, which depends on the intensity of the impinging irradiation. Let us therefore as an example regard the nonlinear absorption caused by the imaginary part of the third order susceptibility.

Nonlinear Absorption Coefficient

In Sect. 13.1.3, we supposed a medium with third-order optical nonlinearity, excited by a monochromatic light wave with the angular frequency ω. Moreover, we supposed that the linear susceptibility is purely real at the given frequency. According to (13.30), no single-photon resonances should therefore occur between populated energy levels. In this case, we found a nonlinear absorption coefficient given by (13.16):

$$\alpha_{nl}(\omega) \approx \frac{3}{4}\frac{\omega}{cn}\mathrm{Im}\chi^{(3)}(\omega = \omega + \omega - \omega)\,|E_0|^2$$

According to (13.18), in the given circumstances the imaginary part of the third order nonlinear susceptibility can only be different from zero when the third order hyperpolarizability has a non-vanishing imaginary part. It is given by (compare (13.37)):

$$\beta^{(3)}_{abcd}\left(\omega = \omega + \omega - \omega\right) \propto \frac{1}{\varepsilon_0 \hbar^2} \sum_n \sum_m \sum_l \sum_k \times \\ \times \left[\frac{p_{nm,a} p_{ml,b} p_{lk,c} p_{kn,d}}{\left(\omega_{mn} - \omega - \mathrm{i}\Gamma_{mn}\right)\left(\omega_{ln} - 2\omega - \mathrm{i}\Gamma_{ln}\right)\left(\omega_{kn} - \omega - \mathrm{i}\Gamma_{kn}\right)} + \dots \right] \tag{13.42}$$

Because single photon resonances have been excluded, the only possibility to come into resonance is to require:

$$2\omega \to \omega_{ln}$$

This condition describes a two-photon resonance, and the corresponding absorption process is called two-photon absorption. It is a process, where the energy of two photons is used instantaneously to bridge the energy gap between the levels n and l. In a somewhat naïve but simple picture, one may imagine that two photons have to "arrive" at the same time, so that the system can utilize their energy instantaneously for an absorption process. The more photons arrive in a certain time interval, the more probable the process becomes. For that reason, the two-photon absorption coefficient depends on the intensity of the impinging light.

Let us re-arrange expression (13.42). We may write:

$$\beta^{(3)}_{abcd}\left(\omega = \omega + \omega - \omega\right) \\ \propto \frac{1}{\varepsilon_0 \hbar^2} \sum_n \sum_l \left\{ \frac{\left[\sum_m \frac{p_{nm,a} p_{ml,b}}{\left(\omega_{mn} - \omega - \mathrm{i}\Gamma_{mn}\right)}\right]\left[\sum_k \frac{p_{lk,c} p_{kn,d}}{\left(\omega_{kn} - \omega - \mathrm{i}\Gamma_{kn}\right)}\right]}{\left(\omega_{ln} - 2\omega - \mathrm{i}\Gamma_{ln}\right)} + \dots \right\}$$

From here we see, that the transition rates for a two-photon absorption process between the levels n and l will not be determined by p_{nl} as in the case of simple absorption, but rather by conglomerates of the type:

$$\left[\sum_k \frac{p_{lk,c} p_{kn,d}}{\left(\omega_{kn} - \omega - \mathrm{i}\Gamma_{kn}\right)} \right] \tag{13.43}$$

Correspondingly, the selection rule would be, that the term (13.43) is different from zero.

By the way, similar expressions hold for the probability of Raman-scattering.

Having shortly discussed the two-photon absorption as one prominent example of multiphoton processes, let us formulate two main conclusions which apply for other multi-photon processes as well. Compared to single photon processes, multiphoton processes

- are subject to *other selection rules* (may be efficient in situations, where single photon processes do not work),
- are subject to *other resonance conditions* (do not require the incident frequency to be tuned to the transition frequency). Hence, there is more freedom in choosing the light source.

Therefore, they are widely used in optical spectroscopy today.

13.3 Resume from Chapter 13

13.3.1 Overview on Main Results

So far, this chapter has dealt with some properties of nonlinear optical susceptibilities, essential for the description of nonlinear optical processes in different media. As this is a book on thin film optics, we will not go into further detail in this field. Nevertheless, some fundamental understanding on nonlinear optics is essential even for a thin film researcher, because in the case of laser applications, nonlinear optical processes may be essential to understand the performance of a coating. For that reason, a simple treatment of nonlinear processes such as the optical Kerr-effect and nonlinear absorption processes have been included into this book.

In fact, we have already dealt with other non-linear optical processes throughout this book. Thus, the simple discussion of the saturation of an optical transition (Sect. 10.7.1) made us familiar with another nonlinear optical process: At high light intensities, resonant excitation may alter the population of the participating quantum states. Then, assumption (13.24) is no more valid, and the intensity-dependence of the diagonal elements of the density matrix has to be taken into account. This is comparably easy to be calculated in the case of a two-level system, but is beyond the scope of this book. We only notice that the general conclusions form Sect. 10.7.1 will remain valid.

Other sources of optical nonlinearities may, for example, arise from sample heating caused by light absorption, which alters the diagonal elements of the density matrix due to thermal population of higher energy levels.

Another remark concerns the general philosophy of our treatment as described in Sect. 2.4. In that section we remarked, that the calculation of an optical signal includes two main parts: the first part deals with the calculation of the corresponding material constants, while in the second part, one has to solve Maxwell's equations in the given experimental geometry, utilizing the previously calculated material constants. Concerning our treatment of nonlinear optics, the only thing we did so far was to give a recipe to calculate

the material constants (nonlinear optical susceptibilities). What about the solution of the wave equation (2.2), considering the existence of nonlinear polarization contributions?

A grammar treatment of nonlinear optical processes will, of course, include the solution of the corresponding wave equation. Again, we will not deal with this theory here. We only mention, that, for example, an efficient frequency conversion by SHG does not only require a large second order susceptibility. Instead, one also has to take care that the phase velocities of the electromagnetic waves at the ground and doubled frequencies are identical (phase-matching). Otherwise, no efficient energy transfer from the incident wave to the SHG signal will occur. These results will be obtained as a natural conclusion from the solution of the corresponding nonlinear wave equation. The interested reader is referred to the literature on nonlinear optics in this context.

Keeping in mind the above-mentioned remarks, we formulate the main results of the current chapter as follows:

- We have become familiar with elemental non-linear optical processes arising from second- and third order optical nonlinearities in the non-linear material equation. In particular, from the structure of the material equation, we were able to identify some utmost important nonlinear optical effects, namely: SHG, SFG, DFG, optical rectification, the Pockels effect, THG, the optical Kerr-effect, and nonlinear absorption.
- Basing on Liouville's equation for the density matrix, we developed a semiclassical perturbation approach for the calculation of nonlinear optical polarizabilities in the dipole approximation. On this basis, we were able to identify two- and higher order photon resonance processes. Additionally, some important selection rules for nonlinear optical processes could be derived.

13.3.2 Problems

1. Give a derivation of (13.18)!

   Answer: In (3.20), you will now have to consider that

   $$P = P^{\text{linear}} + P^{\text{nonlinear}}.$$

   In analogy to Sect.3.2.2, you will find then:

   $$P^{\text{linear}} = (\varepsilon - 1)\,\varepsilon_0 \left(E + \frac{P^{\text{nonlinear}}}{3\varepsilon_0} \right)$$

   That leads to:

$$D = \varepsilon_0 E + P^{\text{linear}} + P^{\text{nonlinear}}$$
$$= \varepsilon_0 \varepsilon E + \frac{\varepsilon + 2}{3} P^{\text{nonlinear}} \equiv \varepsilon_0 \varepsilon E + D^{\text{nonlinear}}$$

Assuming: $D^{\text{nonlinear}} = \varepsilon_0 \chi^{(j)} E^j$ and $P^{\text{nonlinear}} = N \varepsilon_0 \beta^{(j)} E^j_{\text{micr}}$ with $E_{\text{micr}} = \frac{\varepsilon + 2}{3} E$, we obtain (13.18).

2. Make sure that the Rabi-frequency as defined in Sect.13.2.2 has the right dimensionality!

   Answer: Yes, it has.

3. Assuming a Rydberg-like atom, give a physical interpretation of (13.39)!

   Answer: It turns out, that the electric field amplitude in the wave should be much smaller than

   $$\frac{1}{4\pi\varepsilon_0} \frac{e}{a_0^2},$$

   where a_0 is Bohr's radius, and e the elementary charge. For the derivation you should assume that $p_{nm} \approx e a_0$. That means, that nonlinear optical processes become relevant, when the field of the wave becomes comparable to the electric field strength in an atom.

14 Summary

We have now finished our excursion through fundamental aspects of the optical properties of optical thin films and film systems. At this moment, it is worth summarizing the main statements derived in this book, and to emphasize their relation to practical aspects of optical thin film spectroscopy.

First of all, let us recall what has been defined as the subject of this book. This book represents a tutorial for those readers who need a grammar mathematical apparatus for the description of thin film optical spectra. The book is definitely *not* dedicated to thin film design, and it is even not restricted to optical interference coatings. Instead, it should give support to anybody who has to judge the properties of thin films or thin film systems, basing on transmittance and/or reflectance spectra.

In this connection it is worth remembering that a scientist may be involved in thin film optics for quite different reasons. Of course, there is the broad community of researchers in the field of optical interference coatings. But knowledge on the optical properties of thin solid films may be important in other branches as well. Thin film spectroscopy may be helpful to judge the properties of any thin film, no matter whether it is to be applied for optical purposes or not. It may yield important information on the film geometry, stoichiometry, structure and so on. Questions of this type may arise in the fields of optoelectronics, semiconductor physics, or physical chemistry, to give just some examples. For that reason, in this monograph we performed a broad and detailed discussion of optical material properties (parts I, III, and IV). In contrast to books that are specialized on interference coatings, the absorption behaviour of the material has been discussed extensively. This is a must for a thin film spectroscopist, because particularly in analytical tasks, the spectral regions of considerable absorption are much more interesting than the transparency regions. So when regarding, for example, Fig. 3.5, we have to recognize that the thin film engineer, which is involved in the design of interference coatings, and the more analytical spectroscopist may work with the same material, but they will utilize disjunctive spectral regions. In interference coatings design, one will try to apply materials with as possible low absorption losses in the requested spectral region. In chemical analysis, on the contrary, one will particularly focus on absorption features to judge the structure and stoichiometry of a sample.

There are other differences as well. In optical analytics, one will often restrict on single-layer-systems, while the thin film engineer has to consider multilayer systems. Again, both aspects have been considered in this book (part II).

The large number of film interfaces in an optical interference coating, combined with low bulk absorption losses, results in specific loss mechanisms, which are often negligible in the analytics of absorbing single layer systems: In interference coatings theory, interface absorption and interface scattering loss mechanisms may be relevant. We did not deal with this matter in this book, because this is rather a special problem in the field of interference coatings.

It is finally worth noting, that reverse search procedures occur in analytical thin film spectroscopy as well as in optical coatings design. In this book, we only described a few aspects of the analytical task (Sect. 7.4.6). Let us shortly state the main differences concerning reverse searches in optical coatings design and thin film analytics.

In the analytical task, our goal would be to determine the optical constants and the film thickness of a sample from *measured* spectra (in our case $T(\nu)$ and $R(\nu)$). From these spectra, one might try to calculate the optical constants by minimizing a merit function of the type (7.27). The minimization should come to an end, when a fit is achieved within the measurement error bars. As the measurements have been performed with really existing samples, we may expect that there exists at least one solution of this minimization problem (at least one pair of functions $n(\nu)$ and $\alpha(\nu)$), which leads to a fit of the experimental spectra. As a matter of fact, one will often find a *multiplicity* of solutions. Clearly, only one of these solutions corresponds to the physically meaningful solution which is relevant for the concrete sample under consideration, and this solution has to be selected.

In thin film design, the task is to design a coating that fits certain prescribed specifications. Mathematically that leads to the same minimization procedure, with the only difference that the measured spectra have to be replaced by the required specifications (for example the filter characteristics), and the measurement error by the accepted tolerances. The multiplicity of solutions is now of use: If there are different solutions, they correspond to different designs having the same spectral response, and one may choose the design which is most conveniently to be manufactured. On the contrary, there is no guarantee that there exists a solution at all. That means it may be impossible to find a thin film design which meets the specifications - basing on the optical materials available. Despite of the current efforts to refine the mathematical design methods (the development of the so-called *needle optimization* technique is just one example), it is also a challenging task to develop new optical materials with tailored optical properties.

Let us visualize the content of this book in the light of the previous discussion. This is attempted in Table 14.1. Here, the main contents of this book are opposed to the specifics of optical interference coatings practice, and optical

Table 14.1. To the content of this monograph

	Optical interference coatings	Optical thin film spectroscopy for analytical tasks	This book
Motivation	Design a coating which meets a prescribed specification; quality control	Get information on geometry, structure, stoichiometry, band structure and so on	Tutorial, which derives the basic mathematical apparatus
Specifics of the reverse search procedure	Multiplicity of solutions is welcome; existence of the solutions is guaranteed only in special cases	Solution must exist, but there is only one physically meaningful solution	No design skills, examples on reverse search in analytics
Typical spectral region	Outside the regions of intrinsic film material absorption	Absorbing spectral regions	Derivation of equations, valid for spectral regions of significant as well as negligible absorption
Typical models for optical constants	Classical models and their derivations (Cauchy, Sellmeier)	Semiclassical or quantum mechanical models	Classical and semiclassical description
Typical sample geometry	Multilayer stack on a surface or a slab (substrate)	Surface, Slab, or single film on a surface or a slab	Surface, slab, single film on a surface or a slab; multilayer stack on a surface or a slab
Interface or surface losses	May be important	Usually negligible	Not considered

thin film analysis. This is a simplified classification, and exclusions from the mentioned features will surely occur (for example, metal interference filters contain metal films, which necessarily have some absorption). But the main message is, that this book supplies the reader with the basic theoretical apparatus applicable in interference coatings physics as well as analytical thin film spectroscopy. On the other hand, it does not deal with highly specialized questions, such as the description of interface absorptions or a strong quantum mechanical treatment of optical processes in crystals.

Having opposed the content of this book to the requirements of different research fields, let us finally review the main subjects described in this book and relate them to practical problems which are essential in optical thin film research. Again, we will prefer a table (Table 14.2) for this review. Throughout the derivations in this book, we did not always mention the concrete relation between the subject under discussion and practical problems in thin film spectroscopy. Nevertheless, after having read this book, the relation between the third and the first columns of Table 14.2. should be clear. If not, so the

Table 14.2. Practical relevance of the subjects considered in this book

Subject in this book	Section	Relation to thin film practice
Drude's formula	3.1	Free electron contribution to the linear optical constants (for example metal films, highly doped semiconductor films)
Lorentzian oscillator model	3.2	Bound electron contribution to the linear optical constants in the vicinity of a single absorption line, or lattice vibrations
Multioscillator model	4.3	Complicated absorption structures, caused by bound electrons or lattice vibrations Inhomogeneous line broadening
Sellmeier- and Cauchy formulae	4.4	Refractive index dispersion in transparency regions
Mixtures	4.5	Role of film contaminations (extrinsic absorption) Surface roughness with high spatial frequency Columnar film structure, large voids in evaporated coatings Simple treatment of optical anisotropy in terms of depolarisation factors Thermal shifts Subnanometer voids in coatings, produced by ion-assisted methods Properties of composite film materials
Kramers–Kronig-Relations	5	Sum rules for quantitative spectroscopy
Fresnel's equations	6	Interface reflections Polarization effects at oblique incidence Propagating surface plasmon polaritons at metal surfaces
T and R for thick slabs	7.1; 7.4.4	Optical properties of possibly absorbing substrates at any angle of incidence
T and R for thin films	7.2–7.4	Forward and reverse search for a single film on a surface or substrate
Mathematical treatment of refractive index gradients	8.1	Gradient index layers Rugate filters
Matrix formalism	8.2; 9.1	Multilayer coatings
Selection rules	10.4	Interpretation of optical spectra
Semiclassical treatment of the dielectric function	11; 12	Intrinsic thermal shifts Shape of the absorption edge in crystalline films (evaporation) Size effects Shape of the absorption edge in amorphous films (ion-assisted deposition)
Nonlinear susceptibilities	13	Nonlinear refraction at high laser intensities Nonlinear absorption at high laser intensities

reader is referred to literature claimed as recommended for further reading to the relevant chapter. The referring to literature has to be understood in its cumulative version – it always means: the cited reference and references cited therein. Having made these last remarks, we conclude the last chapter of this book.

Bibliography

Chapter 1

[1] Harnessing Light: *Optical Science and Engineering for the 21th Century* (COSE-Report) (National Academy Press, Washington D.C. 1998)

[2] A.H. Guenther (Ed.): *International Trends in Applied Optics* (SPIE-Press Bellingham, 2002)

[3] J.G. Webster (Ed.): *Wiley Encyclopedia of Electrical and Electronics Engineering*, Vol. 15 (John Wiley & Sons, Inc. New York Chichester Weinheim Brisbane Singapore Toronto, 1999) 327–336

Chapters 2–5

General

[1] M. Born and E. Wolf: *Principles of Optics* (Pergamon Press, Oxford London Edinburgh New York Paris Frankfurt, 1968)

[2] Robert H. Good: *Classical Electromagnetism* (Saunders College Publishing, Fort Worth Philadelphia San Diego New York Orlando Austin San Antonio Toronto Montreal London Sydney Tokyo 1999)

[3] L.D. Landau and E.M. Lifschitz: *Lehrbuch der theoretischen Physik*, Bd. II: Klassische Feldtheorie (Akademie Verlag, Berlin 1981) [engl.: Textbook of the theoretical physics, vol. II: Classical field theory]

[4] Д.В. Сивухин: *Общий Курс Физики* IV: *Оптика* (Москва Наука, Главная Редакция Физико-Математической Литературы 1980) (in Russ.) [engl.: D.V. Sivukhin: Physics IV: Optics]

[5] Chun Wa Wong: *Mathematische Physik: Konzepte, Methoden, Übungen* (Spektrum Akademischer Verlag, Heidelberg Berlin Oxford 1994) [engl.: Mathematical physics: Concepts, methods, exercises]

[6] М.Б. Виноградова, О.В. Руденко, А.П. Сухоруков: *Теория Волн* (Москва Наука, Главная Редакция Физико-Математической Литературы 1979) (in Russ.) [engl.: M.B. Vinogradova, O.V. Rudenko, and A.P. Sukhorukov: Wave theory (Moskau Nauka 1979)]

[7] H.-H. Perkampus: *Lexikon Spektroskopie* (VCH Verlagsgesellschaft, Weinheim New York Basel Cambridge 1993) [engl.: Encyclopedia spectroscopy]

[8] Brockhaus abc: *Physik*, Bde. 1 und 2 (VEB F.A. Brockhaus Verlag, Leipzig 1989) [engl.: Brockhaus abc: Physics, vol. 1 and 2]

[9] O. Stenzel: *Das Dünnschichtspektrum* (Akademie-Verlag, Berlin 1996) [engl.: The thin film spectrum]

[10] R.P. Feynman, R.B. Leighton, and M. Sands: *The Feynman Lectures of Physics*, Vol. 2 (Addison-Wesley Publishing Company, Inc. 1964)

[11] R.A. Serway and R.J. Beichner: *Physics: For Scientists and Engineers with Modern Physics*, 5th edn (Saunders College Publishing, Fort Worth Philadelphia San Diego New York Orlando Austin San Antonio Toronto Montreal London Sydney Tokyo 2000)

[12] D. Halliday, R. Resnick, and J. Walker: *Fundamentals of Physics*, 6th edn (John Wiley & Sons, Inc. New York Chichester Weinheim Brisbane Singapore Toronto 2001)

[13] H.J. Hediger: *Infrarotspektroskopie* (Akademie Verlag Ges, Frankfurt/a.M 1971) [engl.: Infrared spectroscopy]

[14] Paul Klocek (Ed.): *Handbook of Infrared Optical Materials* (Marcel Dekker, Inc., New York Basel Hong Kong 1991)

[15] Edward D. Palik (Ed.): *Handbook of Optical Constants of Solids* (Academic Press, Orlando 1998)

[16] K. Kreher: *Festkörperphysik*, Wissenschaftliche Taschenbücher Mathematik Physik, Bd. 103 (Akademie-Verlag, 1973) [engl.: Solid state physics, Academic pocket book, vol. 103]

[17] K. Kreher: *Elektronen und Photonen in Halbleitern und Isolatoren*, Wissenschaftliche Taschenbücher Mathematik Physik, Bd. 291 (Akademie-Verlag, 1986) [engl.: Electrons and photons in semiconductors and isolators, Academic pocket book, vol. 291]

[18] S.H. Wemple: Refractive-index behavior of amorphous semiconductors and glasses, Phys. Rev. B **7**, 3767–3777 (1973)

[19] S.G. Lipson, H.S. Lipson, and D.S. Tannhauser: *Optik* (Springer-Verlag Berlin 1997)

Chapter 3: Further Reading

[1] J.C. Phillips: Ionicity of the chemical bond in crystals, Rev. Mod. Phys. **42**, 317–356 (1970)

[2] Keith D. Bonin and Michael A. Kadar-Kallen: Linear electric-dipole polarizabilities, IJMPB **8**, 3313–3370 (1994)

[3] P. Dub: The influence of a surface monolayer on the s-polarized optical properties of a dielectric. The classical microscopical model, Surface Science **135**, 307–324 (1983)

[4] Amitabha Bagchi, Rubén G. Barrera, and Ronald Fuchs: Local-field effect in optical reflectance from adsorbed overlayers, Phys. Rev. B **25**, 7086–7096 (1982)

[5] Alexander Wokaun: Surface-enhanced electromagnetic processes, Solid State Physics **38**, 223–294 (1984)

[6] Y.R. Shen: *The Principles of Nonlinear Optics* (John Wiley & Sons, Inc., New York Chichester Brisbane Toronto Singapore 1984)

[7] A.S. Davydov: *Quantenmechanik* (VEB Deutscher Verlag der Wissenschaft, Berlin 1978) [engl.: Quantum mechanics]

[8] N. Bloembergen: *Nonlinear Optics* (Addison-Wesley Publishing Company, Inc., 1992)

Chapter 4: Further Reading

[1] Bergmann-Schäfer: *Lehrbuch der Experimentalphysik*, Bd. III: Optik, 9. Aufl. (Walter de Gruyter, Berlin 1993) [engl.: Textbook of experimental physics, vol. III: Optics, 9th edn]

[2] А.Н. Матвеев: *Оптика* Высшая Школа (Москва 1985) (in Russ.) [engl: A.N. Matveev: Optics]

[3] Aleksandra B. Djurišić and E. Herbert Li: Modeling the index of refraction of insulating solids with a modified Lorentz oscillator model, Appl. Opt. **37**, 5291–5297 (1998)

[4] A. Franke, A. Stendal, O. Stenzel, and C. von Borczyskowski: Gaussian quadrature approach to the calculation of the optical constants in the vicinity of inhomogeneously broadened absorption lines, Pure Appl. Opt. **5**, 845–853 (1996)

[5] Michael E. Thomas: A computer code for modeling optical properties of window materials, SPIE 1112: Window and Dome Technologies and Materials, 260–267 (1989)

[6] Olaf Stenzel: Optical absorption of heterogeneous thin solid films, Adv. in Solid State Phys. **39**, 151–160 (1999)

[7] U. Kreibig and M. Vollmer: *Optical Properties of Metal Clusters*, Springer Series in Materials Science, Vol. 25 (Springer-Verlag, Heidelberg 1995)

[8] L.D. Landau and E.M. Lifschitz: *Lehrbuch der theoretischen Physik, Bd. VIII: Elektrodynamik der Kontinua* (Akademie-Verlag, Berlin 1985) [engl.: Textbook of the theoretical physics, vol. VIII: Electrodynamics of continua]

[9] Vladimir M. Shalaev: *Optical Properties of Nanostructured Random Media* (Spinger-Verlag, Berlin Heidelberg New York 2002)

[10] D.E. Aspnes and J.B. Theeten, F. Hottier: Investigation of effective-medium models of microscopic surface roughness by spectroscopic ellipsometry, Phys. Rev. B **20**, 3292–3302 (1979)

[11] W. Theiss: The use of effective medium theories in optical Spectroscopy, Festkörperprobleme/Adv. in Solid State Phys. **33** (1993)

[12] W.A. Weimer and M.J. Dyer: Tunable surface plasmon resonance silver films, Appl. Phys. Lett. **79**, 3164–3166 (2001)

[13] V.A. Markel, V.M. Shalaev, P. Zhang, W. Huynh, L. Tay, T.L. Haslett, and M. Moskovits: Near-field optical spectroscopy of individual surface-plasmon modes in colloid clusters, Phys. Rev. B **59**, 10903–10909 (1999)

[14] S.J. Oldenburg, R.D. Averitt, S.L. Westcott, and N.J. Halas: Nanoengineering of optical resonances. Chem. Phys. Lett. **288**, 243–247 (1998)

[15] Alexander Wokaun: Surface-enhanced electromagnetic processes, Solid State Phys. **38**, 223–294 (1984)

[16] T. Yamaguchi, S. Yoshida, and A. Kinbara: Optical effect of the substrate on the anomalous absorption of aggregated silver films, Thin Solid Films **21**, 173–187 (1974)

[17] Joachim R. Krenn und Franz R. Aussenegg: Nanooptik mit metallischen Strukturen, Physik Journal **1**, Nr. 3, 39–45 (2002) [engl.: Nano-optics with metallic structures]

[18] Johannes Bosbach, Franz Stietz und Frank Träger: Ultraschnelle Elektrodynamik in Nanoteilchen, Physikalische Blätter **57**, Nr. 3, 59–62 (2001) [engl.: Ultra fast electrodynamics in nano-particles]

[19] Franz Stietz und Frank Träger: Monodispersive Metallcluster auf Oberflächen, Physikalische Blätter **55**, Nr. 9, 57–60 (1999) [engl.: Monodispersive metal clusters on surfaces]

[20] U. Kreibig, M. Gartz, and A. Hilger: Mie resonances: Sensors for physical and chemical cluster interface properties; Ber. Bunsenges. Phys. Chem. **101**, 1593–1604 (1997)

[21] Rolf E. Hummel and P. Wißmann (Eds.): *Handbook of Optical Properties, vol. II: Optics of Small Particles, Interfaces, and Surfaces* (CRC Press, Boca Raton New York London Tokyo 1995)

[22] O. Stenzel, S. Wilbrandt, A. Stendal, U. Beckers, K. Voigtsberger, and C. von Borczyskowski: The incorporation of metal clusters into thin organic dye layers as a method for producing strongly absorbing composite layers: an oscillator model approach to resonant metal cluster absorption, J. Phys. D: Appl. Phys. **28**, 2154–2162 (1995)

[23] O. Stenzel, A. Stendal, M. Röder, and C. von Borczyskowski: Tuning of the plasmon absorption frequency of silver and indium nanoclusters via thin amorphous silicon films, Pure Appl. Opt. **6**, 577–588 (1997)

[24] Bangyi Yang, Barbara L. Walden, Russell Messier, and William B. White: Computer simulation of the cross-sectional morphology of thin films, SPIE 821: Modeling of Optical Thin Films, 68–76 (1987)

[25] E.E. Chain and D.M. Byrne: Microstructural information related to thin film optical measurements, Thin Solid Films **181**, 323–332 (1989)

[26] Junzo Ishikawa, Yasuhiko Takeiri, Kiyoshi Ogawa, and Toshinori Takagi: Transparent carbon film prepared by mass-separated negative-carbon-ion-beam deposition, J. Appl. Phys **61**, 2509–2515 (1987)

[27] Eva C. Freeman and William Paul: Optical constants of rf sputtered hydrogenated amorphous Si, Phys. Rev. B **20**, 716–728 (1979)

[28] M.H. Brodsky (Ed.): *Amorphous Semiconductors* (Springer-Verlag, Berlin Heidelberg New York 1979)

Chapters 6–9

General

[1] M. Born and E. Wolf: *Principles of Optics* (Pergamon Press, Oxford London Edinburgh New York Paris Frankfurt 1968)

[2] S.G. Lipson, H.S. Lipson, and D.S. Tannhauser: *Optik* (Springer-Verlag, Berlin 1997)

[3] H.A. Macleod: *Thin-Film Optical Filters* (Adam Hilger Ltd., Bristol 1986)

[4] Hans Kuzmany: *Festkörperspektroskopie – Eine Einführung* (Springer-Verlag, Berlin Heidelberg New York London Paris Tokyo Hong Kong 1989) [engl.: Solid state spectroscopy – Introduction]

[5] Roland R. Willey: *Practical Design and Production of Optical Thin Films* (Marcel Dekker, Inc., New York Basel 2002)

[6] N. Kaiser and H.K. Pulker (Eds.): *Optical Interference Coatings* (Springer-Verlag, Berlin Heidelberg New York 2003)

[7] Ian J. Hodgkinson and Qi Hong Wu: *Birefringent Thin Films and Polarizing Elements* (World Scientific, Singapore New Jersey London Hong Kong 1997)

[8] Brian T. Sullivan and J.A. Dobrowolski: Deposition error compensation for optical multilayer coatings: I. Theoretical description, Appl. Opt. **31**, 3821–3835 (1992)

[9] Brian T. Sullivan and J. A. Dobrowolski: Deposition error compensation for optical multilayer coatings: II. Experimental results – Sputtering system, Appl. Opt. **32**, 2351–2360 (1993)

Chapter 6: Further Reading

[1] W.S. Letochow: *Laserspektroskopie*, Wissenschaftliche Taschenbücher, Bd. 165 (Akademie Verlag, Berlin 1977) [engl.: Laser spectroscopy, Academic pocket book, vol. 165]

[2] H. Ehrenreich, H.R. Philipp, and B. Segall: Optical properties of aluminum, Phys. Rev. **132**, 1918–1928 (1963)

[3] H. Ehrenreich and H.R. Philipp: Optical properties of Ag and Cu, Phys. Rev. **128**, 1622–1629 (1962)

[4] B.R. Coopert and H. Ehrenreich, H. R. Philipp: Optical properties of noble metals. II, Phys. Rev. **138**, A494–A507 (1965)

[5] H. Raether: *Surface Plasmons on Smooth and Rough Surfaces and on Gratings*, Tracts in Modern Physics 111 (Springer-Verlag, Berlin 1988)

[6] R.M.A. Azzam and N.M. Bashara: *Ellipsometry and Polarized Light* (Elsevier, Amsterdam 1987) 269

[7] Michael F. Weber, Carl A. Stover, Larry R. Gilbert, Timothy J. Nevitt, and Andrew J. Ouderkirk: Giant birefringent Optics in multilayer polymer mirrors, Science **287**, 2451–2456 (2000)

[8] Roger Strharsky and John Wheatley: Polymer optical interference filters, Optical and Photonic News, 34–40 (November 2002)

Chapter 7: Further Reading

[1] Enrico Nichelatti: Complex refractive index of a slab from reflectance and transmittance: analytical solution, J. Opt. A: Pure Appl. Opt. **4**, 400–403 (2002)

[2] O. Stenzel: The spectral position of absorbance maxima in ultrathin organic solid films: Dependence on film thickness, phys. stat. sol. (a) **148**, K33 (1995)

[3] B. Harbecke: Coherent and incoherent reflection and transmission of multilayer structures, Appl. Phys. B **39**, 165–170 (1986)

[4] J.H. Dobrowolski, F. C. Ho, and A. Waldorf: Determination of optical constants of thin film coating materials based on inverse synthesis, Appl. Opt. **22**, 3191–3196 (1983)

[5] O. Stenzel, R. Petrich, W. Scharff, V. Hopfe, and A. V. Tikhonravov: A hybrid method for determination of optical thin film constants, Thin Solid Films **207**, 324–329 (1992)

[6] O. Stenzel and R. Petrich: Flexible construction of error functions and their minimization: Application to the calculation of optical constants of absorbing or scattering thin-film materials from spectrophotometric datan J. Phys. D: Appl. Phys. **28**, 978–989 (1995)

[7] D.P. Arndt, R.M.A. Azzam, Jean M. Bennett, J.P. Borgogno, Charles K. Carniglia, William E. Case, J.A. Dobrowolski, Ursula J. Gibson, T. Tuttle Hart, F.C Ho, V.A. Hodgkin, W.P. Klapp, H. Angus Macleod, E. Pelletier, Martin K. Purvis, D.M. Quinn, D.H. Strome, R. Swenson, Paul A. Temple, Tracy F. Thonn: Multiple determination of the optical constants of thin-film coating materials, Appl. Opt. **23**, 3571–3596 (1984)

[8] J.C. Manifacier, J. Gasiot, and J.P. Fillard: A simple method for the determination of the optical constants n, k and the thickness of a weakly absorbing thin film, J. Phys. E: Scientific Instruments **9**, 1002–1004 (1976)

[9] Xuantong Ying, Albert Feldman, and E.N. Farabaugh: Fitting of transmission data for determining the optical constants and thichnesses of optical films, J. Appl. Phys. **67**, 2056–2059 (1990)

[10] R.T. Phillips: A numerical method for determining the complex refractive index from reflectance and transmittance of supported thin films, J. Phys. D: Appl. Phys. **16**, 489–497 (1983)

[11] Eduardo Elizalde, J.M. Frigerio, J. Rivory: Determination of thickness and optical constants of thin films from photometric and ellipsometric measurements, Appl. Opt. **25**, 4557–4561 (1986)

[12] J.P. Borgogno, B. Lazarides, E. Pelletier: Automatic determination of the optical constants of inhomogeneous thin films, Appl. Opt. **21**, 4020–4028 (1982)

[13] P. Grosse and V. Offermann: Analysis of reflectance data using Kramers–Kronig relations, Appl. Phys. A **52**, 138–144 (1991)

[14] Lawrence H. Robins, Edward N. Farabaugh, Albert Feldman: Determination of the optical constants of thin chemical-vapor-deposited diamond windows from 0.5 to 6.5 eV. Proc. SPIE 1534, Diamond Optics IV, 105–116 (1991)

[15] R.W. Tustison: Protective, infrared transparent coatings, Proc. SPIE CR39, 231–240 (1991)

Chapter 8: Further Reading

[1] A. Thelen: *Design of Optical Interference Coatings* (McGraw-Hill Book Company, 1989)

[2] S.A. Furman and A.V. Tikhonravov: *Basics of Optics of Multilayer Systems* (Edition Frontieres, Paris 1992)

[3] Daniel Poitras, Stéphane Larouche, and Ludvik Martinu: Design and plasma deposition of dispersion-corrected multiband rugate filters, Appl. Opt. **41**, 5249–5255 (2002)

[4] P.G. Verly and J.A. Dobrowolski: Iterative correction process for optical thin film synthesis with the Fourier transform method, Appl. Opt. **29**, 3672–3684 (1990)

[5] William H. Southwell: Using apodization functions to reduce sidelobes in rugate filters, Appl. Opt. **28**, 5091–5094 (1989)

[6] W.H. Southwell and Randolph L. Hall: Rugate filter sidelobe suppression using quintic and rugated quintic matching layers, Appl. Opt. **28**, 2949–2951 (1989)

[7] William H. Southwell: Coating design using very thin high- and low-index layers, Appl. Opt. **24**, 457–460 (1985)

[8] Thomas D. Rahmlow, Jr. and Jeanne E. Lazo-Wasem: Rugate and discrete hybrid filter designs, Proc. SPIE 3133, Int. Symp. on Optical Science, Engineering, and Instrumentation, San Diego 1997, 58–64

[9] Alexander V. Tikhonravov: Some theoretical aspects of thin-film optics and their applications, Appl. Opt. **32**, 5417–5426 (1993)

[10] B. Harbecke: Coherent and incoherent reflection and transmission of multilayer structures, Appl. Phys. B **39**, 165–170 (1986)

[11] J.A. Dobrowolski and S.H.C. Piotrowski: Refractive index as a variable in the numerical design of optical thin film systems, Appl. Opt. **21**, 1502–1511 (1982)

[12] J.A. Dobrowolski and D.G. Lowe: Optical thin film synthesis program based on the use of Fourier transforms (T), Appl. Opt. **17**, 3039–3050 (1978)

[13] J.P. Borgogno, P. Bousquet, F. Flory, B. Lazarides, E. Pelletier, and P. Roche: Inhomogeneity in films: limitation of the accuracy of optical monitoring of thin films, Appl. Opt. **20**, 90–94 (1981)

[14] J.A. Dobrowolski: Completely automatic synthesis of optical thin film systems, Appl. Opt. **4**, 937–946 (1965)

[15] А.Г. Свешников, А.В. Тихонравов: *Математичщские Моделирование – Математическое Методы в Задачах Анализа и Синтеза Скоистых Сред*, т. 1, но. 7/1989 (Москва Наука, Главная Редакция Физико-Математической Литературы 1989) [engl.: A.G. Sveshnikov and A.V. Tikhonravov: Mathematical methods in analysis and synthesis tasks in thin film optics]

[16] А.В. Тихонравов: *Математика Кибернетика – Синтез Слоистых Сред* 1987/5; (Издательство Знаниа Москва 1987) [engl.: A.V. Tikhonravov: Mathematical kybernetics – Synthesis of thin film systems]

Chapter 9: Further Reading

[1] O. Stenzel: New challenges in optical coating design, Adv. in Solid State Phys. **43**, 875–888 (2003)

[2] M. Nevière and E. Popov: *Light Propagation in Periodic Media* (Marcel Dekker, Inc., New York Basel 2003)

[3] E. Popov, L. Mashev, and D. Maystre: Theoretical study of the anomalies of coated dielectric gratings, Optica Acta **33**, 607–619 (1986)

[4] S.S. Wang and R. Magnusson: Theory and applications of guided-mode resonance filters, Appl. Opt. **32**, 2606–2613 (1993)

[5] S.S. Wang and R. Magnusson: Multilayer waveguide-grating filters, Appl. Opt. **34**, 2414– 2420 (1995)

[6] A. Sharon, S. Glasberg, D. Rosenblatt, and A.A. Friesem: Metal-based resonant grating waveguide structures, J. Opt. Soc. Am. A **14**, 588–595 (1997)

[7] A.Sharon, D. Rosenblatt, and A.A. Friesem: Resonant grating-waveguide structures for visible and near-infrared radiation, J. Opt. Soc. Am. A **14**, 2985–3993 (1997)

[8] F. Lemarchand, H. Giovannini, and A. Sentenac: Interest of hybrid structures for thin film design: Multilayered subwavelength micrograting, Proc. SPIE 3133, Int. Symp. on Optical Science, Engineering, and Instrumentation, San Diego 1997, 58–64

Chapters 10–12

General

[1] Edward D. Palik (Ed.): *Handbook of Optical Constants of Solids* (Academic Press, Orlando 1998)

[2] H. Paul: *Eine Einführung in die Quantenoptik* (Teubner Studienbücher: Physik 1995) [engl.: Introduction in quantum optics]

[3] В.И. Гавриленко, А.М. Грехов, Д.В. Корбутяк, В.Г. Литовченко: *Оптические Свойства Полупроводников – Справочник* (Киев Хаукова Думка 1987) (in Russ.) [engl.: V.I. Gavrilenko, A.M. Grechov, D.V. Korbutjak, and V.G. Litovcenko: Optical properties of semiconductors – Reference book (Kiev Naukova Dumka 1987)

[4] L.D. Landau and E.M. Lifschitz: *Lehrbuch der theoretischen Physik*, Bd. III: Quantenmechanik, (Akademie-Verlag, Berlin 1979) [engl.: Textbook of the theoretical physics, vol. III: Quantum mechanics]

[5] H. Haken and H.C. Wolf: *Atom- und Quantenphysik: Einführung in die experimentellen und theoretischen Grundlagen* (Springer-Verlag, Berlin Heidelberg New York 1992) [engl.: Atomic and quantum physics: Introduction in experimental and theoretical basics]

[6] H. Haken and H.C. Wolf: *Molekülphysik und Quantenchemie: Einführung in die experimentellen und theoretischen Grundlagen* (Springer-Verlag, Berlin Heidelberg New York 1992) [engl.: Molecular physics and quantum chemistry: Introduction in experimental and theoretical basics]

Chapter 10: Further Reading

[1] C. Kittel and H. Krömer: *Physik der Wärme* (R. Oldenbourg Verlag, München Wien 1989) [engl.: Thermal Physics]

Chapter 11: Further Reading

[1] S. Davydov: *Quantenmechanik* (VEB Deutscher Verlag der Wissenschaft, Berlin 1978) [engl.: Quantum mechanics]

[2] H. Schechtman and W.E. Spicer: Near infrared to vacuum ultraviolet absorption spectra and the optical constants of phthalocyanine and porphyrin films, J. Mol. Spectrosc. **33**, 28–48 (1970)

[3] A. Stendal, U. Beckers, S. Wilbrandt, O. Stenzel, and C. von Borczyskowski: The linear optical constants of thin phthalocyanine and fullerite films from the near infrared up to the UV spectral regions: Estimation of electronic oscillator strength values, J. Phys. B: At. Mol. Opt. Phys. **29**, 2589–2595 (1996)

Chapter 12: Further Reading

[1] C.F. Klingshirn: *Semiconconductor Optics* (Springer-Verlag, Berlin Heidelberg New York 1997)

[2] Hans Kuzmany: *Festkörperspektroskopie – Eine Einführung* (Springer-Verlag, Berlin Heidelberg New York London Paris Tokyo Hong Kong 1989) [engl.: Solid state spectroscopy – Introduction]

[3] C. Weißmantel and C. Hamann: *Grundlagen der Festkörperphysik* (VEB Deutscher Verlag der Wissenschaften, Berlin 1979) [engl.: Fundamentals of solid state physics]

[4] C. Kittel: *Introduction to Solid State Physics* (John Wiley & Sons, Inc., New York London Sydney Toronto 1971)

[5] H. Ibach and H. Lüth: *Festkörperphysik: Einführung in die Grundlagen* (Springer-Verlag, Berlin Heidelberg New York 1990) [engl.: Solid state physics: Introduction in the basics]

[6] V.L. Bonch-Bruevich and S.G. Kalashnikov: *Halbleiterphysik* (VEB Deutscher Verlag der Wissenschaften, Berlin 1982) [engl: Semiconductor physics]

[7] А.С. Давыдов: *Теория Твердого Тела* (Москва Наука, Главная Редакция Физико-Математической Литературы 1976) (in Russ.) [engl.: A.S. Davydov: Theory of solid state (Moskau Nauka 1976)]

[8] Richard Zallen: Symmetry and reststrahlen in elemental crystals, Phys. Rev. **173**, 824–832 (1968)
[9] Claude A. Klein, Thomas M. Hartnett, and Clifford J. Robinson: Critical-point phonon frequencies of diamond, Phys. Rev. B **45**, 12854–12863 (1992)
[10] M.H. Brodsky (Ed.): *Amorphous Semiconductors* (Springer-Verlag, Berlin Heidelberg New York 1979)
[11] Richard Zallen: *The Physics of Amorphous Solids* (John Wiley & Sons, Inc., New York Chichester Brisbane Toronto Singapore 1983)
[12] N.F. Mott and E.A. Davis: *Electronic Processes in Non-Crystalline Materials* (Clarendon Press, Oxford 1979)
[13] J. Tauc, J. Non-Crystall. Solids **97-98**, 149–154 (1987)
[14] Eva C. Freeman and William Paul: Optical constants of rf sputtered hydrogenated amorphous Si, Phys. Rev. B **20** 716–728 (1979)
[15] G.D. Cody, T. Tiedje, B. Abeles, B. Brooks, and Y. Goldstein: Disorder and the optical-absorption edge of hydrogenated amorphous silicon, Phys. Rev. Lett. **47**, 1480–1483 (1981)
[16] T. Datta, John A. Woollam, and W. Notohamiprodjo: Optical-absorption edge and disorder effects in hydrogenated amorphous diamondlike carbon films, Phys. Rev. B **40**, 5956–5960 (1989)
[17] O. Stenzel, R. Petrich, and M. Vogel: The optical constants of the so-called "diamond-like" carbon layers and their description in terms of semiempirical dispersion models, Opt. Mater. **2**, 125–142 (1993)

Chapter 13

[1] M. Schubert and B. Wilhelmi: *Einführung in die nichtlineare Optik I und II* (BSB B. G. Teubner Verlagsgesellschaft, Leipzig 1971) [engl.: Introduction in non-linear optics I and II]
[2] N. Bloembergen: *Nonlinear Optics* (Addison-Wesley Publishing Company, Inc. 1992)
[3] Y.R. Shen: *The Principles of Nonlinear Optics* (John Wiley & Sons, Inc., New York Chichester Brisbane Toronto Singapore 1984)
[4] Eugene Poliakov, Vladimir M. Shalaev, Vladimir Shubin, and Vadim A. Markel: Enhancement of nonlinear processes near rough nanometer-structured surfaces obtained by deposition of fractal colloidal silver aggregates on a plain substrate, Phys. Rev. B **60**, 10739–10742 (1999)
[5] Evgeni Y. Poliakov, Vadim A. Markel, Vladimir M. Shalaev, and Robert Botet: Nonlinear optical phenomena on rough surfaces of metal thin films, Phys. Rev. B **57**, 14901–14913 (1998)

Index

SPRINGER SERIES IN SURFACE SCIENCES

Editors: G. Ertl, and D.L. Mills

Founding Editor: H.K.V. Lotsch

1 **Physisorption Kinetics**
By H. J. Kreuzer, Z. W. Gortel

2 **The Structure of Surfaces**
Editors: M. A. Van Hove, S. Y. Tong

3 **Dynamical Phenomena at Surfaces, Interfaces and Superlattices**
Editors: F. Nizzoli, K.-H. Rieder, R. F. Willis

4 **Desorption Induced by Electronic Transitions, DIET II**
Editors: W. Brenig, D. Menzel

5 **Chemistry and Physics of Solid Surfaces VI**
Editors: R. Vanselow, R. Howe

6 **Low-Energy Electron Diffraction**
Experiment, Theory
and Surface Structure Determination
By M. A. Van Hove, W. H. Weinberg, C.-M. Chan

7 **Electronic Phenomena in Adsorption and Catalysis**
By V. F. Kiselev, O. V. Krylov

8 **Kinetics of Interface Reactions**
Editors: M. Grunze, H. J. Kreuzer

9 **Adsorption and Catalysis on Transition Metals and Their Oxides**
By V. F. Kiselev, O. V. Krylov

10 **Chemistry and Physics of Solid Surfaces VII**
Editors: R. Vanselow, R. Howe

11 **The Structure of Surfaces II**
Editors: J. F. van der Veen, M. A. Van Hove

12 **Diffusion at Interfaces: Microscopic Concepts**
Editors: M. Grunze, H. J. Kreuzer, J. J. Weimer

13 **Desorption Induced by Electronic Transitions, DIET III**
Editors: R. H. Stulen, M. L. Knotek

14 **Solvay Conference on Surface Science**
Editor: F. W. de Wette

15 **Surfaces and Interfaces of Solids**
By H. Lüth*)

16 **Atomic and Electronic Structure of Surfaces**
Theoretical Foundations
By M. Lannoo, P. Friedel

17 **Adhesion and Friction**
Editors: M. Grunze, H. J. Kreuzer

18 **Auger Spectroscopy and Electronic Structure**
Editors: G. Cubiotti, G. Mondio, K. Wandelt

19 **Desorption Induced by Electronic Transitions, DIET IV**
Editors: G. Betz, P. Varga

20 **Scanning Tunneling Microscopy I**
General Principles and Applications to Clean and Adsorbate-Covered Surfaces
Editors: H.-J. Güntherodt, R. Wiesendanger
2nd Edition

21 **Surface Phonons**
Editors: W. Kress, F. W. de Wette

22 **Chemistry and Physics of Solid Surfaces VIII**
Editors: R. Vanselow, R. Howe

23 **Surface Analysis Methods in Materials Science**
Editors: D. J. O'Connor, B. A. Sexton, R. St. C. Smart
2nd Edition

24 **The Structure of Surfaces III**
Editors: S. Y. Tong, M. A. Van Hove,
K. Takayanagi, X. D. Xie

25 **NEXAFS Spectroscopy**
By J. Stöhr

26 **Semiconductor Surfaces and Interfaces**
By W. Mönch
3rd Edition

27 **Helium Atom Scattering from Surfaces**
Editor: E. Hulpke

28 **Scanning Tunneling Microscopy II**
Further Applications
and Related Scanning Techniques
Editors: R. Wiesendanger, H.-J. Güntherodt
2nd Edition

29 **Scanning Tunneling Microscopy III**
Theory of STM
and Related Scanning Probe Methods
Editors: R. Wiesendanger, H.-J. Güntherodt
2nd Edition

30 **Concepts in Surface Physics**
By M. C. Desjonquères, D. Spanjaard*)

31 **Desorption Induced by Electronic Transitions, DIET V**
Editors: A. R. Burns, E. B. Stechel, D. R. Jennison

32 **Scanning Tunneling Microscopy and Its Applications**
By C. Bai
2nd Edition

33 **Adsorption on Ordered Surfaces of Ionic Solids and Thin Films**
Editors: H.-J. Freund, E. Umbach

34 **Surface Reactions**
Editor: R. J. Madix

35 **Applications of Synchrotron Radiation**
High-Resolution Studies of Molecules
and Molecular Adsorbates on Surfaces
Editor: W. Eberhardt

36 **Kinetics of Metal-Gas Interactions at Low Temperatures: Hydriding, Oxidation, Poisoning**
By E. Fromm

37 **Magnetic Multilayers and Giant Magnetoresistance**
Fundamentals and Industrial Applications
Editor: U. Hartmann*)

*) Available as a textbook