

**Integrated Optics**

**Design and Modeling**

**Integrated Optics  
Design and Modeling**

**Reinhard März**



**Artech House  
Boston • London**

For a complete listing of the *Artech House Optoelectronics Library*,  
turn to the back of this book.

## **Library of Congress Cataloging-in-Publication Data**

März, Reinhard.

Integrated optics : design and modeling / Reinhard März.

Includes bibliographical references and index.

ISBN 0-89006-668-X

1. Integrated optics. I. Title

TA1660.M37 1994

621.36'93--dc20

94-37147

CIP

## **British Library Cataloguing in Publication Data**

März, Reinhard

Integrated Optics: Design and Modeling

I. Title

621.36

ISBN 0-89006-668-X

© 1995 ARTECH HOUSE, INC.

685 Canton Street

Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

International Standard Book Number: 0-89006-668-X

Library of Congress Catalog Card Number: 94-37147

10 9 8 7 6 5 4 3 2 1

## **Contents**

### Preface

xi

### Chapter 1. Introduction

1

#### 1.1 History

1

#### 1.2 Materials

2

#### 1.3 Outline of the Book

5

#### References

7

### Chapter 2. Foundations

9

#### 2.1 Helmholtz Equations

9

##### 2.1.1 Maxwell's Equations

9

##### 2.1.2 Vector Helmholtz Equations

12

##### 2.1.3 Scalar Helmholtz Equations

13

##### 2.1.4 Waves in Homogeneous Media

15

##### 2.1.5 Polarization

16

#### 2.2 Dielectric Interfaces

18

##### 2.2.1 Boundary Conditions

18

##### 2.2.2 Reflection and Refraction

21

##### 2.2.3 Fabry-Perot Interferometer

25

##### 2.2.4 Interference Filters

27

#### 2.3 Geometrical Optics

32

##### 2.3.1 Eikonal Equation

32

##### 2.3.2 Ray Equation and Fermat's Principle

34

##### 2.3.3 Caustics

36

#### References

37

### Chapter 3. Waveguide Theory

39

#### 3.1 General Remarks

40

##### 3.1.1 Phenomena of Waveguiding

40

##### 3.1.2 Typical Waveguides in Integrated Optics

42

##### 3.1.3 Notations

44

3.2	Formulations of the Eigenvalue Problem	45	4.7.2	Benchmark Results	125
3.2.1	The Vector $E$ -Field and $H$ -Field Eigenvalue Problems	46	References		129
3.2.2	The $E_z/H_z$ Eigenvalue Problem	47	 Chapter 5.	Mode Conversion	131
3.2.3	Transverse Eigenmodes	48	5.1	Hermite-Gaussian Beams	131
3.2.4	Orthogonality of the Eigenmodes	50	5.1.1	Evolution of Hermite-Gaussian Beams	132
3.3	Power Flux in Waveguides	53	5.1.2	Physical Properties of the Hermite-Gaussian Beam	135
3.4	Dispersion	54	5.2	Coupling of Gaussian Beams	136
3.5	Universal Approximation Schemes	55	5.3	Gaussian Optics	143
3.5.1	The Ritz Variational Principle	55	5.3.1	Transformation Rules for the Beam Parameter	143
3.5.2	Rayleigh-Schrödinger Perturbation Theory	56	5.3.2	Imaging by Optical Systems	148
3.6	Weakly Guiding Waveguides	58	5.3.3	Spherical Aberrations	152
3.6.1	Nearly Uniform Waveguides	59	5.3.4	Chromatic Aberrations	155
3.6.2	Slab-like Waveguide Structures	62	5.4	Fiber-to-Chip Coupling by Microlenses	157
3.7	Analytical Solutions	65	5.4.1	Microlenses	158
3.7.1	The Parabolic Slab	65	5.4.2	One- and Two-Lens Assemblies	162
3.7.2	The Circular Parabolic Fiber	69	5.5	Optical Tapers	163
3.8	Slab Waveguides	74	5.5.1	Local Normal Mode Theory	163
3.9	Separable Eigenvalue Problems	77	5.5.2	Design of Optical Tapers	166
3.9.1	Marcatili's Approach	79	References		168
3.9.2	The Effective Index Method	80	 Chapter 6.	Codirectional Coupling	169
3.10	Geometrical Optics of the Waveguide	82	6.1	Coupled Mode Theory	169
References		85	6.1.1	Eigenmodes of Coupled Waveguide Arrays	170
 Chapter 4.	Beam Propagation	87	6.1.2	Phenomenological Coupled Mode Theory	173
4.1	Fundamental Principles	88	6.1.3	Coupling Coefficients	175
4.1.1	Forward Helmholtz Equations	89	6.1.4	Tapered Coupling	178
4.1.2	Mathematical Assessment	92	6.2	Uniform Directional Couplers	181
4.1.3	Optimized Paraxial Approximations	94	6.2.1	Solution of the Coupled Mode Equations	182
4.1.4	Boundary Conditions	95	6.2.2	Physical Properties of Directional Couplers	184
4.2	Classical BPM	98	6.2.3	Coupling Coefficients	185
4.2.1	Diffraction	98	6.2.4	Transfer Matrix of the Directional Coupler	186
4.2.2	Derivation of the Propagator	101	6.3	Symmetrical Couplers	188
4.2.3	Interpretation of the Classical BPM	103	6.4	Asymmetrical Couplers	191
4.2.4	Numerical Fourier Transformation	105	6.4.1	Filter Characteristic	191
4.2.5	Optimized BPM Propagators	109	6.4.2	Coupling Coefficients	193
4.3	Diffraction Integrals	110	6.4.3	Tapered Coupling	195
4.4	Improved BPM Algorithms	111	6.5	Transfer Matrix Theory	197
4.4.1	Eigenmode Based BPM	112	6.5.1	Elementary Components	197
4.4.2	Finite Differences: FD-BPM	113	6.5.2	Compound Components and Networks	202
4.4.3	Finite Elements: FE-BPM	117	6.6	Mach-Zehnder Devices	205
4.5	Vectorial Beam Propagation Method	120	6.6.1	Mach-Zehnder Interferometer	205
4.6	Bidirectional Beam Propagation	120	6.6.2	Mach-Zehnder Couplers	207
4.7	Application of the BPM	123	6.7	Directional Couplers with Periodic Overlay	209
4.7.1	3D/2D Beam Propagation Method	124			

6.7.1 Some Properties of Periodic Couplers	209	Chapter C. Component Cost Modeling	301
6.7.2 $\Delta\beta$ -Coupler	211	C.1 Yield Modeling	302
6.7.3 $\Delta\kappa$ -Coupplers	212	C.2 Yield Extrapolation	305
References	216	C.3 Cost Extrapolation	306
<b>Chapter 7. Contradirectional Coupling</b>	<b>219</b>	C.4 Example: Transceiver Chip	307
7.1 Phenomenological Coupled Mode Theory	220	C.5 Arrays of Identical Devices	307
7.1.1 Coupled Mode Equations	220	C.6 Integrated Optics vs. Microelectronics	310
7.1.2 Radiation	224	References	311
7.1.3 Dispersion and Filter Curves	227	<b>Chapter D. Mathematical Background</b>	313
7.1.4 Chirped and Tapered Coupling	229	D.1 Vector Algebra and Analysis	313
7.1.5 Mathematical Assessment	230	D.1.1 Coordinate Systems	315
7.2 Bragg Gratings	231	D.1.2 Generalized Cylindrical Coordinates	317
7.2.1 Coupled Mode Propagator	232	D.2 Linear Operators	321
7.2.2 Operation at Oblique Incidence	235	References	328
7.2.3 Compound Bragg Gratings	242	<b>About the Author</b>	329
7.2.4 Curved Bragg Gratings	244	<b>Index</b>	331
7.3 Contradirectional Coupler	247		
References	250		
<b>Chapter 8. Planar Spectrographs</b>	<b>251</b>		
8.1 Theory of Planar Spectrographs	252		
8.1.1 Light-Path Function	252		
8.1.2 Imaging by Focusing Spectrographs	256		
8.1.3 Stigmatization	261		
8.1.4 Efficiency	262		
8.1.5 Spectral Resolution and Free Spectral Range	266		
8.2 Focusing Reflection Gratings	269		
8.2.1 Stigmatic Mountings	270		
8.2.2 Improved Rowland Mountings	273		
8.2.3 Flat-Field Spectrographs	274		
8.3 Phased Arrays	277		
8.3.1 Principle of Operation	277		
8.3.2 Rowland Mountings	279		
8.3.3 Stigmatic Mountings	280		
References	281		
<b>Chapter A. Application of Photonics</b>	<b>283</b>		
<b>Chapter B. Computer Aided Engineering</b>	<b>289</b>		
B.1 Software Environment	289		
B.2 Beam Propagation Method	292		
B.3 Rapid Prototyping	295		
References	300		

## Preface

This monograph is intended to provide an in-depth treatment of integrated optics from a designer's point of view. Most of the book deals with the derivation of methods for the design and modeling of a large class of integrated optical components such as fiber-to-chip couplings, various types of codirectional couplers, Bragg gratings, contradirectional couplers, focusing reflection gratings and phased arrays. Since the book attempts to be self-contained, it also covers the theoretical foundations of integrated optics, in particular waveguide theory and the theory of beam propagation. The supplementary material collected at the end will help both to assess the role of integrated optics within the framework of photonics and to classify the tools presented here within the environment of computer aided engineering.

Within this book we will often make use of a quantum-mechanical notation in order to indicate the mathematical background as well as the high degree of parallelism between "classical" quantum mechanics and integrated optics. Another advantage of this notation is its universal character, i.e., many problems and techniques may be formulated irrespective of the approximations introduced previously.

I would like to express my thanks to Siemens AG for their support of this project. In particular, I owe thanks to my colleagues Dr. Bruno Acklin (Siemens AG, Munich), Dr. Cornelius Cremer (Siemens AG, Munich), Dr. Hans F. Mahlein (Siemens AG, Munich), Dr. Hans-Peter Nolting (Heinrich-Hertz Institut, Berlin), Dr. Meinrad Schienle (Siemens AG, Munich), Dr. Frank Schmidt (Konrad-Zuse Zentrum für Informationstechnik, Berlin) and not least to my brother Dr. Christoph März for many helpful discussions and for their careful reading of parts of the manuscript.

Reinhard März  
Munich  
June 1994

# Chapter 1

## Introduction

### 1.1 HISTORY

The term *integrated optics* was introduced in 1969 when Miller [1] presented the concept of integrated optical circuits that should be integrated on a single substrate. Some theoretical concepts relating to waveguides, co- and contradirectional couplers and other devices could be taken over from the theory of microwaves which had been established 15 years previously.

Most activities in the pioneer days of integrated optics, i.e., in the 1970s and early 1980s, were focussed on demonstrating devices that were compatible with the technologies of integrated optics. The distributed feedback (DFB) laser demonstrated by Kogelnik and Shank in 1972 represents a very prominent milestone along this path. A rich variety of passive components such as directional couplers, Y-branches, waveguide crossings, Bragg gratings, transmission gratings, acousto-optical filters, optical switches and modulators were also demonstrated for the first time. An early 4-channel WDM-transmitter (WDM – wavelength division multiplex) demonstrated by Nakamura et al. represents the first attempt to realize integrated optical circuits.

In parallel to those developments, the theoretical background of integrated optics was successively established. The most spectacular event in this field was the application of the beam propagation method (BPM) to problems of integrated optics by Feit and Fleck in 1978.

The following period was characterized by successive consolidation, i.e., by a stabilization of the technology, an increase in fabrication yield and more quantitative predictions of device behavior. On the one hand, these efforts were aimed at the fabrication of state-of-the-art devices, while on the other they were forced by the development of more sophisticated integrated optical circuits, i.e., by circuits requiring an increasing chip area and/or number of fabrication steps. Prominent examples of integrated optical circuits and highly complex single devices realized during this period by different groups are

- 8x8 Switching Matrix (Ericsson, HHI, NTT)
- Bidirectional Transceiver (AT&T, GMMT, Siemens)
- DWDM<sup>1</sup> Receiver (Bellcore, Philips, Siemens)
- DWDM Transmitter (AT&T)
- Coherent Receiver (AT&T, NTT, HHI)
- Laser + Optical Amplifier (AT&T)
- Laser+Modulator (AT&T, Alcatel)
- Disc Pickup (University of Osaka)
- Tunable 128-Channel Filter (NTT)

Although many activities were focussed on this consolidation, some new components such as focussing reflection gratings, phased arrays, polarization independent acousto-optical filters and optical tapers were demonstrated for the first time.

This trend towards consolidation was accompanied by the development of flexible computer programs that could analyze “arbitrary” waveguide structures. Finite differences (FD), finite elements (FE) and various other numerical methods were used to calculate the eigenmodes of waveguides. Many new transient algorithms were implemented to overcome the lack of stability of the “classical” beam propagation method. The first CAD system for integrated optics was developed.

Today, some integrated optical components have attained a high degree of maturity. Some of them already offer the performance of expensive hybrid assemblies.

## 1.2 MATERIALS

In the previous section we examined the history of integrated optics from a components point of view. We will now look at it from the materials side.

The materials and processing used for the fabrication of integrated optical devices impose restrictions on their design. These boundary conditions concern the dielectric profile, i.e., the available waveguide dimensions and the refractive index contrast.<sup>2</sup> In addition, practical designs aim at devices which are insensitive to the most critical fabrication tolerances. For this purpose the underlying waveguide structure is optimized by a tolerance analysis. It should be noted that the fabrication tolerances represent a rapidly moving target which must be individually assessed for each type of processing.

Figure 1.1 shows the most important classes of materials used for the realization of integrated optical circuits. Only the III-V material systems (InGaAsP/InP and GaAlAs/GaAs), both representing direct semiconductors, allow for monolithic

Figure 1.1 is a table comparing five material systems used in integrated optics. The columns represent various physical and technological parameters:

	Transmitters/Receivers	Passive optics	Wavelength range	Fiber-to-chip loss [dB]	Waveguide loss [dB/cm]	$dn/dT [10^{-4}/K]$	$In_{TE} - n_{TM} [10^{-4}]$	Long term stability	Wafer size
InGaAsP/InP	●	●	T	>2**	2	2	0.1-10	+	2-3"
GaAlAs/GaAs	●	●	I	>2**	2	4	2-10	+	3"
Glass	●	T, I	I	0.4	<0.1	<0.1	0.1-0.5	+	≥4"
LiNbO <sub>3</sub>	●	T, I	I	<1	<0.3	<0.1	400	+	3"
Polymers	●	I*	I	<0.5	0.1-1.5	-1...3	2-50	?	≥8"

Legend:  
 T (transmission) = 1.3/1.5 μm  
 I (interconnects) = <1.1 μm  
 \* - at 1.3 μm waveguide losses > 1 dB/cm, at 1.5 μm > 3 dB/cm  
 \*\* - can be improved by using opt. taper and vertical cavity lasers

Figure 1.1. Material systems of integrated optics.

integration; i.e., for the integration of laser- and photodiodes with passive optical components on a single substrate. In addition, integrated optics covers a great variety of other material systems which allow for the integration of passive components such as power splitters, optical filters and switches.

Figure 1.1 also shows some typical physical and technological parameters such as the temperature dependence  $dn/dT$  of the refractive index model, birefringence  $n_{TE} - n_{TM}$ , fiber-to-chip coupling and waveguide loss as well as the long term stability of the material systems. In the following we will briefly examine a number of material systems:

### InGaAsP/InP

The InGaAsP/InP material system [2] allows for monolithic integration within the 1.3/1.5 μm wavelength region where the glass fibers exhibit extremely low losses. The waveguide structures are formed by epitaxy and dry etching. Due to their high refractive index and the high index contrasts, the waveguide structures are rather

1. DWDM – dense wavelength division multiplex.

2. In Section 3.1.2 we will identify a set of “standard” waveguides for integrated optics and discuss the interdependence of the dielectric profile and the material and processing in more detail.

small (typically  $1 \mu\text{m}$ ). The fiber-to-chip coupling requires spot magnifications of  $m = 3\dots 5$  depending on the underlying waveguide structures.

The evolution of the dielectric constant of the quaternary  $\text{In}_{1-x}\text{Ga}_x\text{As}_y\text{P}_{1-y}$  material system as a function of the composition parameter  $y$  and the photon energy  $E$  can be formulated in terms of a single effective oscillator model [3]

$$\epsilon = 1 + \eta_d + \eta_d \eta^2 + \frac{\eta_d^2 \eta^4}{2(1 - \eta_g^2)} \ln \left( \frac{2 - \eta_g^2 - \eta^2}{\eta_g^2 - \eta^2} \right) \quad (1.1)$$

where

$$\begin{aligned} \eta &= \frac{E/\text{eV}}{3.391 - 1.652y + 0.863y^2 - 0.123y^3} \\ \eta_d &= \frac{28.91 - 9.278y + 5.626y^2}{3.391 - 1.652y + 0.863y^2 - 0.123y^3} \\ \eta_g &= \frac{1.35 - 0.72y + 0.12y^2}{3.391 - 1.652y + 0.863y^2 - 0.123y^3}. \end{aligned}$$

We should note that Equation (1.1) is restricted to quaternary layers which are lattice matched to InP ( $y = 2.197x$ ) and to photon energies below the band edge ( $\eta < \eta_g$ ).

### GaAlAs/GaAs

The GaAlAs/GaAs material system [4] allows for monolithic integration within the  $0.8 \mu\text{m}$  wavelength region. The corresponding components are usually applied to short-range interconnections. The remarks on the fabrication and refractive index profile of InGaAsP/InP also apply to GaAlAs/GaAs waveguide structures.

Highly sophisticated models [5] for the dielectric constant of the ternary material system  $\text{Ga}_{1-x}\text{Al}_x\text{As}$  exist today. They allow us to calculate the material dispersion with a high accuracy over a huge spectral range.

### Glass Materials

Waveguide structures in glass are fabricated today by two different technologies: diffusion and deposition. Ion exchange technology [6] is based on the exchange of ions in special glasses by purely thermal or field-assisted diffusion. Alternatively, waveguide structures can also be fabricated [7] by the deposition of glass using chemical vapor (CVD) or flame hydrolysis deposition (FHD) on a substrate (usually silicon). Waveguide structures in glass exhibit extremely low waveguide losses irrespective of the underlying fabrication technology. Most of the waveguide structures are fiber-matched, i.e., the chips can be butt-coupled to standard single-mode fibers.

The formulas describing the dispersion of glasses are in most cases purely phenomenological. They are usually formulated in terms of a Taylor expansion.

### **LiNbO<sub>3</sub>**

Lithium niobate [4] represents an anisotropic material with high electro-optical and acousto-optical coefficients. The waveguide structures are fabricated by a diffusion process. The anisotropy yields always a significant polarization dependence for the devices.

### **Polymers**

Polymers actually represent a huge class of materials. For purely passive optical components, polycarbonate (used for compact discs) and PMMA (photoresist) are preferred today. Other polymers [8] exhibit a large electro-optical or non-linear coefficient, but often at the expense of a poor long term stability.

For a long time it looked as if the III-V technologies would drive out their competitors, i.e., the glass technologies (SiO<sub>2</sub>/Si and ion exchange), polymers and LiNbO<sub>3</sub>. However, the complexity of fabrication, yield problems and costs of the III-V technologies, their high fiber-to-chip coupling and waveguide losses as well as their significant polarization dependence kept the competition between the material systems open (1994).

### 1.3 OUTLINE OF THE BOOK

We will end this introduction by taking a short tour through the book. The material presented in the following chapters can be divided into two parts: the theoretical foundation, and the design and modeling of integrated optical devices.

The following three chapters form the theoretical background of integrated optics.

- Chapter 2, “Foundations”, covers a collection of fundamental topics which are required for the following chapters. We will start with a derivation of the vectorial and scalar Helmholtz equations. An examination of dielectric interfaces, i.e., a discussion of Snell’s law, Fresnel’s formulas, the Goos-Hänchen shift and the theory of interference filters will follow. At the end of this chapter we will take a look at geometrical optics in order to obtain a feeling for the underlying approximations.
- Chapter 3, “Waveguide Theory”, deals with the theory of dielectric waveguides. We will examine various formulations of the eigenvalue problem, the derivation of two universal approximation schemes — the Ritz variational principle and the Rayleigh-Schrödinger perturbation theory — and the most prominent approximations for weakly guiding waveguides, namely, the “weakly guiding” approximation, the effective index method (EIM) and Marcato’s approach. For later use we will also derive analytical solutions for some types of parabolic waveguide.

- Chapter 4, “Beam Propagation”, focusses on the theory of finite beams propagating through an inhomogeneous half-space. We will start with a derivation of the forward Helmholtz equation. In a second step we will see the equivalence of the paraxial Helmholtz equation and Fresnel’s diffraction integral both representing small angle approximations of the original beam propagation problem. This chapter will include an extensive assessment of the “classical” beam propagation method and various competing algorithms.

In the other four chapters we will concentrate on the design and modeling of integrated optical devices.

- Chapter 5, “Mode Conversion”, deals with the efficient interfacing of waveguides. We will examine the three most important techniques of interfacing: butt coupling, imaging by optical systems and adiabatic mode conversion by optical tapers. Butt coupling and imaging by optical systems are treated within the framework of Gaussian optics, i.e., by calculating the beam propagation and overlap integrals of Hermite-Gaussian beams. Optical tapers are analyzed by means of local normal mode theory.
- Chapter 6, “Codirectional Coupling”, tackles the interaction of guided modes traveling in the same direction. This chapter starts with an examination of the required design tools, i.e., the coupled mode theory describing the operation of uniform couplers, and the transfer matrix theory which turns out to be crucial for the description of compound devices and networks. By using these techniques we will examine the most important devices such as symmetrical and asymmetrical couplers with and without  $\kappa$ -tapering, Y-branches, waveguide crossings, Mach-Zehnder interferometers and directional couplers with a periodic overlay, i.e.,  $\Delta\beta$ - and  $\Delta\kappa$ -couplers.
- Chapter 7, “Contradirectional Coupling”, tackles the interaction of guided modes traveling in the opposite direction via almost periodic structures. After deriving the coupled mode and transfer matrix theories of contradirectional coupling, we will concentrate on a discussion of the rich variety of physical effects occurring in Bragg gratings, such as radiation of optical power into the substrate and superstrate, TE-TM polarization conversion, the Brewster angle for TE-polarization and the Goos-Hänchen shift for finite beams. The chapter ends with a discussion of contradirectional couplers within the framework of a 4-wave coupled mode theory.
- Chapter 8, “Planar Spectrographs”, deals with the design of focussing planar spectrographs. We will first derive the general theory of imaging and aberrations, and follow it by an analysis of efficiency, spectral resolution and free spectral range. We will apply these results to analyze focussing reflection gratings and phased arrays. We will also discuss techniques for constructing spectrographs with stigmatic, i.e., aberration-free, points.

The supplementary material collected at the end of this book shall help to assess the role of integrated optics within the framework of photonics and also to classify the tools presented here within the environment of computer aided engineering.

## REFERENCES

- [1] Miller, S. E., “Integrated Optics: An Introduction”, *The Bell System Technical Journal*, Vol. 48, 1969, pp. 2059–2068.
- [2] Koch, T. L., and Koren, U., “Semiconductor Photonic Integrated Circuits”, *IEEE Journal of Quantum Electronics*, Vol. 27, 1991, pp. 641–653.
- [3] Pearsell, T. P. (ed.), *GaInAsP Alloy Semiconductors*, New York, John Wiley & Sons, 1982.
- [4] Tamir, T. (ed.), *Integrated Optics*, New York, Springer, 1975.
- [5] Jenkins, D. W., “Optical Constants of  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ”, *Journal of Applied Physics*, Vol. 68, 1990, pp. 1848–1853.
- [6] Ramaswamy, R. V., and Srivastava, R., “Ion-Exchanged Glass Waveguides: a Review”, *IEEE Journal of Lightwave Technology*, Vol. 6, 1988, pp. 984–1002.
- [7] Kawachi, M., “Silica Waveguides on Silicon and Their Application to Integrated-Optic Components”, *Optical and Quantum Electronics*, Vol. 22, 1990, pp. 391–416.
- [8] Van Tomme, E., van Daele, P. P., Baets, R. G., and Lagasse, P.E., “Integrated Optical Devices Based on Nonlinear Optical Polymers”, *IEEE Journal of Quantum Electronics*, Vol. 27, 1991, pp. 778–787.

## Chapter 2

### Foundations

In the first chapter of this book we review some fundamental properties of electromagnetic radiation in inhomogeneous isotropic media. This material forms the background for the other topics covered in this book and introduces a common notation.

We start from Maxwell's equations and derive the vector and scalar Helmholtz equations. After a short discussion of waves in homogeneous space and of the corresponding polarization states, we examine the influence of the dielectric interface; i.e., we study the boundary conditions, Snell's law, Fresnel's formula and the Goos-Hänchen shift. The chapter is concluded by a brief derivation of geometrical optics starting from the vector Helmholtz equation. Within this section we will derive the eikonal and ray equations as well as Fermat's principle and give some idea of the applicability of geometrical optics.

#### 2.1 HELMHOLTZ EQUATIONS

##### 2.1.1 Maxwell's Equations

Using Gaussian units, the evolution of the electric field  $\mathbf{E}(t)$  and of the magnetic field  $\mathbf{H}(t)$  in a medium with the magnetization density  $\mathbf{M}(t)$  and the polarization density  $\mathbf{P}(t)$  is given by Maxwell's equations

$$\nabla \times \mathbf{E}(t) = -\frac{1}{c} \dot{\mathbf{H}}(t) - \frac{4\pi}{c} \dot{\mathbf{M}}(t) \quad (2.1)$$

$$\nabla \times \mathbf{H}(t) = \frac{1}{c} \dot{\mathbf{E}}(t) + \frac{4\pi}{c} \dot{\mathbf{P}}(t) + \frac{4\pi}{c} \mathbf{J}(t) \quad (2.2)$$

$$\nabla \cdot \mathbf{E}(t) = -4\pi \nabla \cdot \mathbf{P}(t) + 4\pi \rho(t) \quad (2.3)$$

$$\nabla \cdot \mathbf{H}(t) = -4\pi \nabla \cdot \mathbf{M}(t), \quad (2.4)$$

in which

$$c = 2.998 \cdot 10^8 \frac{m}{s} \quad (2.5)$$

is the phase velocity of light in free space. Within Maxwell's equations  $\rho(t)$  is the charge density, and  $\mathbf{J}(t)$  stands for the current density caused by all charges which are not associated with the polarization density  $\mathbf{P}(t)$ .

For integrated optics, we are usually interested in time-harmonic solutions of Maxwell's equations, which are obtained as a quasi-stationary solution after all relaxation oscillations are completed. Only for some types of nonlinear devices, which are not the subject of our book, will relaxation oscillations themselves be examined. Furthermore, for the rest of this book we will concentrate on nonmagnetic materials ( $\mathbf{M}(t) = \mathbf{0}$ ). We will also assume that all charges and currents are described by the polarization density  $\mathbf{P}(t)$ , i.e.,  $(\rho(t) = 0, \mathbf{J}(t) = \mathbf{0})$ . Finally, we shall restrict ourselves to linear and isotropic media, whose polarization density

$$\mathbf{P}(t) = \int_0^\infty d\tau \chi_e(\tau) \mathbf{E}(t - \tau) \quad (2.6)$$

is related to the electric field via the electric susceptibility  $\chi_e$ , which is a scalar quantity for isotropic media. In general, we can obtain time-harmonic quantities by Fourier transformation, i.e., by

$$f(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt e^{i\omega t} f(t). \quad (2.7)$$

Under these conditions, we arrive at the time-harmonic Maxwell equations<sup>1</sup>

$$\nabla \times \mathbf{E} = ik_0 \mathbf{H} \quad (2.8)$$

$$\nabla \times \mathbf{H} = -ik_0 \epsilon \mathbf{E} \quad (2.9)$$

$$\nabla \cdot \epsilon \mathbf{E} = 0 \quad (2.10)$$

$$\nabla \cdot \mathbf{H} = 0, \quad (2.11)$$

where we introduce the free space propagation constant

$$k_0 = \frac{\omega}{c} = \frac{2\pi}{\lambda_0}, \quad (2.12)$$

the vacuum wavelength  $\lambda_0$  and the dielectric function

$$\epsilon(\mathbf{r}, \omega) = 1 + 4\pi\chi_e(\mathbf{r}, \omega). \quad (2.13)$$

1. To avoid clumsy notation throughout this book we will distinguish a function and its Fourier transform only by using the corresponding variables, e.g.,  $f(t) \xrightarrow{\text{FT}} f(\omega)$ . Furthermore, we will use the time-harmonic Maxwell equations for the rest of the book; i.e., all quantities in these equations are assumed to be Fourier transformed with respect to time without any additional comment.

The real part of the refractive index profile

$$n(\mathbf{r}, \omega) = \sqrt{\epsilon(\mathbf{r}, \omega)} \quad (2.14)$$

relates the wavelength inside the medium  $\lambda = \lambda_0/n$  to the vacuum wavelength  $\lambda_0$ . The imaginary part of the refractive index profile describes the absorption of the optical field.

We should point out here that as a consequence of the principle of causality, which states that no effect will occur before it is caused, the real and imaginary parts of the susceptibility  $\chi_e$  in Equation (2.6) are interdependent via the Kramers-Kronig relations

$$\text{Re}(\chi_e(\omega)) = \frac{2}{\pi} \mathcal{P} \int_0^\infty d\Omega \frac{\Omega \text{Im}(\chi_e(\Omega))}{\Omega^2 - \omega^2} \quad (2.15)$$

$$\text{Im}(\chi_e(\omega)) = \frac{2}{\pi} \mathcal{P} \int_0^\infty d\Omega \frac{\omega \text{Re}(\chi_e(\Omega))}{\Omega^2 - \omega^2}, \quad (2.16)$$

where  $\mathcal{P} \int d\Omega f(\Omega)$  is the principal value of the integral  $\int d\Omega f(\Omega)$ . In physical terms, *the dispersion and absorption in a medium are interdependent*. Thus, strong dispersion occurs only in the spectral vicinity of absorption bands such as the fundamental gap of a III-V semiconductor.

The mean power flux density is given by the real part of the complex Poynting vector

$$\mathbf{S} = \frac{c}{8\pi} (\mathbf{E} \times \mathbf{H}^*). \quad (2.17)$$

This physical quantity is used to determine the power flux inside waveguide structures or the interface between different media. Using the vector identity

$$\nabla \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\nabla \times \mathbf{a}) - \mathbf{a} \cdot (\nabla \times \mathbf{b})$$

we derive Poynting's theorem

$$\nabla \cdot \mathbf{S} = 2i\omega (w_m - w_e). \quad (2.18)$$

Since the mean of a product of functions  $a(t)b(t)$  with respect to time is

$$\langle a(t)b(t) \rangle = \lim_{T \rightarrow \infty} \frac{1}{4T} \int_0^T dt (ae^{-i\omega t} + a^*e^{i\omega t})(be^{-i\omega t} + b^*e^{i\omega t}) = \frac{1}{2} \text{Re}(ab^*),$$

we see that

$$w_m = \frac{1}{16\pi} \mathbf{H} \cdot \mathbf{H}^* \quad (2.19)$$

is the mean density of magnetic energy and

$$w_e = \frac{\epsilon}{16\pi} \mathbf{E} \cdot \mathbf{E}^* \quad (2.20)$$

that of electric energy. We further recognize from Equation (2.18) that, if a region  $\Omega$  contains the same amount of magnetic and electric energy

$$\int_{\Omega} dV w_e = \int_{\Omega} dV w_m,$$

then the inflow and outflow of optical power coincide.

### 2.1.2 Vector Helmholtz Equations

Eliminating either the electric or magnetic field from Maxwell's equations we obtain the vector Helmholtz equations:

$$\nabla \times (\nabla \times \mathbf{E}) = k_0^2 \epsilon \mathbf{E} \quad (2.21)$$

$$\nabla \times \frac{1}{\epsilon} (\nabla \times \mathbf{H}) = k_0^2 \mathbf{H}. \quad (2.22)$$

Utilizing the vector identity  $\nabla \times (\nabla \times \mathbf{A}) = \nabla(\nabla \cdot \mathbf{A}) - \Delta \mathbf{A}$  and the reformulated Maxwell equation  $\nabla \cdot \mathbf{E} = -\nabla(\ln \epsilon) \cdot \mathbf{E}$  we can rewrite the vector Helmholtz equations in the alternative form:<sup>2</sup>

$$\Delta \mathbf{E} + k_0^2 \epsilon \mathbf{E} = -\nabla(\nabla \ln \epsilon \cdot \mathbf{E}) \quad (2.23)$$

$$\Delta \mathbf{H} + k_0^2 \epsilon \mathbf{H} = -\nabla(\ln \epsilon) \times \nabla \times \mathbf{H}. \quad (2.24)$$

In addition to the physical solutions of interest for us the vector Helmholtz equation supports nonphysical, spurious solutions such as magnetic monopoles ( $\nabla \cdot \mathbf{H} \neq 0$ ).

For homogeneous media ( $\nabla \epsilon = 0$ ), the right-hand sides of the vector Helmholtz Equations (2.23) and (2.24) vanish; i.e., the vector components are decoupled. Solutions of the vector Helmholtz equations without a longitudinal component of the electric or magnetic field are called *transverse* or *scalar waves*, any other solution is referred to as a *vector wave* or *hybrid solution*. The electric field of a scalar electric wave is accompanied via the Maxwell Equation (2.8) by the two transverse components of the magnetic field (e.g.,  $H_x, E_y, H_z$ ). Similarly, the magnetic field of a scalar magnetic wave is accompanied by the transverse components of the electric field (e.g.,  $E_x, H_y, E_z$ ). In inhomogeneous media ( $\nabla \epsilon \neq 0$ ), usually all components of the electric and magnetic field couple to each other via the right-hand side of Equations (2.23) and (2.24).

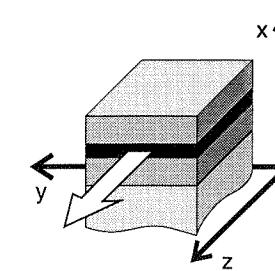


Figure 2.1. Layered medium.

### 2.1.3 Scalar Helmholtz Equations

In layered media we find two types of scalar waves, one transverse electric TE wave and one transverse magnetic TM wave. Both waves have in common that the field component, which is perpendicular to the layers, vanishes. For our further discussions, we will introduce the coordinate system as depicted in Figure 2.1. The  $yz$  plane of this coordinate system is defined by the interfaces between the dielectric layers, i.e.,

$$\nabla \epsilon = \frac{\partial \epsilon}{\partial x} \mathbf{e}_x,$$

the beam propagates mainly in the  $z$  direction. We then obtain the scalar Helmholtz equations

$$\nabla^2 E_y + k_0^2 \epsilon E_y = 0 \quad (2.25)$$

$$H_x = \frac{i}{k_0} \frac{\partial E_y}{\partial z}$$

$$H_z = -\frac{i}{k_0} \frac{\partial E_y}{\partial x}$$

for TE polarization and

$$\nabla^2 H_y - \frac{\partial \ln \epsilon}{\partial x} \frac{\partial H_y}{\partial x} + k_0^2 \epsilon H_y = 0 \quad (2.26)$$

$$E_x = -\frac{i}{k_0 \epsilon} \frac{\partial H_y}{\partial z}$$

$$E_z = \frac{i}{k_0 \epsilon} \frac{\partial H_y}{\partial x}$$

for TM polarization. Sometimes, the scalar Helmholtz equation for TM polarization is used in the alternative form

$$\left( \nabla \frac{1}{\epsilon} \nabla + k_0^2 \right) H_y = 0. \quad (2.27)$$

2. Throughout this book the symbol  $\Delta$  will stand for the vectorial Laplacian operator, and  $\nabla^2$  for the scalar Laplacian. It should be noted that both operators coincide only for Cartesian coordinate systems (see also Appendix D).

For integrated optical devices, large refractive index steps occur only at the chip surface, i.e., along the  $x$ -coordinate of our coordinate system. The lateral confinement of the optical field is achieved by small changes of the refractive index (see Figure 2.2). In this way, the situation shown in Figure 2.2 is typical for integrated optical devices, and in most cases we will use the scalar Helmholtz Equations (2.25) and (2.26) to derive the optical properties. For layered media both formulations of the scalar Helmholtz equation for TM polarization coincide. If the scalar Helmholtz equation is applied (as an approximation) to structures having a weak lateral inhomogeneity, the two forms will produce slightly different results.

To analyze the relative magnitudes of the field components, we consider a TE wave, i.e., an TE polarized eigenmode of the layered medium

$$E_y(x, y, z) = e_y(x, y)e^{iqz}$$

traveling in the  $z$ -direction. We then obtain the following expression for the ratio of the field components of the magnetic field

$$\left| \frac{H_z}{H_x} \right| \approx \left| \frac{\lambda}{2\pi} \frac{1}{E_y} \frac{\partial E_y}{\partial x} \right|;$$

i.e., the transverse field component  $H_x$  dominates over the longitudinal  $H_z$  as long as the amplitude is not changed significantly over lengths of  $\lambda/2\pi$ . This condition is usually satisfied in integrated optics. To summarize, *For a TE wave, the electric field is strictly perpendicular to the wave vector, the magnetic field is approximately perpendicular to the wave vector*. For TM waves, the electric and magnetic fields exchange their roles.

We will now try to get a physical idea of the hybrid character of the waves propagating through integrated optical devices. For this purpose, we consider the coupling mechanism described by the terms on the right-hand of vector Helmholtz Equations (2.23) and (2.24). We recognize that the coupling takes effect only in the immediate spatial vicinity of an inhomogeneity  $\nabla\epsilon \neq 0$ . For a more detailed examination, we consider a stratified medium with a TE-like wave propagating in

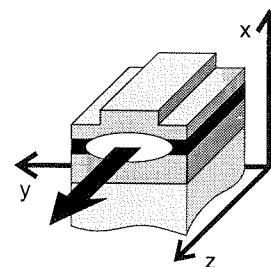


Figure 2.2. Typical integrated optical chip.

the  $z$ -direction, i.e.,  $\frac{\partial E_y}{\partial z} \approx ik_0\sqrt{\epsilon}E_y$ . In addition, the structure has a weak lateral inhomogeneity (see Figure 2.2). According to the vector Helmholtz Equation (2.23), we find the following relation for the non-dominating  $z$ -component of the electric field

$$\Delta E_z - k_0^2\epsilon E_z \approx -ik_0\sqrt{\epsilon} \frac{\partial \ln \epsilon}{\partial y} E_y$$

We expect that in addition to the dominating vector component  $E_y$ , the non-dominating field contributions  $E_x$  and  $E_z$  will occur at a lateral inhomogeneity. To estimate the coupling into the non-dominating contributions, we compare the magnitudes of the two coupling terms:<sup>3</sup>

$$\left| \frac{k_0\sqrt{\epsilon} \frac{\partial \ln \epsilon}{\partial y}}{k_0^2\epsilon} \right| = \left| \frac{\lambda}{2\pi} \frac{1}{\epsilon} \frac{\partial \epsilon}{\partial y} \right|.$$

Over the distance of one wavelength the solutions obtained using the vector Helmholtz equations depend only little on the behavior of the dielectric profile; i.e., the vector Helmholtz equation takes the mean over that region. The hybrid character of the waves will therefore be negligible, if the condition

$$\left| \frac{\Delta\epsilon}{\epsilon} \right| \approx 2 \left| \frac{\Delta n}{n} \right| \ll 2\pi \quad (2.28)$$

is satisfied. For small changes in the composition of the material, used for lateral confinement, this condition is excellently fulfilled. For confinement by etched ribs at the chip surface, however, the hybrid character of the waves becomes significant at the lateral boundary of the ribs. Such vector waves can be used to construct purely passive polarization converters.

### 2.1.4 Waves in Homogeneous Media

In a homogeneous medium ( $\epsilon = \text{const}$ ) the scalar Helmholtz equations for TE and TM polarization coincide, and we obtain

$$\nabla^2 \Psi + k_0^2 \epsilon \Psi = 0, \quad (2.29)$$

where  $\Psi(\mathbf{r})$  can be any Cartesian vector component of the electric or magnetic field. The plane wave

$$\Psi(\mathbf{r}) = e^{i\mathbf{q}\cdot\mathbf{r}} \quad (2.30)$$

with the dispersion relation

$$q^2 = \mathbf{q} \cdot \mathbf{q} = k_0^2 \epsilon \quad (2.31)$$

is a standard solution of the scalar Helmholtz equation. The wave vector  $\mathbf{q}$  is perpendicular to the phase fronts of the plane wave. Its magnitude is obviously

3. The following chapter comprises a much more detailed discussion of weakly guiding waveguides.

media dependent. We could have found this solution in a more elegant way by Fourier transformation of the scalar Helmholtz equation with respect to the three space coordinates

$$\Psi(\mathbf{q}) = \frac{1}{(2\pi)^{3/2}} \iiint_{-\infty}^{+\infty} d^3r e^{-i\mathbf{q}\cdot\mathbf{r}} \Psi(\mathbf{r})$$

which leads to

$$(\mathbf{q} \cdot \mathbf{q} - k_0^2 \epsilon) \Psi(\mathbf{q}) = 0.$$

This Fourier transformation in fact represents a decomposition of the entire field into a complete set of plane waves, which have all the same frequency but propagate in different directions.

Another interesting solution of the scalar Helmholtz equation in a homogeneous space is the spherical wave

$$\Psi(r) = \frac{e^{\pm iqr}}{r}, \quad (2.32)$$

which has a spherical symmetry ( $r = \sqrt{\mathbf{r} \cdot \mathbf{r}}$ ). Using the Laplace operator in spherical coordinates, we obtain

$$\nabla^2 \Psi = \frac{1}{r^2} \frac{\partial}{\partial r} \left( r^2 \frac{\partial \Psi}{\partial r} \right) = -q^2 \Psi,$$

i.e., the dispersion relation (2.31). The upper sign in the Equation (2.32) corresponds to an outgoing wave, whereas the lower sign denotes an incoming wave.

For both types of solution – plane and spherical waves – the local wave vector is perpendicular to the phase fronts and to both the electric and magnetic field. As an illustration Figure 2.3 shows the electric and magnetic fields along a straight line that runs parallel to the wave vector.

Later on, we will discuss another type of wave propagating through a homogeneous medium, the Gaussian beam, in much more detail. This wave is of particular interest for integrated optics, since it acts as an approximation for the focus of lenses on the one hand and of waveguide eigenmodes on the other hand.

### 2.1.5 Polarization

As we saw in the previous section, the electric field of a plane wave traveling in the  $z$ -direction has in general two (transverse) vector components with amplitudes  $A_x = |A| \cos \alpha$  and  $A_y = |A| \sin \alpha e^{-i\delta\varphi}$  which are phase shifted by  $\delta\varphi$  with respect to each other. Using the phase factor  $\varphi = k(z - z_0)$  of the electric field as the parameter describing the evolution of the wave, we obtain

$$\text{Re}(\mathbf{E}) = \frac{1}{2} (\mathbf{E} + \mathbf{E}^*) = |A| \begin{pmatrix} \cos \alpha \cos \varphi \\ \sin \alpha \cos(\varphi - \delta\varphi) \end{pmatrix} \quad (2.33)$$

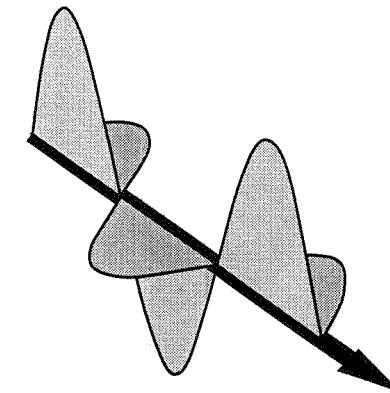


Figure 2.3. Electric and magnetic fields in a homogeneous medium.

for the real part of the electric field. Since both vector components are periodic functions of the parameter  $\varphi$ ,  $\text{Re}(\mathbf{E})$  will describe a closed curve. If  $\delta\varphi = m\pi$  this curve will reduce to a straight line. This scenario is called linear polarization. For  $\delta\varphi = (2m - 1)\pi/2$  we arrive at a standard ellipse, whose principal axes are parallel to the axes of our coordinate system. If, in addition, the amplitudes of both vector components have the same absolute value  $\alpha = \pi/4$ , the ellipse reduces to a circle. Such a wave is circularly polarized. From these results we assume that  $\text{Re}(\mathbf{E})$  will in general describe an ellipse (see Figure 2.4), which is rotated around its center.

To verify this assumption, we consider a standard ellipse with principal axes  $B_x = B \cos \beta$  and  $B_y = B \sin \beta$  and try to obtain the entire representation (2.33) of  $\text{Re}(\mathbf{E})$  for an arbitrary phase factor  $\varphi$  by rotation of this standard ellipse through an angle  $\psi$ , i.e.,

$$\begin{pmatrix} \cos \psi & \sin \psi \\ -\sin \psi & \cos \psi \end{pmatrix} \begin{pmatrix} \cos \alpha \cos \varphi \\ \sin \alpha \cos(\varphi - \delta\varphi) \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} \cos \beta \cos(\beta - \varphi_0) \\ \sin \beta \sin(\beta - \varphi_0) \end{pmatrix}$$

must hold for the coefficients of both  $\cos \varphi$  and  $\sin \varphi$  simultaneously. Solving the resulting four equations, we obtain

$$\tan(2\psi) = \tan(2\alpha) \cos(\delta\varphi) \quad (2.34)$$

$$\sin(2\beta) = \sin(2\alpha) \sin(\delta\varphi). \quad (2.35)$$

The calculations leading to (2.34) and (2.35) are elementary, but somewhat lengthy.

We have assumed here that the propagation constants of both vector components of the electric field coincide. This is not true in integrated optical waveguides, as we will see in Chapter 4. In consequence, we observe a continuous change of the polarization state of the optical wave between linear and elliptical polarization.

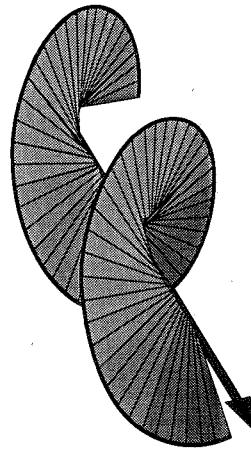


Figure 2.4. Electric field for elliptical polarization.

## 2.2 DIELECTRIC INTERFACES

### 2.2.1 Boundary Conditions

We will start the discussion of dielectric interfaces with an examination of the behavior of the optical field at a dielectric interface, i.e., at an abrupt transition between two different media  $I$  and  $II$  with dielectric constants  $\epsilon_I$  and  $\epsilon_{II}$ . To derive equations for the field components at the dielectric interface, we construct a flat box  $R$  with the surface  $\partial R$  that encloses a piece of the interface (see Figure 2.5). From Gauss divergence theorem we conclude that

$$\int_R dV \nabla \cdot \epsilon \mathbf{E} = \int_{\partial R} d\mathbf{F} \cdot \epsilon \mathbf{E} = 0,$$

where  $d\mathbf{F}$  is an (oriented) surface element of the box. With vanishing thickness of the box, the contributions from its side wall will also vanish. If we now orient the top and bottom surface of the box both being of size  $\partial R_{\parallel}$ , in the same way, we arrive at the equation

$$\int_{\partial R_{\parallel}} d\mathbf{F} (\epsilon_{II} \mathbf{E}_{II} - \epsilon_I \mathbf{E}_I) = 0,$$

where  $\mathbf{E}_I$  is the electric field on the medium  $I$  side of the dielectric interface and  $\mathbf{E}_{II}$  that on the medium  $II$  side. Since this equation holds for an arbitrary area  $\partial R_{\parallel}$ , we conclude that the normal components of the electric field are related by

$$\epsilon_{II} E_{\perp,II} = \epsilon_I E_{\perp,I}. \quad (2.36)$$

Next, we carry out the same procedure for the magnetic field and observe that the normal component of the magnetic field is continuous at the dielectric interface.

To derive equations for the tangential components of the electric and magnetic field at the dielectric interface we consider a flat strip  $A$  with the boundary  $\partial A$  that is perpendicular to the interface and encloses a piece of it (see Figure 2.5). Using Stokes's theorem, we now conclude that

$$\int_A d\mathbf{F} \cdot \nabla \times \mathbf{E} = \int_{\partial A} d\mathbf{r} \cdot \mathbf{E} = ik_0 \int_A d\mathbf{F} \cdot \mathbf{H},$$

where  $d\mathbf{r}$  is an (oriented) line element of the strip. With vanishing width of the strip, the contributions from both sides will vanish. Furthermore, the surface integral on the right-hand side of this equation will vanish provided the magnetic field  $\mathbf{H}$  is finite within the whole area  $A$  of the strip. Thus, we obtain

$$\int_{\partial A_{\parallel}} d\mathbf{r} (\mathbf{E}_{II} - \mathbf{E}_I) = 0,$$

where  $\partial A_{\parallel}$  is the path on the dielectric interface. Since this path can be chosen arbitrarily, we conclude that the tangential components of the electric field are continuous on the dielectric interface. In the same way, we can show that the tangential components of the magnetic field are continuous.

Up to now, we have shown that all field components except the normal component of the electric field are continuous at a dielectric interface. Next, we will derive conditions for the derivatives of the field components at the dielectric interface. For this purpose, we consider an arbitrary scalar function  $f$ , which satisfies the dielectric boundary condition

$$f_{II} = cf_I,$$

where  $c$  is constant over the whole interface. Its tangential derivatives are then related by

$$\nabla_{\parallel} f \Big|_{II} = c \nabla_{\parallel} f \Big|_I.$$

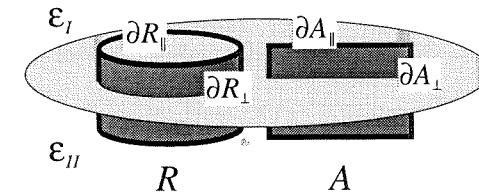


Figure 2.5. Boundary conditions.

Obviously, the tangential gradient  $\nabla_{\parallel}$  of a function is continuous, if and only if the function itself is continuous at the dielectric interface. Thus, the tangential derivatives of all field components except the normal component of the electric field are continuous. For the latter field component we obtain

$$\epsilon_{II} \nabla_{\parallel} E_{\perp} \Big|_{II} = \epsilon_I \nabla_{\parallel} E_{\perp} \Big|_I. \quad (2.37)$$

Using Maxwell's equations we will now elaborate conditions for the derivatives of the other field components. With

$$\nabla \cdot (\epsilon \mathbf{E}) \Big|_{II} = 0 = \nabla \cdot (\epsilon \mathbf{E}) \Big|_I$$

we obtain a condition for the normal derivative of the normal component of the electric field

$$\epsilon_{II} \nabla_{\perp} E_{\perp} \Big|_{II} - \epsilon_I \nabla_{\perp} E_{\perp} \Big|_I = (\epsilon_I - \epsilon_{II}) \nabla_{\parallel} \cdot \mathbf{E}_{\parallel}. \quad (2.38)$$

Utilizing an analog procedure, we observe that the normal derivative of the normal component of the magnetic field is continuous at the dielectric interface. Finally, we derive equations for the normal derivatives of the tangential field components at the interface. From the dielectric boundary condition  $\mathbf{H}_{II} = \mathbf{H}_I$  and Maxwell's equations, we obtain

$$\nabla \times \mathbf{E} \Big|_{II} = \nabla \times \mathbf{E} \Big|_I.$$

For our further discussion, we introduce a local coordinate system having two tangential coordinates and a third one that is parallel to the normal vector of the dielectric interface. We then obtain two equations

$$[\nabla_{\parallel} E_{\perp} - \nabla_{\perp} E_{\parallel}]_{II} = [\nabla_{\parallel} E_{\perp} - \nabla_{\perp} E_{\parallel}]_I.$$

Utilizing the dielectric boundary conditions for these tangential derivatives, we obtain

$$\nabla_{\perp} E_{\parallel} \Big|_{II} - \nabla_{\perp} E_{\parallel} \Big|_I = \frac{(\epsilon_{II} - \epsilon_I)}{\epsilon_I} \nabla_{\parallel} E_{\perp} \Big|_I. \quad (2.39)$$

Thus, the normal derivatives of the tangential components of the electric field are continuous only if the normal component of the electric field is constant along the interface. This condition, in turn, can be satisfied only if the normal component of the electric field is everywhere zero. From the dielectric boundary condition  $E_{\parallel,II} = E_{\parallel,I}$ , we conclude that the tangential components of the curl of the magnetic fields are related by

$$\frac{1}{\epsilon_{II}} (\nabla \times \mathbf{H})_{\parallel} \Big|_{II} = \frac{1}{\epsilon_I} (\nabla \times \mathbf{H})_{\parallel} \Big|_I.$$

In consequence, we obtain two slightly different equations

$$\frac{1}{\epsilon_{II}} [\nabla_{\parallel} H_{\perp} - \nabla_{\perp} H_{\parallel}]_{II} = \frac{1}{\epsilon_I} [\nabla_{\parallel} H_{\perp} - \nabla_{\perp} H_{\parallel}]_I$$

**Table 2.1**  
Continuous (–) and discontinuous (•) dielectric boundary conditions for the vector components of the electric and magnetic fields.

	$\mathbf{E}_{\parallel}$	$\mathbf{E}_{\perp}$	$\mathbf{H}_{\parallel}$	$\mathbf{H}_{\perp}$
1	–	•	–	–
$\nabla_{\parallel}$	–	•	–	–
$\nabla_{\perp}$	○	•	•	–

and end up with the dielectric boundary conditions for the normal derivative of the tangential field components of the magnetic field

$$\frac{1}{\epsilon_{II}} \nabla_{\perp} \mathbf{H}_{\parallel} \Big|_{II} - \frac{1}{\epsilon_I} \nabla_{\perp} \mathbf{H}_{\parallel} \Big|_I = \left( \frac{1}{\epsilon_I} - \frac{1}{\epsilon_{II}} \right) \nabla_{\parallel} H_{\perp} \Big|_I. \quad (2.40)$$

In general, it is clear that neither the normal derivative of the tangential components of the electric field nor those of the magnetic field are continuous at a dielectric interface.

Let us finally summarize the rather formal, but important results of this section. Table 2.1 represents the dielectric boundary conditions. Within that table, a bullet (•) stands for a quantity that is discontinuous at the interface, while a minus (–) designates a continuous vector component. The open circle (○) denotes the component that can be both continuous or discontinuous at the interface. In inspecting the table, we recognize an asymmetry between the number of discontinuous components for the electric and magnetic fields. This asymmetry is due to the restriction to nonmagnetic materials that we made at the beginning of this chapter.

## 2.2.2 Reflection and Refraction

We will now turn our interest to the behavior of a plane wave at a planar dielectric interface between two media *in* and *out* with dielectric constants  $\epsilon_{in}$  and  $\epsilon_{out}$ . Figure 2.6 shows the plane of incidence, which is defined by the normal vector of the dielectric interface and the wave vector  $\mathbf{q}_{in}$  of the incident wave. The wave vectors  $\mathbf{q}_{refl}$  and  $\mathbf{q}_{out}$  for the reflected and the transmitted wave respectively are also located in the plane of incidence. Figure 2.6 implies that continuity of the tangential

field components at the dielectric interface can be obtained only if the tangential components of the wave vectors of all waves coincide. If we elaborate these equations, we obtain the law of reflection

$$\theta_{\text{refl}} = \theta_{\text{in}} \quad (2.41)$$

and Snell's law of refraction

$$\sqrt{\epsilon_{\text{out}}} \sin \theta_{\text{out}} = \sqrt{\epsilon_{\text{in}}} \sin \theta_{\text{in}}. \quad (2.42)$$

On the basis of these equations, we can determine the direction of propagation of the reflected and the transmitted wave.

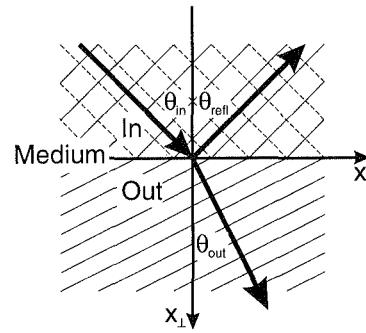


Figure 2.6. Reflection and refraction of a plane wave at a planar dielectric interface.

In a second step, we will derive formulas for the amplitudes of the reflected and refracted wave at the dielectric interface. We shall consider two largely different linear polarizations, namely s- and p-polarization. For s-polarization, the electric field is perpendicular (in German *senkrecht*) to the plane of incidence, for p-polarization the electric field is parallel to the plane of incidence. To allow a unified treatment of both polarizations, we shall now define the characteristic admittance  $\eta$  via the normal component of Poynting's vector

$$S_{\perp} = \frac{c}{8\pi} [\mathbf{E} \times \mathbf{H}^*]_{\perp} = \frac{c}{8\pi} \eta |\mathbf{E}_{\parallel}|^2, \quad (2.43)$$

i.e., via the power flux through the dielectric interface. Using Maxwell's equations we obtain

$$\eta = \begin{cases} \sqrt{\epsilon - \epsilon_{\text{in}} \sin^2 \theta_{\text{in}}} & \text{for s-polarization} \\ \epsilon / \sqrt{\epsilon - \epsilon_{\text{in}} \sin^2 \theta_{\text{in}}} & \text{for p-polarization.} \end{cases} \quad (2.44)$$

Using the reflection and transmission coefficients, i.e., the relative amplitudes for the reflected electric field  $r = E_r/E_i$  and for the transmitted electric field  $t = E_t/E_i$  we can now write

$$\begin{aligned} t &= 1 + r \\ \eta_{\text{out}} t &= \eta_{\text{in}} (1 - r) \end{aligned} \quad (2.45)$$

for the dielectric boundary conditions of the tangential electric and magnetic field components at the dielectric interface. Solving these equations we arrive at Fresnel's formulas:

$$r = \frac{\eta_{\text{in}} - \eta_{\text{out}}}{\eta_{\text{in}} + \eta_{\text{out}}} \quad (2.46)$$

for the reflection coefficient and

$$t = \frac{2\eta_{\text{in}}}{\eta_{\text{in}} + \eta_{\text{out}}} \quad (2.47)$$

for the transmission coefficient. The reflectance, i.e., the relative reflected power, is given by

$$R = \frac{\text{Re}(S_{\perp}^{\text{refl}})}{\text{Re}(S_{\perp}^{\text{in}})} = |r|^2. \quad (2.48)$$

In the same way, the transmittance, i.e., the relative transmitted power, is determined by

$$T = \frac{\text{Re}(S_{\perp}^{\text{out}})}{\text{Re}(S_{\perp}^{\text{in}})} = \frac{\text{Re}(\eta_{\text{out}})}{\text{Re}(\eta_{\text{in}})} |t|^2. \quad (2.49)$$

Obviously, we can also formulate the above equations using the characteristic impedance  $\zeta = 1/\eta$ .

We will now discuss Fresnel's formula in some more detail. For lossless media ( $\sqrt{\epsilon} > 0$ ) the characteristic admittance  $\eta_{\text{in}}$  is always a real quantity. If the characteristic admittance of the output medium is a real quantity – we will see very soon that this is not always true – both reflection and transmission coefficients are also real quantities. Thus, no phase shift will occur between the incident, reflected and transmitted waves at the dielectric interface. The transmittance is then given by

$$T = \frac{4\eta_{\text{in}}\eta_{\text{out}}}{(\eta_{\text{in}} + \eta_{\text{out}})^2} = 1 - R. \quad (2.50)$$

Equation (2.50) can be regarded as the conservation law for optical power at the interface.

If the output medium is less dense than the input medium ( $\sqrt{\epsilon_{\text{out}}} < \sqrt{\epsilon_{\text{in}}}$ ) or if it is an ideal metal, i.e.,  $\epsilon$  is real and  $\epsilon \ll 0$ , the characteristic admittance of the output medium will become purely imaginary at a critical angle, which is defined by

$$\sin \theta_c = \sqrt{\frac{\epsilon_{\text{out}}}{\epsilon_{\text{in}}}}. \quad (2.51)$$

Snell's law of refraction implies that total reflection, i.e., no refraction, will occur beyond that angle ( $\theta > \theta_c$ ). In the regime of total reflection the reflection coefficient (2.46) can be written as

$$r = e^{i\phi_r}, \quad (2.52)$$

in which  $\phi_r$  is a pure phase shift. Thus, reflection becomes perfect ( $R = 1$ ). Incidentally, the transmittance as defined in Equation (2.49) also vanishes. This indicates that we will find only an evanescent wave on the output side of the dielectric interface. The phase shift between incident and reflected wave at the dielectric interface is given by

$$\phi_r = \begin{cases} -2 \tan^{-1}(\eta_{\text{out}}/\eta_{\text{in}}) & \text{for s-polarization} \\ \pi - 2 \tan^{-1}(\eta_{\text{in}}/\eta_{\text{out}}) & \text{for p-polarization.} \end{cases} \quad (2.53)$$

It is polarization dependent; i.e., linear polarized light is converted into elliptically polarized light, provided the former is not s- or p-polarized.

One more consequence of the phase shift between the incident and the reflected wave is that finite beams will be laterally shifted. This phenomenon is known as the Goos-Hänchen shift (see Figure 2.7). This shift can be calculated using the angular plane-wave decomposition of incident and reflected beams at the interface, i.e.,

$$\begin{aligned} E_i(r_{\parallel}, r_{\perp} = 0) &= \int_{-\infty}^{+\infty} d\beta E_i(\mathbf{q}, r_{\perp} = 0) e^{i\mathbf{q}\mathbf{r}_{\parallel}} \\ E_r(r_{\parallel}, r_{\perp} = 0) &= \int_{-\infty}^{+\infty} d\beta E_i(\mathbf{q}, r_{\perp} = 0) e^{i(\mathbf{q}\mathbf{r}_{\parallel} + \phi_r)}. \end{aligned}$$

The beam is assumed to be a nearly plane wave with a wave vector  $(\mathbf{q}_{\parallel}, q_{\perp})$ ; i.e., its optical power is then concentrated at  $\mathbf{q} = \mathbf{q}_{\parallel}$ . On the basis of Taylor's expansion of the phase shift at this point

$$\phi_r(\mathbf{q}) = \phi_r(\mathbf{q}_{\parallel}) + (\mathbf{q} - \mathbf{q}_{\parallel}) \cdot \nabla_q \phi_r|_{\mathbf{q}_{\parallel}} + \dots \quad (2.54)$$

we see that the Goos-Hänchen shift is given by

$$d_{\text{GH}} = |\nabla_q \phi_r|_{\mathbf{q}_{\parallel}}. \quad (2.55)$$

The shape of beams having a wide angle spectrum will be distorted. We will point out here that *the Goos-Hänchen shift is caused by the phase shift at reflection*. Therefore, this shift is associated with interference filters and Bragg gratings.

From Equation (2.48) we recognize that the reflectivity at a dielectric interface vanishes if the equation  $\eta_{\text{out}} = \eta_{\text{in}}$  holds. For s-polarization this condition implies  $\epsilon_{\text{out}} = \epsilon_{\text{in}}$ , i.e., that there is no dielectric interface. For p-polarization, however, we always find an angle – Brewster's angle  $\theta_B$  – where the reflectance vanishes. It is determined by

$$\tan \theta_B = \sqrt{\frac{\epsilon_{\text{out}}}{\epsilon_{\text{in}}}}. \quad (2.56)$$

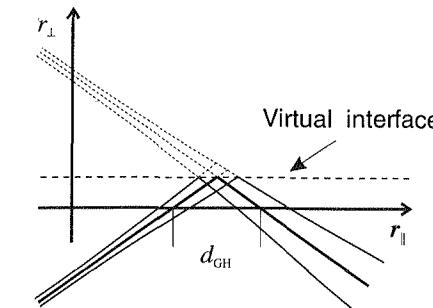


Figure 2.7. Goos-Hänchen shift.

It is easy to show that Brewster's angle is always smaller than the limiting angle for total reflection.

Figure 2.8 shows as an example the reflectance of a glass-air interface as a function of the angle of incidence for a plane wave, which is incident from the glass side  $n_{\text{in}} = 1.52$ . Since the output medium is denser, beyond  $\theta_c = 41.1^\circ$ , total reflection will occur for s- and p-polarization. For p-polarization, Brewster's angle is found at  $\theta_B = 33.3^\circ$ . At normal incidence the reflectance coincides for both polarizations. This behavior is expected, since characteristic admittance reduces to the refractive index at normal incidence  $q_{\parallel} = 0$ .

### 2.2.3 Fabry-Perot Interferometer

In this section, we will examine the simplest layered structure, a single homogeneous layer. We will derive expressions for the reflection coefficient by summing the amplitudes of all partial reflections and refractions. This procedure is called *Airy's summation*.

We denote the coefficients describing the reflection and transmission at the first interface by  $r_{\text{in}}$  and  $t_{\text{in}}$ , the coefficients for the second interface by  $r_{\text{out}}$  and  $t_{\text{out}}$ . It is important to distinguish carefully between the reflection and transmission coefficients  $(r, t)$  for the outgoing waves and those  $r^{(-)}, t^{(-)}$  for the returning waves. Using Fresnel's formulas we easily show that

$$r^{(-)} = -r \quad (2.57)$$

$$tt^{(-)} - rr^{(-)} = 1. \quad (2.58)$$

The influence of passing a homogeneous layer of thickness  $d$  and dielectric constant  $\epsilon$  is described by the equivalent phase thickness

$$\delta = k_0 d \sqrt{\epsilon - \epsilon_{\text{in}} \sin^2 \theta_{\text{in}}}. \quad (2.59)$$

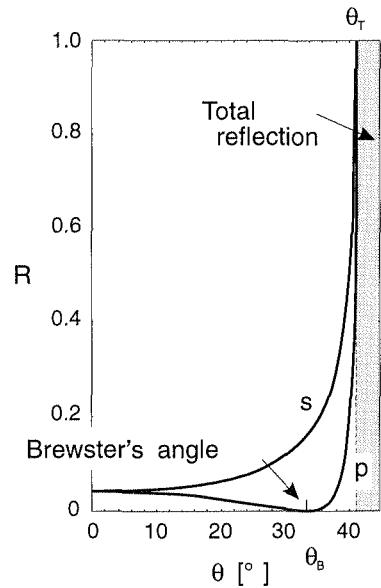


Figure 2.8. Reflectance vs. angle of incidence for a glass-air interface.

For transparent media and angles below the critical angle of total reflection, the coefficient  $e^{i\delta}$  will be a pure phase factor, whereas for lossy material, attenuation  $|e^{i\delta}| < 1$  will take place. Keeping these notations in mind and summing up the contributions for the multiple reflections we obtain the following relation for the reflection coefficient of the Fabry-Perot interferometer:

$$\begin{aligned} r_{\text{FP}} &= r_{\text{in}} + t_{\text{in}}r_{\text{out}}t_{\text{in}}^{(-)}e^{2i\delta} \sum_{l=0}^{\infty} \left( r_{\text{in}}^{(-)}r_{\text{out}}e^{2i\delta} \right)^l \\ &= r_{\text{in}} + \frac{t_{\text{in}}t_{\text{in}}^{(-)}r_{\text{out}}e^{2i\delta}}{1 - r_{\text{in}}^{(-)}r_{\text{out}}e^{2i\delta}} \\ &= \frac{r_{\text{in}} + r_{\text{out}}e^{2i\delta}}{1 + r_{\text{in}}r_{\text{out}}e^{2i\delta}} \end{aligned} \quad (2.60)$$

Figure 2.9 shows the first view partial waves of the Airy summation.

If we restrict ourselves to symmetric arrangements  $\eta_{\text{out}} = \eta_{\text{in}}$ , i.e.,  $r = -r_{\text{in}} = r_{\text{out}}$  we arrive at a general equation for the reflectance of a Fabry-Perot interferometer:

$$R_{\text{FP}} = |r_{\text{FP}}|^2 = \left| \frac{r[1 - \exp(2i\delta)]}{1 - |r|^2 \exp(2i(\phi_r + \delta))} \right|^2 \quad (2.61)$$

where  $\phi_r$  denote the phase angles of  $r$ . For lossless media and angles below the critical angle of total reflection ( $|e^{i\delta}| = 1, \phi_r = 0$ ), Equation (2.61) simplifies further

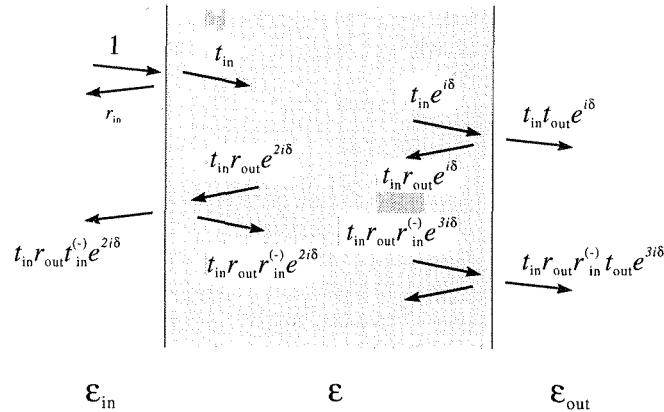


Figure 2.9. The first view partial waves of Airy's summation.

to

$$R_{\text{FP}} = \frac{F^2 \sin^2 \delta}{(2\pi)^2 + F^2 \sin^2 \delta} \quad (2.62)$$

where

$$F = \frac{\pi \sqrt{R}}{1 - R} \quad (2.63)$$

designates the finesse of the Fabry-Perot interferometer. It represents a measure for the spectral resolution of the device. Obviously, no reflection occurs if the condition  $\delta = m\pi$  is satisfied. A more detailed analysis of Equation (2.61) shows that the transmittance does not completely vanish even in the regime of total reflection. This phenomenon is called *tunneling*, in conformity to the terminology of quantum mechanics.

## 2.2.4 Interference Filters

In the last section we used Airy's summations to analyze the Fabry-Perot interferometer, which is in fact the most elementary interference filter. We could obviously use the same procedure for more complicated filters, too. However, for such cases we will derive an alternative approach based on characteristic matrices.

Using the fundamental solutions of Maxwell's equations in a homogeneous space, which takes both forward and reverse waves into account, we find

$$\begin{pmatrix} E_{\text{in}} \\ iH_{\text{in}} \end{pmatrix} = \mathcal{M} \begin{pmatrix} E_{\text{out}} \\ iH_{\text{out}} \end{pmatrix}. \quad (2.64)$$

The characteristic matrix  $\mathcal{M}$  with the matrix elements  $m_{ij}$  describing a single ho-

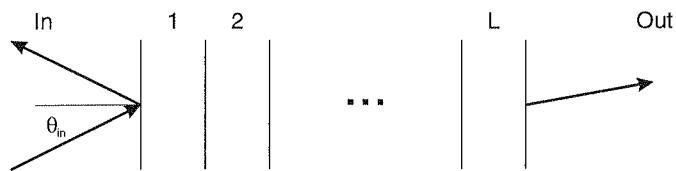


Figure 2.10. Schematic drawing of an interference filter.

mogeneous layer of thickness  $d$  and the dielectric constant  $\epsilon$  is given by<sup>4</sup>

$$\mathcal{M} = \begin{pmatrix} \cos \delta & -\sin \delta / \eta \\ \eta \sin \delta & \cos \delta \end{pmatrix} = \begin{pmatrix} \cosh(i\delta) & -\sinh(i\delta)/(i\eta) \\ -i\eta \sinh(i\delta) & \cosh(i\delta) \end{pmatrix} \quad (2.67)$$

with the characteristic admittance  $\eta$ . The phase thickness is again

$$\delta = k_0 d \sqrt{\epsilon - \epsilon_{in} \sin^2 \theta_{in}}$$

(see Equation (2.59)) where  $\theta_{in}$  is the angle of incidence and  $\epsilon_{in}$  the dielectric constant of the input medium. Since the tangential components of both the electric and the magnetic field are continuous at a dielectric interface, we can describe an arbitrary interference filter (see Figure 2.10) consisting of  $L$  layers by stacking up the characteristic matrices, i.e.,

$$\mathcal{M} = \prod_{l=1}^L \mathcal{M}_l = \mathcal{M}_1 \mathcal{M}_2 \dots \mathcal{M}_L. \quad (2.68)$$

For the definition of the admittance  $Y$ , the interference filter is regarded as an effective medium; i.e., the filter admittance is defined by the equation

$$\begin{pmatrix} E_{in} \\ iY E_{in} \end{pmatrix} = \mathcal{M} \begin{pmatrix} E_{out} \\ i\eta_{out} E_{out} \end{pmatrix}, \quad (2.69)$$

which has the solution

$$iY = \frac{m_{21} + im_{22}\eta_{out}}{m_{11} + im_{12}\eta_{out}}. \quad (2.70)$$

4. Sometimes it makes sense to use the “inverse” formulation, i.e.,

$$\begin{pmatrix} H_{in} \\ iE_{in} \end{pmatrix} = \tilde{\mathcal{M}} \begin{pmatrix} H_{out} \\ iE_{out} \end{pmatrix}. \quad (2.65)$$

which is based on the characteristic impedance  $\zeta = 1/\eta$  and

$$\tilde{\mathcal{M}} = \begin{pmatrix} \cos \delta & -\sin \delta / \zeta \\ \zeta \sin \delta & \cos \delta \end{pmatrix} = \begin{pmatrix} \cosh(i\delta) & -\sinh(i\delta)/(i\zeta) \\ -i\zeta \sinh(i\delta) & \cosh(i\delta) \end{pmatrix} \quad (2.66)$$

Matching the tangential field components at the first dielectric interface of the interference filter, we obtain the following relation for the reflection coefficient:

$$r_{IF} = \frac{\eta_{in} - Y}{\eta_{in} + Y}. \quad (2.71)$$

Obviously, we can utilize Equations (2.47), (2.48) and (2.49) for the transmission coefficient, reflectance and transmittance, respectively, if we merely replace the admittance of the output medium  $\eta_{out}$  by that of the interference filter  $Y$ .

We will briefly list here some properties of lossless interference filters which are characterized by a real dielectric function with  $\epsilon > 0$ . A detailed examination may be found in the textbooks by Mcleod [1] and Thelen [2].

### Properties of the Characteristic Matrix

The characteristic matrix is a unimodular 2x2-matrix, i.e.,  $\det \mathcal{M} = 1$ .

The unimodular matrices form a group, hence it is sufficient to show this property for a single layer. For a lossless single layer we find

$$\det \mathcal{M} = \begin{cases} \cos^2 \delta + \sin^2 \delta & \text{for } \epsilon_{in} \sin^2 \theta_{in} \leq \epsilon \\ \cosh^2(i\delta) - \sinh^2(i\delta) & \text{for } \epsilon_{in} \sin^2 \theta_{in} > \epsilon, \end{cases} \quad (2.72)$$

i.e., we arrive in both cases at  $\det \mathcal{M} = 1$ . The inverse of a unimodular matrix is

$$\begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}^{-1} = \begin{pmatrix} m_{22} & -m_{12} \\ -m_{21} & m_{11} \end{pmatrix}, \quad (2.73)$$

the eigenvalues are

$$\mu_{1/2} = \frac{\text{Tr } \mathcal{M}}{2} \pm \sqrt{\left(\frac{\text{Tr } \mathcal{M}}{2}\right)^2 - 1}. \quad (2.74)$$

### Effect of Quarter- and Half-Wave Layers ( $\delta = N\pi/2$ )

For particular wavelengths, where the optical thickness is an integral number of quarter-waves, the characteristic matrix is given by

$$\mathcal{M} = \begin{cases} (-1)^{N/2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \text{for even } N \\ (-1)^{(N+1)/2} \begin{pmatrix} 0 & 1/\eta_j \\ -\eta_j & 0 \end{pmatrix} & \text{for odd } N, \end{cases} \quad (2.75)$$

i.e., half-wave layers are optically transparent. A convenient notation is obtained by defining interference filters at a design wavelength in terms of quarter-wave layers. The characters  $H$ ,  $L$  and  $M$  usually refer to quarter-wave layers of high, low and intermediate refractive index, respectively.

## Periodic Layered Media

The characteristic matrix of an  $N$ -period multilayer is given by

$$\mathcal{M}^N = U_{N-1} \left( \frac{\text{Tr } \mathcal{M}}{2} \right) \mathcal{M} - U_{N-2} \left( \frac{\text{Tr } \mathcal{M}}{2} \right) \mathcal{E}, \quad (2.76)$$

where  $\mathcal{E}$  is the unity matrix.  $U_N(x)$  designate the Chebyshev polynomials of second kind, i.e.,

$$U_N(\cos \theta) = \frac{\sin((N+1)\theta)}{\sin \theta}. \quad (2.77)$$

Spectral regions with  $|\text{Tr } \mathcal{M}| > 2$  are called *stopbands*, the other regions are denoted as *passbands*. We will see that *an infinite periodic stack of multilayers ( $N \rightarrow \infty$ ) as a perfect reflector ( $R = 1$ )* acts within the stopbands. The widths of the stopbands represent a measure of the refractive index contrast of the basic period. With decreasing contrast, the width of a stopband tends to zero. A necessary condition for the occurrence of a stopband is therefore

$$q_l = \frac{2\pi}{\Lambda_l} \sqrt{\epsilon - \epsilon_{\text{in}} \sin^2 \theta_{\text{in}}} = \frac{l\pi}{\Lambda} \quad (2.78)$$

in which  $\Lambda$  is the period, i.e., the thickness of the basic period. If, however, any of the layers forming the basic period is transparent for a certain order  $l$ , the stopband cannot be observed. The stack  $(HL)^N$ , for example, has no even orders for that reason. In our discussion of contradirectional coupling we will show (Chapter 7), that this equation is in fact the Bragg condition that governs any contradirectional coupler.

The envelopes of the filter curve are defined by the envelopes of all multilayers which can be formed using one basic period. Figure 2.11 shows the reflectance as a function of the wavelength. One stopband and the two envelopes are indicated. To prove the (emphasized) theorem, we take advantage of the invariance of  $\text{Tr } \mathcal{M}$  and  $\det \mathcal{M}$  with respect to unitary transformations and represent both the trace  $\text{Tr } \mathcal{M} = \mu_1 + \mu_2$  and the determinant  $\det \mathcal{M} = \mu_1 \mu_2$  in terms of the eigenvalues  $(\mu_1, \mu_2)$  of the characteristic matrix  $\mathcal{M}$ . If the condition  $\text{Tr } \mathcal{M} > 1$  is satisfied, we can conclude that either  $|\mu_1| > 1$  or  $|\mu_2| > 1$  holds. Thus the trace of the stacked matrix

$$\lim_{N \rightarrow \infty} |\text{Tr } \mathcal{M}^N| = \lim_{N \rightarrow \infty} |\mu_1^N + \mu_2^N| \rightarrow \infty$$

tends to infinity. Using the polar representation of  $Y$  we see that this leads, in turn, to a purely imaginary admittance and finally to total reflection ( $R = 1$ ).

## Symmetric Layer Stacks

A symmetric stack of layers  $\mathcal{M} = \mathcal{M}_1 \mathcal{M}_2 \dots \mathcal{M}_L \dots \mathcal{M}_2 \mathcal{M}_1$  can be replaced by a single equivalent layer with the characteristic matrix<sup>5</sup>

$$\mathcal{M} = \begin{pmatrix} \cos \Delta & \sin \Delta / H \\ -H \sin \Delta & \cos \Delta \end{pmatrix}. \quad (2.79)$$

5. This equivalent layer is, of course, highly dispersive.

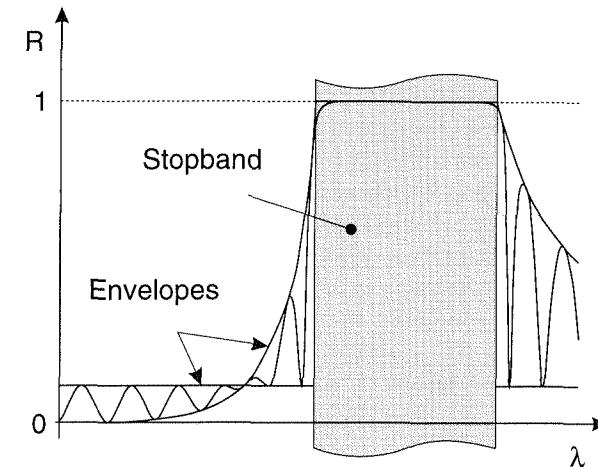


Figure 2.11. Filter characteristic, envelopes and stopband of an interference filter.

The equivalent phase thickness  $\Delta$  is given by

$$\cos \Delta = m_{11}. \quad (2.80)$$

It is either purely real or purely imaginary. The equivalent admittance  $H$  is defined by

$$H = \sqrt{-\frac{m_{21}}{m_{12}}}. \quad (2.81)$$

To prove the existence of such a representation, we consider a symmetric three-layer stack and show only that  $m_{11} = m_{22}$ . The characteristic matrix of an  $N$ -period multilayer is

$$\mathcal{M} = \begin{pmatrix} \cos(N\Delta) & \sin(N\Delta)/H \\ -H \sin(N\Delta) & \cos(N\Delta) \end{pmatrix}. \quad (2.82)$$

For the symmetric layer stack there is an analytic expression for the two envelopes. One envelope

$$R_H = \left( \frac{\eta_{\text{in}} \eta_{\text{out}} - H^2}{\eta_{\text{in}} \eta_{\text{out}} + H^2} \right)^2, \quad (2.83)$$

while the other one,

$$R_L = \left( \frac{\eta_{\text{in}} - \eta_{\text{out}}}{\eta_{\text{in}} + \eta_{\text{out}}} \right)^2, \quad (2.84)$$

is a straight line giving the reflectance at an interface between the input and the output medium. If both media coincide, this envelope lies at  $R_L = 0$ . If the input and output media are different, both envelopes can cross each other. Figure 2.11 shows

an example of the filter curve, both envelopes and one stopband. Due to its poor crosstalk attenuation, the filter of Figure 2.11 is useless for practical applications.

Figure 2.12 shows a simple design and a schematic diagram of the filter curve for the four most important types of interference filters – anti-reflection coating, short-wave-pass, long-wave-pass and bandpass. It should be pointed out that the edge filters – long-wave pass and short-wave pass – should be operated at the fundamental order since the behavior of the filter curve close to the stopband changes at higher orders. More sophisticated designs can be obtained by refining the filter curves of such simple designs, i.e., by varying the thickness of certain layers to reduce the ripples of the filter curve in its passband. A more detailed discussion of this topic is given in the textbooks by Mcleod [1] and Thelen [2].

## 2.3 GEOMETRICAL OPTICS

In this section we will discuss the conditions and methods of deriving a ray representation of the optical field. In physical terms, we will show how to switch from an entire field theory describing the evolution of the electric and magnetic field  $\mathbf{E}(\mathbf{r}, t)$  and  $\mathbf{H}(\mathbf{r}, t)$  to a particle theory dealing with rays that are regarded as trajectories of virtual particles.

### 2.3.1 Eikonal Equation

The whole procedure is restricted mainly by the condition that the wavelength inside the medium must be significantly smaller than the smallest characteristic length  $L$  of both the optical field and the inhomogeneities, i.e.,

$$\bar{q}L \gg 1$$

with  $\bar{q}$  being the mean wave number in the inhomogeneous medium. If this condition is satisfied, we expect that the optical field will look locally like a plane wave. We therefore expand the electric field into a power series (Debye expansion)

$$\mathbf{E} = e^{ik_0 S(\mathbf{r})} \sum_{m=0}^{\infty} \left( \frac{1}{ik_0} \right)^m \mathbf{E}_m(\mathbf{r}), \quad (2.85)$$

which delivers locally plane waves. In Equation (2.85) the factor  $1/ik_0$  acts as smallness parameter in the sense mentioned previously. The exponent  $k_0 S(\mathbf{r})$  describes the phase portrait of the optical field; i.e., the equation  $S(\mathbf{r}) = \text{const}$  can be used to define its phase fronts and  $\nabla S(\mathbf{r})$  acts as the local wave vector.

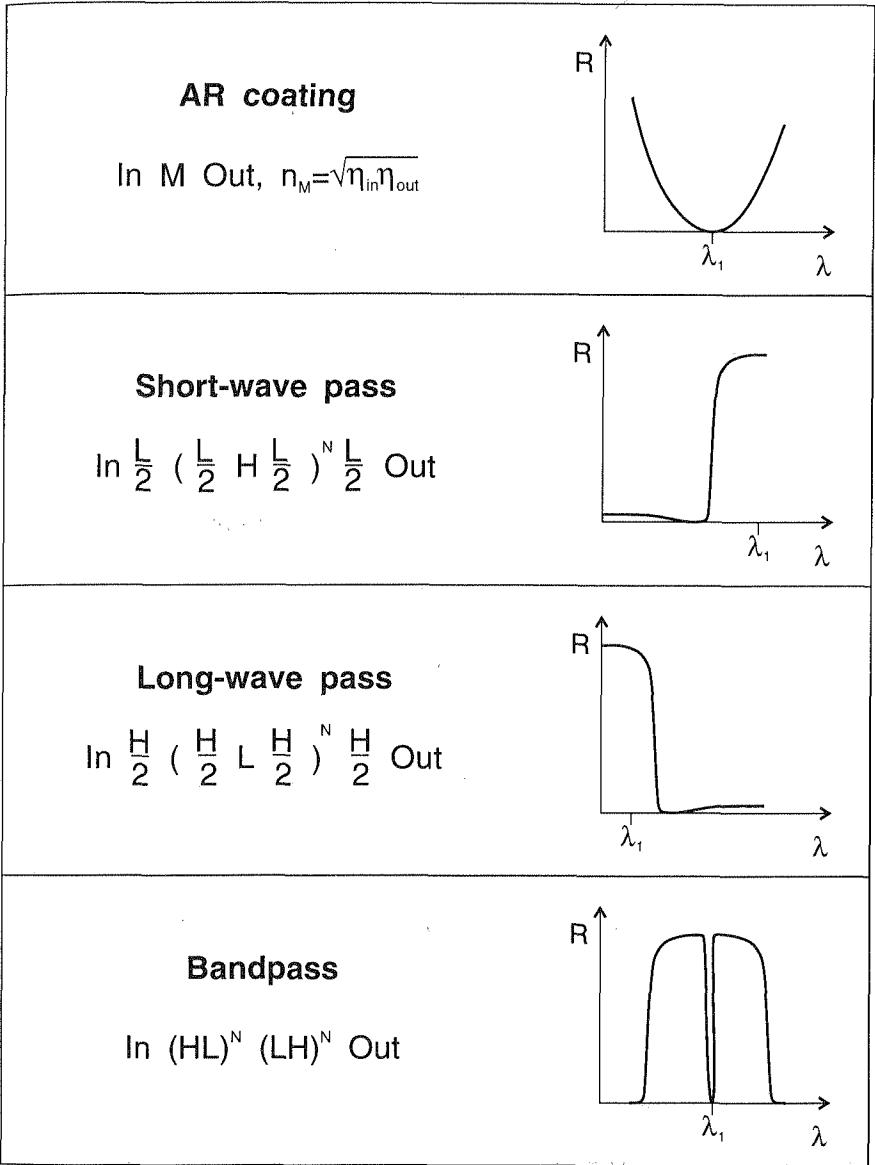


Figure 2.12. Several types of interference filter.

If we repeat the same procedure for the magnetic field, use Maxwell's equations and keep in mind that

$$\nabla \times \mathbf{E} = e^{ik_0 S(\mathbf{r})} \sum_{m=0}^{\infty} \left( \frac{1}{ik_0} \right)^m [ik_0 \nabla S(\mathbf{r}) + \nabla] \times \mathbf{E}_m(\mathbf{r})$$

then we can derive a hierarchy of first order partial differential equations, which corresponds to the powers  $m = 0, \dots, \infty$  of the smallness parameter  $1/ik_0$ . We obtain

$$\begin{aligned} \nabla S \times \mathbf{H}_m + \epsilon \mathbf{E}_m + \nabla \times \mathbf{H}_{m-1} &= 0 \\ \nabla S \times \mathbf{E}_m - \mathbf{H}_m + \nabla \times \mathbf{E}_{m-1} &= 0 \end{aligned} \quad (2.86)$$

with  $\mathbf{E}_{-1}(\mathbf{r}) = \mathbf{H}_{-1}(\mathbf{r}) = 0$ . These two leading equations are used to determine  $S$ ,  $\mathbf{E}_0$  and  $\mathbf{H}_0$ . The higher order equations are used to successively determine  $\mathbf{E}_m$  and  $\mathbf{H}_m$ . However, we should clearly point out here that Debye's expansion is an asymptotic series. Thus, the accuracy of the results cannot be improved by taking an increasing number of contributions into account.

After this brief discussion of Debye's expansion, we start to solve the two leading equations of the hierarchy (2.86). By eliminating  $\mathbf{H}_0$ , we obtain a set of linear equations

$$\mathcal{P} \mathbf{E}_0 = \nabla S \times (\nabla S \times \mathbf{E}_0) + \epsilon \mathbf{E}_0 = 0 \quad (2.87)$$

for the leading contribution  $\mathbf{E}_0$  of the electric field. The solution of the homogeneous set of linear equations requires that the condition

$$\det \mathcal{P} = -\epsilon [(\nabla S)^2 - \epsilon] = 0$$

is satisfied. This equation has two essentially different solutions. One ( $\epsilon = 0$ ) corresponds to a longitudinal wave and a second one

$$(\nabla S)^2 = \epsilon(\mathbf{r}) \quad (2.88)$$

to a transverse wave. Equation (2.88), which describes evolution of the phase portrait, is denoted as an eikonal equation, in which the Greek word *eikon* refers to an image.

### 2.3.2 Ray Equation and Fermat's Principle

To derive the ray equation, we take advantage of the fact that the eikonal equation is a Hamilton-Jacobi equation, i.e., that it is a first order partial differential equation of the form

$$\begin{aligned} \mathcal{H}(\mathbf{r}, \mathbf{p}) &= \mathbf{p}^2 - \epsilon(\mathbf{r}) = 0 \\ \mathbf{p} &= \nabla S. \end{aligned}$$

According to the theory of characteristics the solution of the eikonal equation is then equivalent to the solution of Hamilton's canonical equations

$$\begin{aligned} \dot{\mathbf{r}} &= \nabla_p \mathcal{H} = \mathbf{p} \\ \dot{\mathbf{p}} &= -\nabla_r \mathcal{H} = \nabla \epsilon, \end{aligned} \quad (2.89)$$

which are (in contrast to (2.88)) a set of ordinary differential equations acting in phase space. We see from Equation (2.89) that rays traveling through homogeneous media ( $\nabla \epsilon = 0$ ) are straight lines. In inhomogeneous media the trajectories have a more complex shape.

The solution of the eikonal equation can be expressed in terms of the solution of Hamilton's canonical equations (2.89) by

$$S(t) - S(0) = \int_0^t d\tau \mathbf{p} \cdot \nabla_p \mathcal{H} = \int_0^t d\tau \mathbf{p}(\tau)^2. \quad (2.90)$$

Hamilton's canonical Equations (2.89) can be interpreted as Newton's equations of motion for a single particle in an effective force field  $U = -\epsilon(\mathbf{r})$ . This analogy shows us why geometrical optics is understood in terms of a particle theory, where the ray acts as the trajectory of a virtual particle. In full analogy to classical mechanics we can eliminate the momentum  $\mathbf{p}$  from Hamilton's canonical equations in order to obtain the ray equation

$$\ddot{\mathbf{r}} = \nabla \epsilon(\mathbf{r}). \quad (2.91)$$

The ray equation is often reformulated using the path element  $ds = \dot{\mathbf{r}} dt$  as a parameter. The resulting equation is

$$\frac{d^2 \mathbf{r}}{ds^2} = \nabla n(\mathbf{r}) \quad (2.92)$$

with the refractive index profile  $n(\mathbf{r})$ .

Equation (2.89) shows via  $\dot{\mathbf{r}} \parallel \nabla S$  that *the rays travel perpendicularly to the phase fronts, i.e., tangentially to the local wave vector*. By scalar multiplication of Equation (2.86) with  $\mathbf{E}_0$  and with the local wave vector ( $\nabla S$ ), we obtain the equations

$$\dot{\mathbf{r}} \cdot \mathbf{E}_0 = \dot{\mathbf{r}} \cdot \mathbf{H}_0 = \mathbf{E}_0 \cdot \mathbf{H}_0 = 0; \quad (2.93)$$

i.e., the electric and magnetic field are perpendicular to each other and to the local wave vector. In this approximation for a general inhomogeneous medium, we encounter again the results we have derived for plane waves in a homogeneous medium. As for plane waves, Poynting's vector is parallel to the wave vector, i.e., *(for isotropic media) the energy of a ray travels along the ray*. It can be shown that Snell's law and Fresnel's formulas can be applied locally.

The higher order terms  $\mathbf{E}_m(\mathbf{r})$  and  $\mathbf{H}_m(\mathbf{r})$  of the Debye expansion take the longitudinal vector component and diffraction effects into account. The corresponding differential equations are known as transport equations.

By considering the ray equation as an Euler-Lagrange equation, we can formulate, in analogy to Hamilton's principle of classical mechanics, a minimum-action principle

$$\delta \left( \int_0^t d\tau \mathbf{p} \cdot \nabla_p \mathcal{H} \right) = 0$$

with the boundary conditions  $\mathbf{r}(0) = \mathbf{r}_{\text{start}}$  and  $\mathbf{r}(t) = \mathbf{r}_{\text{end}}$  for geometrical optics. Using the eikonal Equation (2.88) and Hamilton's canonical Equations (2.89) we obtain Fermat's principle:

$$\delta \left( \int_0^t d\tau \epsilon(\mathbf{r}(\tau)) \right) = \delta \left( \int_0^s d\sigma n(\mathbf{r}(\sigma)) \right) = 0 \quad (2.94)$$

The second form of the variational principle uses again the optical path as a parameter of propagation. Fermat's principle is usually formulated as follows. *The ray will take the minimum time to travel from  $\mathbf{r}_{\text{start}}$  to  $\mathbf{r}_{\text{end}}$ .* Strictly speaking, Fermat's principle requires only that the path integral is stationary, and the interpretation of Fermat's principle given here holds only for nondispersive media. Fermat's principle can replace Maxwell's equations as a foundation of geometrical optics.

### 2.3.3 Caustics

We will now discuss the distribution of rays in space. For this purpose we consider a family of rays  $\mathbf{r} = \mathbf{r}(\mathbf{r}_0(\xi, \eta), \tau)$  starting from a surface  $\mathbf{r}_0(\xi, \eta)$ . The ray coordinates  $(\xi, \eta, \tau)$  represent a geodesic coordinate system in which  $\tau$  describes the evolution

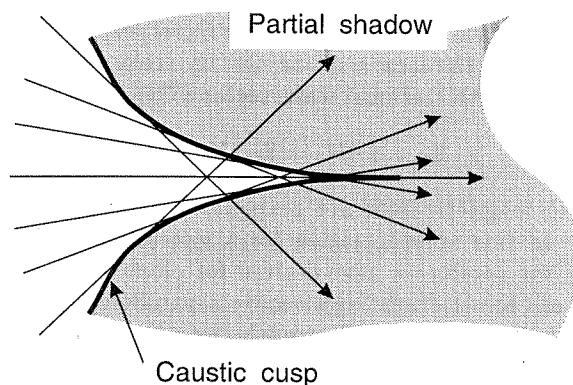


Figure 2.13. Caustic cusp at a focal point.

of the rays and  $\xi$  and  $\eta$  span the transverse direction. The Jacobian for mapping from the ray coordinates  $\xi$ ,  $\eta$  and  $\tau$  to the space coordinates  $x$ ,  $y$  and  $z$  is given by

$$J = \det \left( \frac{\partial(x, y, z)}{\partial(\xi, \eta, \tau)} \right). \quad (2.95)$$

The mapping is singular if the Jacobian vanishes ( $J = 0$ ). A surface for which the Jacobian vanishes is called a *caustic*. It is the envelope of this family of rays. At a caustic geometrical optics predicts an abrupt change of the optical field. This contrasts with the conditions we formulated at the beginning of this section. In consequence, geometrical optics breaks down at caustics. The physical reason for this breakdown is diffraction occurring in a small region around the caustic.

As an example Figure 2.13 shows rays and a *caustic cusp* at a focal point which were produced by a lens with spherical aberrations. More complex-shaped caustics such as *swallowtails*, *hyperbolic* or *elliptic umbilics*, *butterflies* etc. can occur in mirrors and lenses with aberrations as well as in inhomogeneous waveguides. These topologies are, however, rarely observed because of diffraction, which smears the boundary. A detailed discussion of caustics is found in the book by Kravtsov and Orlov [3].

## REFERENCES

- [1] Macleod, H. A., *Thin-Film Optical Filters*, Bristol, Hilger, 1986.
- [2] Thelen, A., *Design of Optical Interference Coatings*, New York, McGraw-Hill, 1989.
- [3] Kravtsov, Y. I. and Orlov, Y. I., *Geometrical Optics of Inhomogeneous Media*, Berlin, Springer, 1990.
- [4] Stratton, J. A., *Electromagnetic Theory*, New York, McGraw-Hill, 1941.
- [5] Born, M., and Wolf, E., *Principles of Optics*, 3rd ed., Oxford, Pergamon Press, 1965.
- [6] Jackson, J. D., *Classical Electrodynamics*, 2nd ed., New York, John Wiley & Sons, 1962.

## Chapter 3

### Waveguide Theory

The fundamental idea behind integrated optics is to handle light by waveguides and not by free-space optics. The first ideas to handle electromagnetic waves at all by waveguides date back to Debye and Hondros (1910). If we examine some classes of integrated optical devices – tapers and geodesic lenses, branches and star couplers, directional couplers, planar spectrographs and phased arrays – we see that there is no exception to this rule. In all these devices the optical field is guided by dielectric waveguide structures, i.e., by properly shaped dielectric profiles. Metal-clad structures represent an alternative way for waveguiding that is popular for microwaves, but are rarely used in integrated optics as the material is too lossy in the optical wavelength region.

The main results of waveguide theory are the number of guided modes for a given dielectric profile and both the effective index and the optical field for each mode. A small number of devices call for the analysis of the radiation field and, in particular, for leaky modes, i.e., for resonances in the radiation field. These results, in turn, form the basis of device modeling, e.g., of the coupled-mode theories describing directional couplers or Bragg gratings.

Waveguide analysis requires the solution of partial differential equations, i.e., a reasonable amount of computer power (with respect to both elapsed memory and time). In consequence, a lot of work in the pioneering days of integrated optics was focussed on approximations such as the effective index method (EIM) or Marcatili's method, which helped to respect the limits of earlier mainframes. With the increasing power of computers, increasingly numerical solutions such as finite element (FE), finite differences (FD) and method of lines (MoL) algorithms are used to analyze waveguide structures. Nevertheless, approximate methods in general and the effective index method in particular represent workhorses for design and modeling of optical waveguide structures even today.

We will start the main thrust of the argumentation within this chapter with the formulation and discussion of the eigenvalue problem for the transverse field components of the electric and magnetic fields. Using the Rayleigh-Schrödinger

perturbation theory, we will in a second step derive the simplified equations which hold for weakly guiding waveguides. By applying the perturbation theory once more, we will obtain the effective index method (EIM) and Marcatili's method. The rest of the chapter comprises some analytical results and a brief discussion of the geometrical optical approach to the waveguide problem.

### 3.1 GENERAL REMARKS

A waveguide structure is a cylindrical dielectric profile with an arbitrary cross section. A typical waveguide structure consists of one or more waveguides (see Figure 3.1) which are associated with the maxima of the dielectric profile. In addition, typical waveguide structures for integrated optics will be non-metallic ( $\text{Re}(\epsilon) > 1$ ), since all known metals are too lossy in the optical wavelength region. We introduce one longitudinal coordinate  $z$  and two transverse coordinates  $\mathbf{r}_t$  to describe the waveguide structures. Obviously, the dielectric profile describing a waveguide structure

$$\epsilon = \epsilon(\mathbf{r}_t)$$

depends only on the two transverse coordinates.

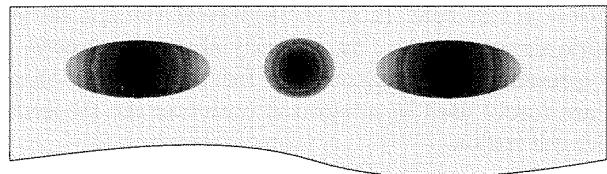


Figure 3.1. Cross section of a waveguide structure.

#### 3.1.1 Phenomena of Waveguiding

An eigenmode  $m$  of a waveguide structure is a propagating or evanescent wave which maintains its transverse shape during propagation. Thus, a forward-propagating eigenmode is given by<sup>1</sup>

$$\begin{aligned}\mathbf{E}^{(m)}(\mathbf{r}_t, z) &= \mathbf{e}^{(m)}(\mathbf{r}_t) e^{iq_m z} \\ \mathbf{H}^{(m)}(\mathbf{r}_t, z) &= \mathbf{h}^{(m)}(\mathbf{r}_t) e^{iq_m z}.\end{aligned}\quad (3.1)$$

1. To avoid clumsy notation, we shall restrict our further discussion to forward-propagating waves. Section 3.2.3 shows the relations between forward and reverse waves.

Three different parameters are commonly used to describe the propagation characteristics of an eigenmode  $m$ . One is the propagation constant  $q_m$  used previously, the second one is the effective refractive index

$$n_m = \frac{q_m}{k_0}, \quad (3.2)$$

the third one is the effective dielectric constant

$$\epsilon_m = n_m^2 \quad (3.3)$$

which will turn out to be the true eigenvalue that we are seeking.

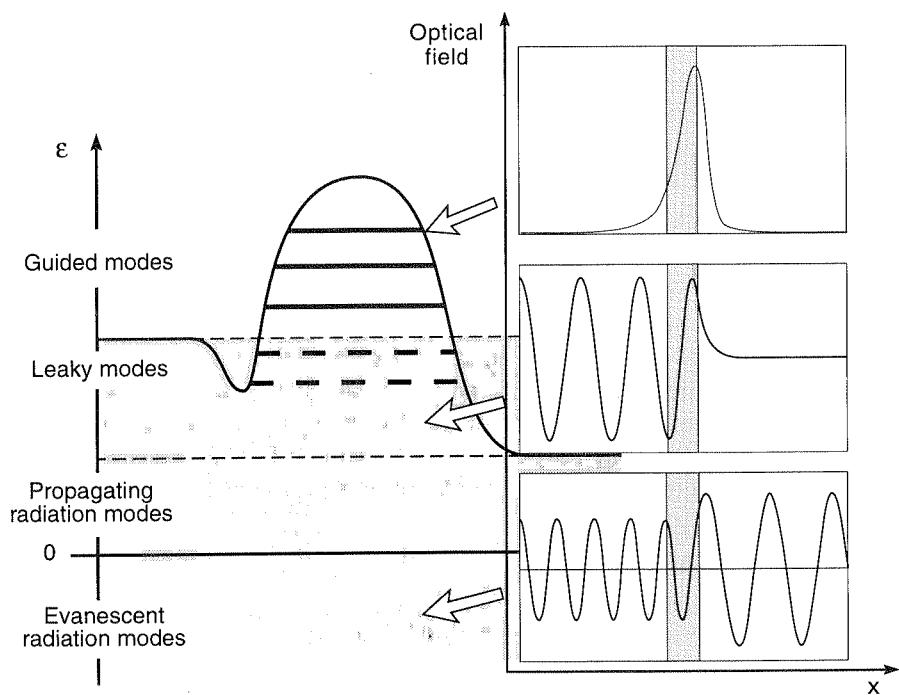


Figure 3.2. Various types of eigenmode supported by a waveguide.

Before we go into a detailed examination of the eigenvalue problems we will give a brief overview of the phenomena which are typical for passive optical waveguides ( $\text{Im}(\epsilon(\mathbf{r}_t)) = 0$ ). For that purpose, Figure 3.2 shows the dielectric profile of a hypothetical waveguide and various types of eigenmode supported by the waveguide structure.

- A waveguide does not support any eigenmodes whose eigenvalue is greater than the maximum of the dielectric function  $\epsilon_m < \max(\epsilon(r_t))$ .
- The guided waves form the discrete part of the spectrum. In the transverse plane the optical power associated with them vanishes at infinity, and their eigenvalues are in the range  $\epsilon_{\text{MAX}} > \epsilon_m \geq \max(\epsilon(r_b))$  where  $\max(\epsilon(r_b))$  is the maximum of the dielectric function along the “boundary”  $r_b$  located at infinity. The cutoff of the eigenmode  $m$  is defined by

$$\epsilon_m = \max(\epsilon(r_b)). \quad (3.4)$$

We should point out here that waveguide structures exist which do not support any guided modes.

- The continuous part of the spectrum is formed by the radiation modes whose eigenvalues are restricted to the range  $\epsilon_m < \max(\epsilon(r_b))$ . Radiation modes show oscillatory behavior at least along a fraction of the boundary. According to their eigenvalues they are classified into *propagating* ( $\epsilon \geq 0$ ) and *evanescent radiation modes* ( $\epsilon < 0$ ).
- Guided and radiation modes form a complete set of functions*; i.e., any input field of the waveguide can be represented by these modes. The eigenmodes of a waveguide with a lossless dielectric profile ( $\text{Im}(\epsilon(r_t)) = 0$ ) are real and the eigenmodes are orthogonal functions.
- Leaky waves are not eigenmodes of the waveguides but resonances in its continuous spectrum. In physical terms, they are guided modes below cutoff. However, the optical fields associated with them diverge at least at a part of the “boundary”; i.e., we should never try to compute the optical power carried by these waves. The resonances are called *leaky modes* since they cannot keep the optical power within the core of the waveguide.

### 3.1.2 Typical Waveguides in Integrated Optics

Figure 3.3 shows a collection of waveguides used in integrated optics today. The waveguide is defined lithographically in all examples, usually by contact or projection lithography, sometimes by direct electron beam writing.

The fabrication of the first four types of waveguide – rib, buried-rib, strip and strip loaded waveguides – involves usually an etching process such as wet chemical etching, reactive ion etching (RIE) or reactive ion beam etching (RIBE). The choice of process and the required aspect ratio (etching depth/structure size) determines the shape of the waveguide, its roughness and, importantly, the fabrication tolerances, especially the waveguide width. For the III-V semiconductor material systems – mainly InGaAsP/InP and GaAlAs/GaAs – the layers shown for these waveguides are grown by epitaxy, usually by metal organic vapor phase epitaxy

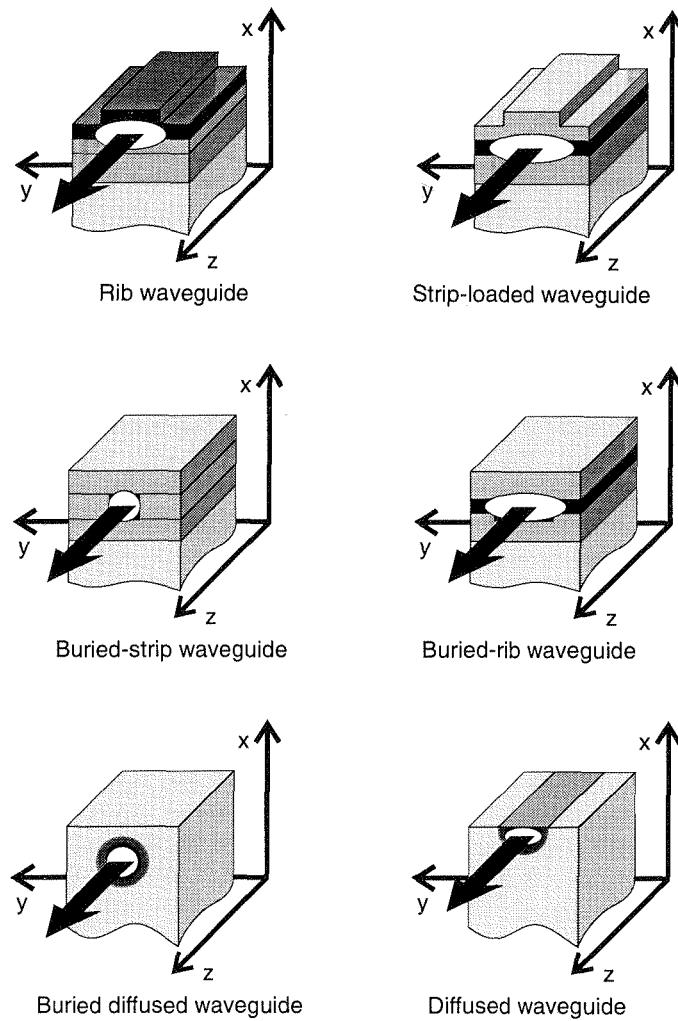


Figure 3.3. Some types integrated optical waveguides.

(MOVPE) or metal organic molecular beam epitaxy (MOMBE) or sometimes still by liquid phase epitaxy (LPE). For the silica-on-silicon material systems ( $\text{SiO}_2/\text{Si}$ ), in contrast, the layers are fabricated by deposition – various types of chemical vapor deposition (CVD) or flame hydrolysis – or by thermal oxidation. The choice of the fabrication process defines the ranges of available layer thicknesses and material compositions (refractive indices). In addition, it determines the roughness of the interfaces and the fabrication tolerances of the material composition and the layer thickness. Diffused and buried-diffused waveguides, the last two types of waveguide,

can be fabricated by ion exchange in glass and by diffusion processes in LiNbO<sub>3</sub>. The fabrication of a buried-diffused waveguide requires a second diffusion process such as a field-assisted ion exchange. Obviously, the whole shape of the dielectric profile and its fabrication tolerances depend on the diffusion process.

The typical size of a single-mode waveguide, i.e., a waveguide which supports only one guided mode per polarisation, is another relevant criterion of classification. It determines the ease of fabrication and fiber-chip coupling and the size of integrated optical circuits. We will show in the following (see Section 3.10) that the cutoff thickness  $d_c$  of the single-layer slab is given by

$$d_c = \frac{\lambda}{2\sqrt{\epsilon_L - \epsilon_B}} \quad (3.5)$$

where  $\epsilon_B$  is the background dielectric constant and  $\epsilon_L$  that of the waveguide layer. We can use this simple formula to estimate the smallest dimension of a waveguide (usually its thickness). To avoid extremely elliptical near fields, the transverse dimensions (usually the waveguide width) are not more than twice as large. Depending on the underlying material system and global layout considerations, the quantity  $\sqrt{\epsilon_L - \epsilon_B}$  is in the range of

$$\sqrt{\epsilon_L - \epsilon_B} = 0.1 \dots 1.$$

The lower bound  $\sqrt{\epsilon_L - \epsilon_B} \approx 0.1$  corresponds to fiber-matched waveguides, i.e., to waveguides whose spot size coincides with that of a single-mode fiber. Diffused waveguides and most waveguides realized in SiO<sub>2</sub>/Si are typical members of this class. The upper bound  $\sqrt{\epsilon_L - \epsilon_B} > 0.7$  indicates strongly guiding waveguides which allow for bends with extremely small radii of curvature and are suited to small chips. Most III-V semiconductor waveguides are representatives of this second class. The shape of a rib, buried-rib or strip-loaded waveguide is chosen to increase the widths of single-mode waveguides and to enhance the ease of fabrication. Some material systems such as SiON/Si and polymers allow for the adjustment of  $\sqrt{\epsilon_L - \epsilon_B}$  within a wide range.

Since different devices call for different waveguide properties it is difficult to define a standard waveguide even for one material system. For the III-V material systems in contrast, we observe an increasing number of integrated circuits which are based on more than one type of waveguide.

### 3.1.3 Notations

To avoid misinterpretations of the equations within this and the following chapters we will add a brief comment on the notation.

For general cylindrical coordinate systems, the differential operators can be written as

$$\nabla\phi = \nabla_t\phi + \frac{\partial\phi}{\partial z}\mathbf{i}_z \quad (3.6)$$

$$\nabla^2\phi = \nabla_t^2\phi + \frac{\partial^2\phi}{\partial z^2} \quad (3.7)$$

$$\nabla \cdot \mathbf{F} = \nabla_t \cdot \mathbf{F}_t + \frac{\partial F_z}{\partial z} \quad (3.8)$$

$$\nabla \times \mathbf{F} = \nabla_t \times \mathbf{F}_t - \mathbf{i}_z \times \left( \nabla_t F_z - \frac{\partial \mathbf{F}_t}{\partial z} \right) \quad (3.9)$$

$$\Delta\mathbf{F} = \Delta_t\mathbf{F}_t + \frac{\partial^2\mathbf{F}}{\partial z^2} + \nabla_t^2 F_z \mathbf{i}_z, \quad (3.10)$$

i.e., they can be kept coordinate-free with respect to the two transverse dimensions. To ensure that the vectorial and scalar Laplacians cannot be mixed up, we will always write  $\nabla^2$  and  $\nabla_t^2$  for the scalar and  $\Delta$  and  $\Delta_t$  for the vectorial Laplacians. A general representation of the differential operators and the elaborations of the most important cylindrical coordinate systems – Cartesian, circular and elliptical coordinates – are given in the mathematical appendix (Appendix D).

For formulating the eigenvalue and beam propagation problems, we will make extensive use of the quantum-mechanical notation (and terminology) as it is more compact and makes the mathematical background more transparent. Furthermore, it allows for the elaboration of fundamental techniques such as perturbation and coupled mode theory irrespective of the underlying operators.

In the following, we will always denote a linear operator by a calligraphic letter (e.g.,  $\mathcal{H}$ ,  $\mathcal{U}$ ), the expression  $|\phi\rangle$  stands for a Dirac vector which is related to the optical field via the projection  $\phi(\mathbf{r}) = \langle \mathbf{r} | \phi \rangle$ . The eigenmode belonging to an eigenvalue  $\epsilon_m$  is denoted by  $|m\rangle$ . The eigenmodes of a degenerated eigenvalue  $m$  are written as a formal vector  $|\mathbf{m}\rangle$  in the eigenspace. The scalar product of two optical fields  $\psi$  and  $\phi$ , usually written as an overlap integral, becomes simply  $\langle \psi | \phi \rangle$ . For more information, the mathematical appendix contains a collection of theorems which apply to the theory of Hilbert spaces and distributions.

## 3.2 FORMULATIONS OF THE EIGENVALUE PROBLEM

In this and the next section we will formulate three different sets of equations describing the eigenvalue problem of an arbitrary waveguide. In all three cases, a reduced eigenvalue problem for only two vector components will turn out to be equivalent to the original problem (comprising six vector components). In addition to the obvious benefits for a numerical treatment, these formulations have the advantage that neither magnetic monopoles  $\nabla \cdot \mathbf{H} \neq 0$  nor nonphysical electric fields  $\nabla \cdot \epsilon \mathbf{E} \neq 0$  can occur. In contrast, it should be noted that the full eigenvalue problems of the three-component vectorial Helmholtz equations, do allow for such solutions.

### 3.2.1 The Vector **E**-Field and **H**-Field Eigenvalue Problems

Using Equation (3.1) and the two vector Helmholtz equations<sup>2</sup>

$$\begin{aligned}\Delta \mathbf{E} + k_0^2 \epsilon \mathbf{E} &= -\nabla (\nabla(\ln \epsilon) \cdot \mathbf{E}) \\ \Delta \mathbf{H} + k_0^2 \epsilon \mathbf{H} &= -\nabla(\ln \epsilon) \times \nabla \times \mathbf{H}\end{aligned}$$

we obtain two eigenvalue problems

$$\frac{1}{k_0^2} [\Delta_t + k_0^2 \epsilon + \nabla_t \nabla_t (\ln \epsilon) \cdot] \mathbf{e}_t^{(m)} = \epsilon_m \mathbf{e}_t^{(m)} \quad (3.11)$$

$$\frac{1}{k_0^2} [\Delta_t + k_0^2 \epsilon + \nabla_t (\ln \epsilon) \times \nabla_t \times] \mathbf{h}_t^{(m)} = \epsilon_m \mathbf{h}_t^{(m)} \quad (3.12)$$

for the transverse field components of the electric and magnetic fields. Equation (3.11) is often called the *vector E-field formulation* of the eigenvalue problem, Equation (3.12) is its *vector H-field formulation*. Since the transverse field components do not couple to the longitudinals, we suppress the equations for the longitudinal field components here.

We will next show that all solutions of Equations (3.11) and (3.12) are eigenmodes of the full Maxwell equations, i.e., that they satisfy Equations (2.8) to (2.11). Using the two divergence theorems  $\nabla \cdot \epsilon \mathbf{E} = 0$  and  $\nabla \cdot \mathbf{H} = 0$ , we can conclude that if the two transverse vector components of either the electric or the magnetic field ( $\mathbf{e}_t^{(m)}$  or  $\mathbf{h}_t^{(m)}$ ) already satisfy Equation (3.1), then the corresponding longitudinal field components  $e_z^{(m)}$  and  $h_z^{(m)}$  will also do so. The longitudinal components can then be written as

$$\begin{aligned}-iq_m e_z^{(m)} &= \nabla_t \cdot \mathbf{e}_t^{(m)} + \nabla_t (\ln \epsilon) \cdot \mathbf{e}_t^{(m)} \\ -iq_m h_z^{(m)} &= \nabla_t \cdot \mathbf{h}_t^{(m)}.\end{aligned} \quad (3.13)$$

If, however, all components of the electric field satisfy Equation (3.1), we see by using Maxwell's equation  $\nabla \times \mathbf{E} = ik_0 \mathbf{H}$  that the magnetic field components will then also satisfy this basic assumption. If all magnetic field components satisfy Equation (3.1), then it also holds for the electric field because of  $\nabla \times \mathbf{H} = -ik_0 \epsilon \mathbf{E}$ . By elaborating these equations, we obtain after some algebra the following relations between the transverse vector components of the electric and magnetic field:

$$\begin{aligned}k_0^2 n_m \mathbf{e}_t^{(m)} &= \mathbf{i}_z \times \left[ \nabla_t \times \frac{1}{\epsilon} \nabla_t \times \mathbf{h}_t^{(m)} - k_0^2 \mathbf{h}_t^{(m)} \right] \\ -k_0^2 n_m \mathbf{h}_t^{(m)} &= \mathbf{i}_z \times \left[ \nabla_t \times \nabla_t \times \mathbf{e}_t^{(m)} - k_0^2 \epsilon \mathbf{e}_t^{(m)} \right]\end{aligned} \quad (3.14)$$

From a purely theoretical point of view, both vector **E**-field and **H**-field formulations are completely equivalent. However, the vector **H**-field formulation is

preferred for the calculation of numerical results, since both transverse magnetic field components are continuous at a dielectric interface. For the transverse **E**-field, in contrast, only the field component parallel to the dielectric interface is continuous.<sup>3</sup>

At the end of this section, we should note that both the vector **E**-field (3.11) and **H**-field formulations (3.12) represent standard eigenvalue problems

$$\mathcal{H} |m\rangle = \epsilon_m |m\rangle \quad (3.15)$$

for the effective dielectric constant  $\epsilon_m$  and not for the effective refractive index  $n_m$ . We will discuss the spectrum of the Hamiltonians belonging to the vector **E**-field and **H**-field formulations in Section 3.2.4 in more detail.

### 3.2.2 The $E_z/H_z$ Eigenvalue Problem

There is an alternative formulation of the eigenvalue problem in terms of the two longitudinal field components  $E_z$  and  $H_z$ . To derive the corresponding equations, we start from Maxwell's equations  $ik_0 \epsilon \mathbf{E} = -\nabla \times \mathbf{H}$  and  $ik_0 \mathbf{H} = \nabla \times \mathbf{E}$  and arrive in a first step at

$$\begin{aligned}-ik_0 \epsilon \mathbf{e}_t^{(m)} &= \mathbf{i}_z \times (ik_0 n_m \mathbf{h}_t^{(m)} - \nabla_t h_z^{(m)}) \\ ik_0 \mathbf{h}_t^{(m)} &= \mathbf{i}_z \times (ik_0 n_m \mathbf{e}_t^{(m)} - \nabla_t e_z^{(m)}).\end{aligned} \quad (3.16)$$

By eliminating the transverse field components on the right-hand side of this equation we can express them in terms of the longitudinals  $e_z^{(m)}$  and  $h_z^{(m)}$ :

$$\begin{aligned}k_0 \mathbf{e}_t^{(m)} &= \frac{i}{\epsilon - n_m^2} (n_m \nabla_t e_z^{(m)} - \mathbf{i}_z \times \nabla_t h_z^{(m)}) \\ k_0 \mathbf{h}_t^{(m)} &= \frac{i}{\epsilon - n_m^2} (n_m \nabla_t h_z^{(m)} + \epsilon \mathbf{i}_z \times \nabla_t e_z^{(m)})\end{aligned} \quad (3.17)$$

By inserting these expressions in turn into the  $z$ -components of the eigenvalue Equations (3.11) and (3.12)

$$[\Delta_t + k_0^2 (\epsilon - n_m^2)] e_z^{(m)} = -iq_m \nabla_t (\ln \epsilon) \cdot \mathbf{e}_t^{(m)} \quad (3.18)$$

$$[\Delta_t + k_0^2 (\epsilon - n_m^2)] h_z^{(m)} = \nabla_t (\ln \epsilon) \cdot (\nabla_t h_z^{(m)} - ik_0 n_m \mathbf{h}_t^{(m)}), \quad (3.19)$$

we obtain the desired equations for the longitudinal field components

$$\begin{aligned}[\Delta_t + k_0^2 (\epsilon - n_m^2) - n_m^2 C_S^{(m)}] e_z^{(m)} &= C_C^{(m)} h_z^{(m)} \\ [\Delta_t + k_0^2 (\epsilon - n_m^2) - \epsilon C_S^{(m)}] h_z^{(m)} &= -\epsilon C_C^{(m)} e_z^{(m)}\end{aligned} \quad (3.20)$$

3. The boundary conditions of electric and magnetic fields at a dielectric interface are examined in Section 2.2.

2. See Equations (2.23) and (2.24) in Section 2.1

where both self- and cross-coupling of the vector components

$$\begin{aligned}\mathcal{C}_S \phi &= \frac{1}{\epsilon - n_m^2} \nabla_t (\ln \epsilon) \cdot \nabla_t \phi \\ \mathcal{C}_C \phi &= \frac{n_m}{\epsilon - n_m^2} [\nabla_t (\ln \epsilon) \times \nabla_t \phi]_z\end{aligned}$$

is caused by the transverse gradient of the dielectric profile. The transverse field components can be easily calculated using Equations (3.17). We see that the  $E_z/H_z$  formulation leads to an implicit eigenvalue problem for the effective refractive index  $n_m$ . Obviously, for general dielectric profiles the  $E_z/H_z$  formulation of the eigenvalue problem is inferior to the vector  $\mathbf{E}$ -field and  $\mathbf{H}$ -field formulations presented previously.

For step-index waveguide structures, however, where neither self- nor cross-coupling occurs ( $\mathcal{C}_S = \mathcal{C}_C = \mathcal{O}$ ), we obtain two independent eigenvalue problems

$$\frac{1}{k_0^2} (\Delta_t + k_0^2 \epsilon) \phi^{(m)} = \epsilon_m \phi^{(m)} \quad (3.21)$$

for the longitudinal field components ( $\phi^{(m)} = e_z^{(m)}$  and  $\phi^{(m)} = h_z^{(m)}$ ). We should note here that both  $e_z^{(m)}$  and  $h_z^{(m)}$  are continuous at the dielectric interface. For hybrid modes ( $e_z^{(m)} \neq 0$  and  $h_z^{(m)} \neq 0$ ), both field components are described by solutions of the same eigenvalue problem. For transverse modes, one of the longitudinal field components vanishes ( $e_z^{(m)} = 0$  or  $h_z^{(m)} = 0$ ).

### 3.2.3 Transverse Eigenmodes

*Every waveguide supports transverse eigenmodes.* For a discussion of this theorem, we introduce geodesic coordinates, i.e., a right-handed orthogonal cylindrical coordinate system  $(u_\perp, u_\parallel, z)$  which is chosen such that the dielectric profile

$$\epsilon = \epsilon(u_\perp)$$

depends only on the normal coordinate  $u_\perp$ . Figure 3.4 shows as an illustration the transverse coordinates  $u_\perp$  and  $u_\parallel$  for an elliptic waveguide.

A transverse electric mode (TE mode) is characterized by the condition  $e_z = 0$ . Using the  $z$ -component of the vector  $\mathbf{E}$ -field eigenvalue Equation (3.18)

$$(\Delta_t + k_0^2 \epsilon - q_m^2) e_z^{(m)} = -iq_m \nabla_\perp (\ln \epsilon) e_\perp^{(m)}$$

we see that the normal field component  $e_\perp = 0$  must also vanish for a transverse electric mode. Thus, the TE modes are solutions of the scalar eigenvalue problem

$$\frac{1}{k_0^2} (\Delta_t + k_0^2 \epsilon) e_\parallel^{(m)} = \epsilon_m e_\parallel^{(m)} \quad (3.22)$$

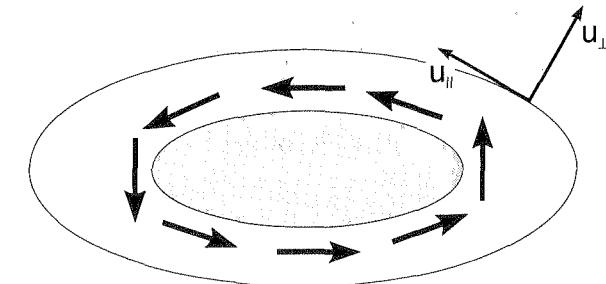


Figure 3.4. Waveguide structure and electric/magnetic field for a TE/TM mode.

for the parallel field component  $e_\parallel^{(m)}$ . Using Maxwell's equation  $ik_0 \mathbf{H} = \nabla \times \mathbf{E}$ , we can easily derive the following expressions for the magnetic field components

$$\begin{aligned}h_\perp^{(m)} &= -\sqrt{\epsilon_m} e_\parallel^{(m)} \\ h_\parallel^{(m)} &= 0 \\ h_z^{(m)} &= -\frac{i}{k_0} \nabla_\perp e_\parallel^{(m)}.\end{aligned} \quad (3.23)$$

Thus, TE modes have only three non-vanishing field components  $e_\parallel$ ,  $h_\perp$  and  $h_z$ .

A transverse magnetic mode (TM mode) is characterized by the condition  $h_z = 0$ . Using the  $z$ -component of the vector  $\mathbf{H}$ -field eigenvalue problem (3.19)

$$(\Delta_t - \nabla_\perp (\ln \epsilon) \nabla_\perp + k_0^2 \epsilon - q_m^2) h_z^{(m)} = -iq_m \nabla_\perp (\ln \epsilon) h_\perp^{(m)}$$

we can again conclude that the normal field component  $h_\perp = 0$  must vanish for TM modes. Analog to the TE case, we obtain the following scalar eigenvalue problem:

$$\frac{1}{k_0^2} (\Delta_t - \nabla_\perp (\ln \epsilon) \nabla_\perp + k_0^2 \epsilon) h_\parallel^{(m)} = \epsilon_m h_\parallel^{(m)} \quad (3.24)$$

for the parallel field component of the magnetic field. Using Maxwell's equation  $-ik_0 \epsilon \mathbf{E} = \nabla \times \mathbf{H}$  we easily derive the field components:

$$\begin{aligned}e_\perp^{(m)} &= \frac{\sqrt{\epsilon_m}}{\epsilon} h_\parallel^{(m)} \\ e_\parallel^{(m)} &= 0 \\ e_z^{(m)} &= \frac{i}{k_0 \epsilon} \nabla_\perp h_\parallel^{(m)}.\end{aligned} \quad (3.25)$$

Obviously, the TM mode also has only three non-vanishing field components.

For weakly guiding waveguides, the diameter of the eigenmodes will be large in comparison to the wavelength of the eigenmode inside the material. Thus, the optical field does not vary significantly over one wavelength, i.e.,

$$|\nabla_t \phi| \ll |q_m \phi|.$$

In that case, the longitudinal field components ( $e_z^{(m)}$  and  $h_z^{(m)}$ ) will be significantly smaller than the transverse ones.

We should note that *the fundamental mode of a waveguide structure is usually a hybrid (vectorial) mode.*

### 3.2.4 Orthogonality of the Eigenmodes

*Any lossless waveguide structure is described by a Hermitian Hamiltonian. Its effective dielectric constants are real, the eigenmodes form an orthogonal set of functions. For lossy waveguide structures, the eigenmodes belonging to the Hamiltonian and its adjoint counterpart form an orthogonal set of functions.*

To prove this theorem, we introduce Maxwell's equations for the complex conjugated dielectric profile  $\epsilon^*$ . The electric and magnetic fields which are solutions of these equations are indicated by a bar on top of the corresponding symbol ( $\bar{\mathbf{E}}$  and  $\bar{\mathbf{H}}$ ). Let us now consider the generalized Poynting vector

$$\tilde{\mathbf{S}}_{lm} = \mathbf{E}^{(l)} \times \bar{\mathbf{H}}^{(m)*} + \bar{\mathbf{E}}^{(m)*} \times \mathbf{H}^{(l)} \quad (3.26)$$

whose divergence

$$\nabla \cdot \tilde{\mathbf{S}}_{lm} = -ik_0(\epsilon - \epsilon^*)\mathbf{E}^{(l)} \cdot \bar{\mathbf{E}}^{(m)*} \quad (3.27)$$

vanishes if the waveguide structure is lossless  $\epsilon = \epsilon^*$ . In that case the barred and unbarred fields coincide. For our further discussion, we will assume lossless waveguides and suppress the bars in Equation (3.26). By integrating the above equation over the cross section  $A_\infty$  of the waveguide structure we obtain

$$\int_{A_\infty} da \nabla \cdot \tilde{\mathbf{S}}_{lm} = \int_{A_\infty} da \mathbf{i}_z \cdot \frac{\partial \tilde{\mathbf{S}}_{lm}}{\partial z} + \int_{A_\infty} da \nabla_t \cdot \tilde{\mathbf{S}}_{lm} = 0. \quad (3.28)$$

Using the two-dimensional form of the divergence theorem we can express the second term on the right-hand side as a line integral:

$$\int_{A_\infty} da \nabla_t \cdot \tilde{\mathbf{S}}_{lm} = \oint_{\partial A_\infty} dl \cdot \tilde{\mathbf{S}}_{lm}. \quad (3.29)$$

For guided modes vanishing at infinity, this line integral vanishes. We can apply the same argument to radiation modes if we regard them as guided modes of a metal-clad waveguide whose cross-section becomes infinite. As a first result, we

can conclude that a generalized Poynting vector for a lossless waveguide structure satisfies the equation

$$\int_{A_\infty} da \mathbf{i}_z \cdot \frac{\partial \mathbf{S}_{lm}}{\partial z} = 0. \quad (3.30)$$

It should be noted here that only the transverse field components  $\mathbf{e}_t$  and  $\mathbf{h}_t$  contribute to Equation (3.30). Let us now consider two different eigenvalues  $\epsilon_l \neq \epsilon_m$  corresponding to either the vector  $\mathbf{E}$ -field or  $\mathbf{H}$ -field formulation of the eigenvalue problem, i.e., Equations (3.11) or (3.12). Each of these eigenvalues leads to two possible propagation constants  $\pm q_m$ , which correspond to eigenmodes propagating forward and backward. Using Equation (3.14), we can easily derive transformation rules for the transverse components of amplitudes

$$\begin{cases} \mathbf{e}_t \\ \mathbf{h}_t \end{cases} \xrightarrow{q \rightarrow -q} \begin{cases} \mathbf{e}_t \\ -\mathbf{h}_t \end{cases}$$

under the transformation  $q_m \rightarrow -q_m$ . By inserting these results into Equation (3.30) for all possible modes for the eigenvalues  $\epsilon_l$  and  $\epsilon_m$ , we arrive at two essentially different equations

$$\int_{A_\infty} da \mathbf{i}_z \cdot (\mathbf{h}_t^{(m)*} \times \mathbf{e}_t^{(l)} \pm \mathbf{h}_t^{(l)} \times \mathbf{e}_t^{(m)*}) = 0.$$

If we simply add these two equations we obtain the orthogonality relations

$$\int_{A_\infty} da \mathbf{i}_z \cdot (\mathbf{h}_t^{(m)*} \times \mathbf{e}_t^{(l)}) = 0. \quad (3.31)$$

In a third step we will use the orthogonality relations to define scalar products for the vector  $\mathbf{E}$ -field and  $\mathbf{H}$ -field formulations of the eigenvalue problem. We start with the vector  $\mathbf{E}$ -field formulation and rewrite the orthogonality relations using Equation (3.14), i.e.,

$$\begin{aligned} \int_{A_\infty} da \mathbf{i}_z \cdot (\mathbf{h}_t^{(m)*} \times \mathbf{e}_t^{(l)}) &= \int_{A_\infty} da \mathbf{e}_t^{(l)} \cdot (\mathbf{i}_z \times \mathbf{h}_t^{(m)*}) \\ &= \frac{1}{n_m^*} \int_{A_\infty} da (\mathcal{S}_E \mathbf{e}_t^{(m)*}) \cdot \mathbf{e}_t^{(l)} \end{aligned}$$

in which the linear operator  $\mathcal{S}_E$  is given by

$$\mathcal{S}_E \mathbf{e}_t = \frac{1}{k_0^2} \nabla_t \times \nabla_t \times \mathbf{e}_t - \epsilon \mathbf{e}_t. \quad (3.32)$$

Using the vectorial form of Green's theorem in two dimensions

$$\begin{aligned} & \int_{A_\infty} da [\mathbf{F}_t \cdot (\nabla_t \times \phi \nabla_t \times \mathbf{G}_t) - \mathbf{G}_t \cdot (\nabla_t \times \phi \nabla_t \times \mathbf{F}_t)] \\ &= \int_{\partial A_\infty} d\mathbf{l} (\mathbf{G}_t \times \phi \nabla_t \times \mathbf{F}_t - \mathbf{F}_t \times \phi \nabla_t \times \mathbf{G}_t) \end{aligned}$$

and keeping in mind that the electric field vanishes at infinity<sup>4</sup> we see that for lossless media the linear operator  $\mathcal{S}_E$  is Hermitian with respect to a scalar product of the type  $\int da \mathbf{e}_{tb}^* \cdot \mathbf{e}_{tc}$ . Thus, the form

$$\langle b | c \rangle = \int_{A_\infty} da (\mathcal{S}_E \mathbf{e}_{tb}^*) \cdot \mathbf{e}_{tc} \quad (3.33)$$

itself is Hermitian ( $\langle b | c \rangle = \langle b | c \rangle^*$ ). It can then serve as a scalar product, since the orthogonality relations also hold for this form. By carrying out the whole procedure for the vector  $\mathbf{H}$ -field we obtain

$$\langle b | c \rangle = \int_{A_\infty} da (\mathcal{S}_H \mathbf{h}_{tb}^*) \cdot \mathbf{h}_{tc} \quad (3.34)$$

with the Hermitian operator

$$\mathcal{S}_H \mathbf{h}_t = \nabla_t \times \frac{1}{k_0^2 \epsilon} (\nabla_t \times \mathbf{h}_t) - \mathbf{h}_t. \quad (3.35)$$

In the last step of this discussion we will show that a Hamiltonian describing lossless waveguide structures is Hermitian with respect to the corresponding scalar products (3.33) and (3.34). For this purpose, we consider the expression

$$\langle \mathcal{H}m | n \rangle - \langle m | \mathcal{H}n \rangle = (\epsilon_m^* - \epsilon_n) \langle m | n \rangle.$$

Due to the orthogonality of eigenmodes (3.31), the right-hand side of this expression vanishes if  $|m\rangle$  and  $|n\rangle$  are two eigenmodes belonging to two different eigenvalues. Thus, we can conclude that

$$\langle \mathcal{H}m | n \rangle = \langle m | \mathcal{H}n \rangle,$$

i.e., that the Hamiltonian is Hermitian for lossless waveguide structures. Using the same expression for  $|m\rangle = |n\rangle$ , we can further conclude that  $\epsilon_m^* = \epsilon_m$ , i.e., that the effective dielectric constants are real quantities. Nevertheless, even for lossless waveguide structures the propagation constants of eigenmodes become purely imaginary if the corresponding effective dielectric constants are negative. Such eigenmodes are called evanescent modes.

4. For radiation modes, see the preceding argument.

For lossy waveguide structures, the bra-vector  $\langle m |$  denotes an eigenmode of the adjoint eigenvalue problem

$$\mathcal{H}^\dagger |m\rangle = \bar{\epsilon}_m |m\rangle.$$

Let us now consider the expression

$$\langle \mathcal{H}^\dagger m | n \rangle - \langle m | \mathcal{H}n \rangle = 0 = (\bar{\epsilon}_m^* - \epsilon_n) \langle m | n \rangle.$$

Obviously, the eigenmodes belonging to different eigenvalues ( $\bar{\epsilon}_m^* - \epsilon_n$ ) are orthogonal, i.e.,  $\langle m | n \rangle = 0$  for  $m \neq n$ . The eigenvalues of the Hamiltonian and its adjoint counterpart

$$\bar{\epsilon}_m^* = \epsilon_m$$

are conjugated complex to each other.

### 3.3 POWER FLUX IN WAVEGUIDES

For an eigenmode  $m$  with an effective propagation constant  $q_m = \text{Re}(q_m) + i\text{Im}(q_m)$  the projection of Poynting's vector onto the waveguide axis is given by

$$S_m(\mathbf{r}_t, z) = \frac{c}{8\pi} \mathbf{e}_t^{(m)} \times \mathbf{h}_t^{(m)*} \cdot \mathbf{i}_z e^{-\text{Im}(q_m)z}. \quad (3.36)$$

Its real part  $\text{Re}(S_m)$  represents the optical power carried along the waveguide. The evolution of power flux is governed only by the imaginary part of the effective propagation constant. Obviously, we get

$$\text{Im}(q_m) \begin{cases} > 0 & \text{for loss} \\ = 0 & \text{for power conservation} \\ < 0 & \text{for gain.} \end{cases} \quad (3.37)$$

Even in the cases of gain and loss, however, the transverse shape of the optical field is conserved during propagation.

We should note here that the total optical power carried along the waveguide is related to the scalar product

$$\int_{A_\infty} da \text{Re}(S_m) \propto \frac{1}{q_m} \text{Re}(\langle m | m \rangle)$$

of the eigenmode with itself (see Equations (3.33) and (3.34)).

The confinement factor of an eigenmode  $m$ ,

$$\eta_m(A) = \frac{\int_A da \text{Re}(S_m)}{\int_{A_\infty} da \text{Re}(S_m)}, \quad (3.38)$$

relates the optical power flux concentrated in the area  $A$  of the transverse plane to the total optical power carried by the eigenmode  $m$ . The confinement factor is a very intuitive quantity which is often used to estimate the modal behavior or to observe the power flow inside a waveguide structure. Thus the formula

$$\text{Im}(n_m) \approx \int_A da \eta_m^0(\mathbf{r}_t) \text{Im}(n(\mathbf{r}_t))$$

allows us to estimate the gain and loss of guided modes for a weakly active waveguide. This result can easily be derived for nearly lossless dielectric profiles in the framework of the first-order Rayleigh-Schrödinger perturbation theory (see Section 3.5).

### 3.4 DISPERSION

Three different effects contribute to the dispersion of optical waveguides: material dispersion, waveguide dispersion and – if more than one guided mode is excited – intermodal dispersion. *Material dispersion* is due to the dispersion of the media which form the waveguide. *Waveguide dispersion*, in contrast, is due to the change of propagation characteristics with the wavelength. It also occurs for waveguides formed by competely non-dispersive media.

The optical field of the eigenmode  $m$  varies as  $\exp(i[k_0 n_m z - \omega t])$ . Thus, the phase fronts propagate with a phase velocity

$$c_m = \frac{c}{n_m} \quad (3.39)$$

along the waveguide.

The group velocity of a pulse is defined as

$$v_m = \frac{\partial \omega}{\partial (k_0 n_m)} = c_m \left( 1 - \frac{\lambda}{n_m} \frac{\partial n_m}{\partial \lambda} \right)^{-1}. \quad (3.40)$$

We see that dispersion ( $n_m = n_m(\lambda)$ ) results in a wavelength dependence of the group velocity, i.e., in the spreading of optical pulses.

The group delay  $\tau_m(L)$  of the eigenmode  $m$  for a transmission line of length  $L$  is given by

$$\tau_m(L) = \frac{L}{v_m} = \frac{L}{c_m} \left( 1 - \frac{\lambda}{n_m} \frac{\partial n_m}{\partial \lambda} \right). \quad (3.41)$$

For multi-mode waveguides and fibers the varying group delays result in *intermodal dispersion*, i.e., in a spreading of pulses which occurs even for single-frequency transmission.

Polarization dispersion is an interesting variant of intermodal dispersion. It occurs in any waveguide whose TE- and TM-like eigenmodes are not degenerated.

Due to the different group delays the phase shift between the transverse vector components of a traveling wave varies along the waveguide, i.e., the polarization state of the wave is not conserved.<sup>5</sup> For example, a linear polarized wave is converted into a elliptic polarized when it passes through a elliptical fiber provided it has the proper length and its principal axes are tilted against the axis of polarization. A technical application of polarization conversion is polarization scrambling, where a piece of elliptical fiber enables the operation of polarization sensitive transmission systems (usually coherent systems) with standard single-mode fibers at the expense of 3 dB extra loss.

### 3.5 UNIVERSAL APPROXIMATION SCHEMES

We will now discuss two methods – the Ritz variational principle and the Rayleigh-Schrödinger perturbation theory – which are frequently used to derive approximate eigenvalues and eigenmodes of Hamiltonians. Both schemes are universal; i.e., they are independent of the waveguide geometry and/or of any previous approximations.

#### 3.5.1 The Ritz Variational Principle

The Ritz variational principle can be applied to any Hermitian Hamiltonian  $\mathcal{H}$  which has a limited spectrum. Since the spectra of the operators in integrated optics have an upper limit, we will concentrate on this case.<sup>6</sup> Let

$$\mathcal{H} |m\rangle = \epsilon_m |m\rangle$$

be the abstract eigenvalue problem of the Hermitian operator  $\mathcal{H}$  and

$$\epsilon_0 = \max\{\epsilon_m\}$$

the dielectric constant of its fundamental mode. Then, the mean dielectric constant of the operator  $\mathcal{H}$  calculated using a normalized test field  $|\phi_{\text{test}}\rangle$  will always be smaller than (or equal to) the dielectric constant of the fundamental mode, i.e.,

$$\langle \phi_{\text{test}} | \mathcal{H} \phi_{\text{test}} \rangle = \bar{\epsilon} \leq \epsilon_0. \quad (3.42)$$

In physical terms, the higher the mean dielectric constant the better the approximation of the fundamental mode. To prove this crucial theorem, we expand the test field

$$|\phi_{\text{test}}\rangle = \sum_m \phi_{\text{test}}^{(m)} |m\rangle$$

5. A detailed examination of polarization states is found in Section 2.1.5.

6. Using the transformations  $\mathcal{H} \rightarrow -\mathcal{H}$  and  $\epsilon \rightarrow -\epsilon$ , we recover the Ritz variational principle of quantum mechanics.

into eigenmodes of the Hamiltonian  $\mathcal{H}$  and directly obtain

$$\langle \phi_{\text{test}} | \mathcal{H} \phi_{\text{test}} \rangle = \sum_m \epsilon_m |\phi_{\text{test}}^{(m)}|^2 \leq \epsilon_0 \sum_m |\phi_{\text{test}}^{(m)}|^2 = \epsilon_0.$$

If we apply this theorem to a family of test fields with  $n$  arbitrarily chosen parameters  $\alpha_i$ , we can conclude that the best approximation of the fundamental mode (within the framework of the underlying test functions) is achieved if the mean dielectric constant becomes the absolute maximum, i.e.,

$$\max_{\alpha_i} (\langle \phi_{\text{test}}(\alpha_i) | \mathcal{H} \phi_{\text{test}}(\alpha_i) \rangle) = \bar{\epsilon} \leq \epsilon_0. \quad (3.43)$$

In most cases, it is sufficient to guarantee that the first derivative of the mean dielectric constant

$$\frac{\partial \langle \phi_{\text{test}}(\alpha_i) | \mathcal{H} \phi_{\text{test}}(\alpha_i) \rangle}{\partial \alpha_i} = 0 \quad (3.44)$$

with respect to any of the parameters  $\alpha_i$  vanishes.

If a test function is orthogonal to the fundamental mode, we can easily show that the mean dielectric constant must be even smaller than the eigenvalue of the first excited mode. If it is orthogonal to the first  $n - 1$  eigenmodes of the Hamiltonian  $\mathcal{H}$

$$\langle \phi_{\text{test}} | 0 \rangle = \dots = \langle \phi_{\text{test}} | n - 1 \rangle = 0$$

the mean dielectric constant must be even smaller than the eigenvalue of the  $n$ th excited mode. Thus, we can also apply Ritz's method to approximate the excited modes of the Hamiltonian  $\mathcal{H}$  provided that a proper set of test fields can be constructed (e.g., using symmetries of the Hamiltonian). Nevertheless, Ritz's method is usually applied to study the fundamental mode.

In integrated optics the Ritz variational principle is usually applied to introduce equivalent waveguide structures, i.e., those with analytical solutions which replace the original waveguide structures. The most prominent example of this kind is given by the equivalent parabolic profiles for slab or circular waveguides.<sup>7</sup> The procedure works as follows: First, we choose an equivalent waveguide and select a set of variational parameters. We then consider the ground state of this waveguide as a test field and determine the best set of parameters using the variational principle (3.43).

### 3.5.2 Rayleigh-Schrödinger Perturbation Theory

We can divide many eigenvalue problems into an “essential” part described by the Hamiltonian  $\mathcal{H}_0$  and a (small) perturbation described by  $\mathcal{H}_1$ . This formulation

makes sense if the eigenvalue problem of the unperturbed Hamiltonian  $\mathcal{H}_0$  can be solved exactly or at least much more efficiently than the full eigenvalue problem.

To formulate the Rayleigh-Schrödinger perturbation theory, we consider the family of Hamiltonians

$$\mathcal{H}(\vartheta) = \mathcal{H}_0 + \vartheta \mathcal{H}_1 \quad (3.45)$$

where  $\vartheta$  acts as a parameter of “smallness”. Thus, the second term  $\vartheta \mathcal{H}_1$  represents a “small” perturbation in comparison to the unperturbed Hamiltonian  $\mathcal{H}_0$ ; i.e., we expect that the eigenmodes of the full Hamiltonian are “similar” to those of the unperturbed eigenvalue problem

$$\mathcal{H}_0 |l^{(0)}\rangle = \epsilon_l^{(0)} |l^{(0)}\rangle.$$

Thus, we prefer to express the eigenvalues and eigenmodes of the full Hamiltonian  $\mathcal{H}$  in terms of eigenvalues and eigenmodes of the unperturbed Hamiltonian. To account for degenerated eigenvalues we have introduced the vector  $|l^{(0)}\rangle$  which for the moment describes a freely chosen eigenmode of the eigenspace  $l$ . We further assume that the perturbation is analytic,<sup>8</sup> i.e., that we can expand the eigenvalues

$$\epsilon_l = \sum_{n=0}^{\infty} \epsilon_l^{(n)} \vartheta^n \quad (3.46)$$

and eigenmodes

$$|l\rangle = \sum_{n=0}^{\infty} \vartheta^n |l^{(n)}\rangle \quad (3.47)$$

of the full Hamiltonian into power series. For the sake of simplicity, we dispense for the moment with normalization of the eigenmodes  $|l\rangle$  and choose the corrections  $|l^{(n)}\rangle$  such that they are orthogonal to the eigenmodes of the unperturbed Hamiltonian, i.e.,

$$\langle l^{(0)} | l^{(n)} \rangle = 0. \quad (3.48)$$

By inserting the expansions into the eigenvalue problem for the full Hamiltonian, we obtain

$$\sum_{n=0}^{\infty} (\mathcal{H}_0 + \vartheta \mathcal{H}_1) \vartheta^n |l^{(n)}\rangle = \sum_{m,n=0}^{\infty} \vartheta^{m+n} \epsilon_l^{(m)} |l^{(n)}\rangle.$$

Since this equation must hold for any  $\vartheta$ , we can derive a hierarchy of equations which correspond to the powers of the smallness parameter  $\vartheta$ , i.e.,

$$\begin{aligned} \vartheta^0 : \quad & \mathcal{H}_0 |l^{(0)}\rangle = \epsilon_l^{(0)} |l^{(0)}\rangle \\ \vartheta^1 : \quad & \mathcal{H}_0 |l^{(1)}\rangle + \mathcal{H}_1 |l^{(0)}\rangle = \epsilon_l^{(0)} |l^{(1)}\rangle + \epsilon_l^{(1)} |l^{(0)}\rangle \\ & \vdots \\ \vartheta^m : \quad & \mathcal{H}_0 |l^{(m)}\rangle + \mathcal{H}_1 |l^{(m-1)}\rangle = \sum_{n=0}^m \epsilon_l^{(n)} |l^{(m-n)}\rangle. \end{aligned} \quad (3.49)$$

<sup>7</sup> This substitution is often referred to as a *Gaussian approximation*.

<sup>8</sup> We should note here that this assumption is not always satisfied. The attractive electron-electron interaction in BCS superconductors is a prominent example of a non-analytic perturbation in quantum mechanics.

The first equation in (3.49) ( $O(\vartheta^0)$ ) naturally represents the eigenvalue problem of the unperturbed Hamiltonian  $\mathcal{H}_0$ . If we multiply the second one ( $O(\vartheta^1)$ ) by the eigenmodes  $|l^{(0)}\rangle$  belonging to the eigenspace  $l$  of the unperturbed Hamiltonian, we obtain a sequence of eigenvalue problems

$$\left| \langle l^{(0)} | \mathcal{H}_1 l^{(0)} \rangle - \epsilon_l^{(1)} \mathcal{E} \right| = 0 \quad (3.50)$$

for the first-order corrections of the eigenvalues  $\epsilon_l^{(1)}$ . In addition, we find the proper eigenmodes  $|l_\nu^{(0)}\rangle$  for the corrected (and in many cases now non-degenerated) eigenvalues  $\epsilon_{l\nu}^{(1)}$ . The eigenvalues corrected up to the first order turn out to be the mean dielectric constants in which the eigenmodes of the unperturbed Hamiltonian act as the test field

$$\epsilon_{l\nu} \approx \epsilon_l^{(0)} + \vartheta \epsilon_{l\nu}^{(1)} = \langle l_\nu^{(0)} | \mathcal{H}_1 l_\nu^{(0)} \rangle. \quad (3.51)$$

Applying Ritz's variational principle to Equation (3.51), we can conclude that *the effective dielectric constant of the fundamental mode obtained by a first-order perturbation theory is always smaller than the true eigenvalue.*

To calculate the corrected eigenmodes, we multiply the second equation with all other eigenmodes of  $\mathcal{H}_0 (\mathbf{m} \neq \mathbf{l})$  and obtain a second sequence of equations

$$\langle \mathbf{m}^{(0)} | l^{(1)} \rangle = \frac{\langle \mathbf{m}^{(0)} | \mathcal{H}_1 l^{(0)} \rangle}{\epsilon_l^{(0)} - \epsilon_m^{(0)}} \quad (3.52)$$

which allows us to expand corrections of the eigenmodes  $|l^{(1)}\rangle$  into the properly chosen eigenmodes of the unperturbed Hamiltonian we have calculated earlier. Due to the resonance denominator, the correction caused by the other eigenmodes will decrease with their increasing spectral distance. We have now elaborated a first-order perturbation theory.

We will stop the successive solution of (3.49) here for two reasons. First, the calculation of further corrections results in lengthy expressions and, much more importantly, the power series of Rayleigh-Schrödinger perturbation theory are in most cases asymptotic expansions; i.e., the best results are obtained if only one or two corrections are taken into account.

Perturbation theory has a couple of applications in integrated optics. First, it can be used to treat any type of weakly perturbed waveguides and fibers. Second, *for weakly guiding waveguides* (see Section 3.3) *the coupling of the vector components via  $\nabla_t(\ln \epsilon)$  can be treated as a perturbation*. The vector correction formulas will turn out to be corrections according to a first-order perturbation theory.

### 3.6 WEAKLY GUIDING WAVEGUIDES

Today, most integrated optical devices are analyzed without taking the hybrid nature of their eigenmodes into account. We will show in this section that weakly

guiding waveguides characterized by a low dielectric contrast ( $\Delta\epsilon \ll \bar{\epsilon}$ ) and large near fields ( $\lambda/\sqrt{\epsilon} \ll \text{spotsize}$ ) can be treated in this way. It should be noted that the reduced eigenvalue problems derived here form the basis for the treatment of most waveguide structures in integrated optics. Only for a few devices, such as the purely passive polarization converter whose operation is governed by the chip-to-air interface, must the hybrid nature of the eigenmodes be taken into account.

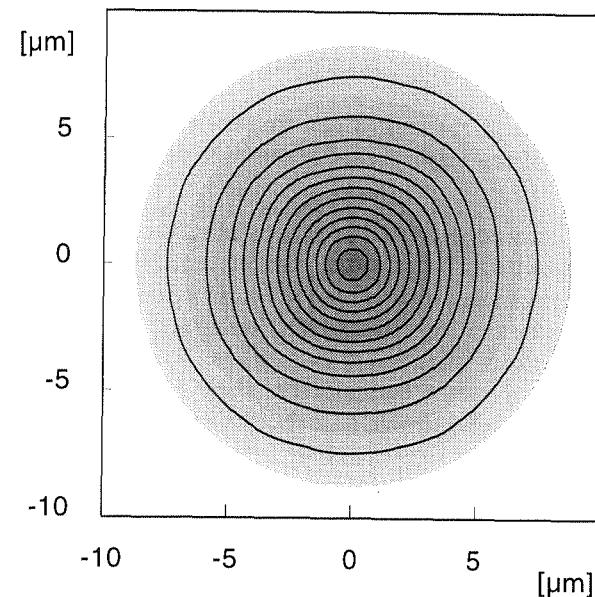


Figure 3.5. Nearly uniform waveguide and its fundamental mode (contours).

#### 3.6.1 Nearly Uniform Waveguides

We will first discuss the nearly uniform dielectric profiles (see Figure 3.5) whose variation  $\Delta\epsilon$  is small over the whole transverse plane compared to the mean value  $\bar{\epsilon}$

$$|\Delta\bar{\epsilon}| \ll \bar{\epsilon}$$

In addition, the near field of an eigenmode  $\phi$  varies little over one wavelength inside the material, i.e.,

$$|\nabla_t \phi| \ll |k_0 \sqrt{\epsilon} \phi|$$

where  $\phi$  is any field component of the electric or magnetic field. We can then expect that the two conditions

$$|\mathcal{H}_e \mathbf{e}_t| = |\nabla_t (\nabla_t (\ln \epsilon) \cdot \mathbf{e}_t)| \ll |k_0^2 \Delta \epsilon \mathbf{e}_t| \quad (3.53)$$

and

$$|\mathcal{H}_h \mathbf{h}_t| = |\nabla_t (\ln \epsilon) \times (\nabla_t \times \mathbf{h}_t)| \ll |k_0^2 \Delta \epsilon \mathbf{h}_t| \quad (3.54)$$

are also satisfied.<sup>9</sup> In physical terms; the eigenmodes of weakly guiding waveguides should be only weakly hybrid; i.e., we can then treat the coupling of vector components as a perturbation within the framework of a Rayleigh-Schrödinger perturbation theory.

Using the Equations (3.11) and (3.12), we see easily that the vector **E**-field and **H**-field formulations of the unperturbed eigenvalue problem

$$\frac{1}{k_0^2} (\Delta_t + k_0^2 \epsilon) \mathbf{e}_{t0}^{(m)} = \epsilon_m^{(0)} \mathbf{e}_{t0}^{(m)} \quad (3.55)$$

and

$$\frac{1}{k_0^2} (\Delta_t + k_0^2 \epsilon) \mathbf{h}_{t0}^{(m)} = \epsilon_m^{(0)} \mathbf{h}_{t0}^{(m)} \quad (3.56)$$

coincide. For Cartesian vector components, these equations decouple completely and we obtain the eigenvalue problem of the scalar Helmholtz equation:

$$\mathcal{H}_s \phi^{(m)} = \frac{1}{k_0^2} (\nabla_t^2 + k_0^2 \epsilon) \phi^{(m)} = \epsilon_m^{(0)} \phi^{(m)} \quad (3.57)$$

where  $\phi^{(m)}$  represents any of the transverse Cartesian components of the electric ( $e_x, e_y$ ) or the magnetic field ( $h_x, h_y$ ).

Before we continue the elaboration of perturbation theory, we will briefly discuss the eigenvalue problem of the scalar Helmholtz equation. Obviously, Equation (3.57) can be easily transformed into the two-dimensional time-independent Schrödinger equation

$$\left( -\frac{\hbar^2}{2m} \nabla_t^2 + V(\mathbf{r}_t) \right) \psi_m = E_m \psi_m.$$

In consequence, both mathematical methods and analytical solutions obtained within the framework of classical quantum mechanics can be applied to weakly guiding waveguides.

Applying Green's theorem for the transverse plane

$$\int_{A_\infty} da (\phi \nabla_t^2 \psi - \psi \nabla_t^2 \phi) = \int_{\partial A_\infty} dl \cdot (\phi \nabla_t \psi - \psi \nabla_t \phi)$$

9. This argument also holds for step-index structures, since we can replace  $\nabla_i \epsilon \approx k_0 \epsilon$  without changing the spectrum significantly.

and keeping in mind that the optical field of an eigenmode  $\phi^{(m)}$  vanishes at infinity we see that the scalar Laplacian  $\nabla_t^2$  is a Hermitian operator with respect to the scalar product

$$\langle b | c \rangle = \int_{A_\infty} da \mathbf{f}_{tb}^*(\mathbf{r}_t) \cdot \mathbf{f}_{tc}(\mathbf{r}_t) \quad (3.58)$$

where  $\mathbf{f}_{ta}(\mathbf{r}_t)$  denotes the transverse components of either the electric or the magnetic field. We can also derive this scalar product from the more general results obtained in Section 3.2.4. If we neglect all derivatives of the dielectric profile ( $\nabla_i \epsilon = 0$ ), the vectorial scalar products for both the vector **E**-field and **H**-field formulations become

$$\langle b | c \rangle = \int_{A_\infty} da \mathbf{f}_{tb}^*(\mathbf{r}_t) \cdot \mathcal{H}_s \mathbf{f}_{tc}(\mathbf{r}_t).$$

Since the Hermitian operator  $\mathcal{H}_s$  occurring in this scalar product is the Hamiltonian of the scalar Helmholtz equation, we can replace the vectorial scalar product in a second step by the expression (3.58). We can further conclude that *the unperturbed Hamiltonian  $\mathcal{H}_s$  describing a lossless, nearly uniform waveguide structure is Hermitian; i.e., its eigenvalues are real and its eigenmodes form an orthogonal set of functions.*

As expected from these considerations, the eigenvalues of the unperturbed Hamiltonian are degenerated with respect to polarization. Their hybrid character is introduced via the perturbations  $\mathcal{H}_e$  and  $\mathcal{H}_h$  of the Hamiltonian. The first order corrections  $\epsilon_m^{(1)}$  of the effective dielectric constants  $\epsilon_m^{(0)}$  according to the Rayleigh-Schrödinger perturbation theory are solutions of the matrix eigenvalue problem (see Equation (3.50)):

$$|\langle \mathbf{m}^{(0)} | \mathcal{H}_1 \mathbf{m}^{(0)} \rangle - \epsilon_m^{(1)} \mathcal{E}| = 0 \quad (3.59)$$

where  $\mathcal{H}_1$  describes the perturbation ( $\mathcal{H}_e$  or  $\mathcal{H}_h$ ) and  $|\mathbf{m}^{(0)}\rangle$  stands for an arbitrary set of eigenmodes spanning the eigenspace of the unperturbed effective dielectric constant  $\epsilon_m^{(0)}$ . If the effective dielectric constants, for example, are degenerated only with respect to polarization, the eigenspace will be two-dimensional and we can use the two eigenmodes which are linearly polarized parallel to the axes of the Cartesian coordinate system to span the eigenspace.

We have seen in Section 3.5.2 that the effective dielectric constant obtained using the first-order Rayleigh-Schrödinger perturbation theory is smaller than or equal to the true eigenvalue. We can conclude straightaway that *the effective dielectric constants calculated by a vector correction formula are always smaller than or equal to the true dielectric constants.*

For the vector **E**-field formulation, the matrix elements are

$$\langle m_\mu^{(0)} | \mathcal{H}_e m_\nu^{(0)} \rangle = \int_{A_\infty} da \mathbf{e}_{t0}^{(m\mu)*} \cdot \nabla_t [\nabla_t (\ln \epsilon) \cdot \mathbf{e}_{t0}^{(m\nu)}] \quad (3.60)$$

in which the eigenmodes  $\mathbf{e}_{t0}^{(m\nu)}$  represent any orthogonal set of eigenmodes spanning the eigenspace of the eigenvalue  $\epsilon_m^{(0)}$ . In the same way we obtain the matrix elements

$$\langle m_\mu^{(0)} | \mathcal{H}_h m_\nu^{(0)} \rangle = \int_{A_\infty} da \mathbf{h}_{t0}^{(m\mu)*} \cdot \nabla_t (\ln \epsilon) \times (\nabla_t \times \mathbf{h}_{t0}^{(m\nu)}) \quad (3.61)$$

for the vector  $\mathbf{H}$ -field formulation.

The solutions of the eigenvalue problem (3.59) give the leading corrections of the scalar waves. For this reason, Equation (3.59) and related formulations are called *vector correction formulas*. The calculated eigenmodes  $|m_\mu^{(0)}\rangle$  form a proper basis within the eigenspace; i.e., they determine the “optical axes”. The corrections caused by the other eigenmodes are

$$\langle \mathbf{m}^{(0)} | \mathbf{l}^{(1)} \rangle = \frac{\langle \mathbf{m}^{(0)} | \mathcal{H}_1 \mathbf{l}^{(0)} \rangle}{\epsilon_l^{(0)} - \epsilon_m^{(0)}}. \quad (3.62)$$

Due to the resonant character of the term on the right-hand side, only closely spaced eigenvalues will contribute to these corrections.

We can calculate the longitudinal field components  $e_z$  and  $h_z$  using Equations (3.13) derived in Section 3.2. The required transverse field components are calculated by perturbation theory. Since the near fields of weakly guiding waveguides do not vary significantly over one wavelength, the longitudinal field components are significantly smaller than the dominating transverse field components. Typically, the amplitudes differ by more than one order of magnitude.

### 3.6.2 Slab-like Waveguide Structures

We will now discuss waveguides which guide their eigenmodes strongly along the  $x$ -direction and weakly along the  $y$ -direction of a properly chosen coordinate system (see Figure 3.6). Such waveguide structures are typical of a huge class of devices in semiconductor material systems (e.g., InGaAsP/InP).

The dielectric profile of such waveguide structures does not vary significantly along straight lines with  $x = \text{const}$ , i.e.,

$$|\Delta\epsilon|_{x=\text{const}} \ll \bar{\epsilon}.$$

For reasons stated in the previous section, we can again assume that the derivative  $\partial\phi/\partial y$  is small in comparison to the wave number inside the material, even for step-index structures. The lateral extent of the near field (along the  $y$ -direction) is then large compared with the wavelength inside the material, and we can expect that the lateral derivatives of the field components (indicated by character  $\phi$ )

$$\left| \frac{\partial\phi}{\partial y} \right| \ll |k_0 \sqrt{\epsilon} \phi|$$

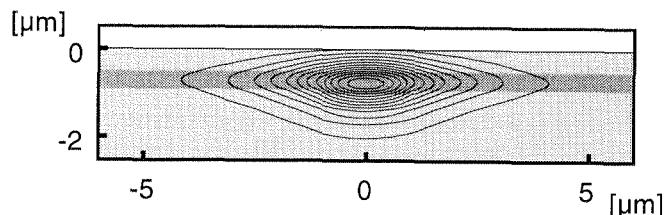


Figure 3.6. Buried-rib waveguide, a typical slab-like waveguide structure.

are small compared with the wave number inside the material. Under such circumstances, the terms

$$\mathcal{H}_e \mathbf{e}_t = \nabla_t \nabla_t (\ln \epsilon) \cdot \mathbf{e}_t - \mathbf{i}_x \frac{\partial}{\partial x} \frac{\partial (\ln \epsilon)}{\partial x} e_x \quad (3.63)$$

and

$$\mathcal{H}_h \mathbf{h}_t = \nabla_t (\ln \epsilon) \times \nabla_t \times \mathbf{h}_t + -\mathbf{i}_y \frac{\partial (\ln \epsilon)}{\partial x} \frac{\partial h_y}{\partial x} \quad (3.64)$$

can be regarded as perturbations in the sense of a Rayleigh-Schrödinger perturbation theory.

We start the discussion of the unperturbed eigenvalue problem with its vector  $\mathbf{E}$ -field formulation

$$\begin{aligned} \frac{1}{k_0^2} \left( \nabla^2 + k_0^2 \epsilon + \frac{\partial}{\partial x} \frac{\partial (\ln \epsilon)}{\partial x} \right) e_{x0}^{(m)} &= \epsilon_m^{(0)} e_{x0}^{(m)} \\ \frac{1}{k_0^2} \left( \nabla^2 + k_0^2 \epsilon \right) e_{y0}^{(m)} &= \epsilon_m^{(0)} e_{y0}^{(m)}. \end{aligned}$$

Obviously, the two transverse vector components are decoupled, but they are solutions of two different eigenvalue equations. Thus the unperturbed eigenvalue problem has two non-degenerated solutions, one with  $e_{x0}^{(m)} = 0$  and a second one with  $e_{y0}^{(m)} = 0$ . Using Equation (3.13), we see that the longitudinal field component

$$-iq_m e_{z0}^{(m)} = \frac{\partial e_{y0}^{(m)}}{\partial y} + \frac{\partial (\ln \epsilon)}{\partial y} e_{y0}^{(m)}$$

can be neglected in the zeroth approximation; i.e., the first solution of the eigenvalue problem is a nearly transverse electric mode. Since we now expect that the second solution is a nearly transverse magnetic mode, we turn our interest to the vector  $\mathbf{H}$ -field formulation

$$\frac{1}{k_0^2} \left( \nabla^2 + k_0^2 \epsilon \right) h_{x0}^{(m)} = \epsilon_m^{(0)} h_{x0}^{(m)}$$

$$\frac{1}{k_0^2} \left( \nabla^2 + k_0^2 \epsilon - \frac{\partial (\ln \epsilon)}{\partial x} \frac{\partial}{\partial x} \right) h_{y0}^{(m)} = \epsilon_m^{(0)} h_{y0}^{(m)}$$

of the eigenvalue problem, which for the same reasons delivers one solution with  $h_{x0}^{(m)} = 0$  and another one with  $h_{y0}^{(m)} = 0$ . The longitudinal field component for the solution with  $h_{x0}^{(m)} = 0$

$$-iq_m h_{z0}^{(m)} = \frac{\partial h_{y0}^{(m)}}{\partial y}$$

can again be neglected in the zeroth approximation; i.e., the second solution is a nearly transverse magnetic mode. To sum up; *the unperturbed Hamiltonian of a slab-like waveguide structure supports only transverse eigenmodes. The TE modes are solutions of the scalar eigenvalue problem*

$$\frac{1}{k_0^2} (\nabla^2 + k_0^2 \epsilon) e_{y0}^{(m)} = \epsilon_{\text{TE},m}^{(0)} e_{y0}^{(m)} \quad (3.65)$$

and the TM modes are those of

$$\frac{1}{k_0^2} \left[ \nabla^2 + k_0^2 \epsilon - \frac{\partial(\ln \epsilon)}{\partial x} \frac{\partial}{\partial x} \right] h_{y0}^{(m)} = \epsilon_{\text{TM},m}^{(0)} h_{y0}^{(m)}. \quad (3.66)$$

It will be shown in the next section that Equations (3.65) and (3.66) are generalizations of the eigenvalue equations for a slab waveguide.

Using partial integration, we can easily show that the operator

$$\frac{\partial(\ln \epsilon)}{\partial x} \frac{\partial}{\partial x}$$

is Hermitian with respect to the simplified scalar product (3.58); i.e., we can also use this scalar product for slab-like waveguide structures. We can further conclude that *the unperturbed Hamiltonian describing the TE and TM modes of a lossless slab-like waveguide structure is Hermitian; i.e., its eigenvalues are real and its eigenmodes form an orthogonal set of eigenmodes.*

Since the eigenmodes of the unperturbed Hamiltonian (TE and TM modes) are not degenerated we need not be concerned about the diagonalization of the eigenspaces within the framework of perturbation theory. For the TE modes we obtain the vector correction formula

$$\epsilon_{\text{TE},m}^{(1)} = \int_{A_\infty} da e_{y0}^{(m)*} \frac{\partial}{\partial y} \left[ \frac{\partial(\ln \epsilon)}{\partial y} e_{y0}^{(m)} \right]. \quad (3.67)$$

The corresponding formula for the TM case is given by

$$\epsilon_{\text{TM},m}^{(1)} = \int_{A_\infty} da h_{y0}^{(m)*} \frac{\partial(\ln \epsilon)}{\partial y} \frac{\partial h_{y0}^{(m)}}{\partial x}. \quad (3.68)$$

The corrections of eigenmodes and the longitudinal field components are calculated in the same way as described in the previous section.

### 3.7 ANALYTICAL SOLUTIONS

Closed solutions of the eigenvalue problem are known only for a few dielectric profiles. We will restrict our discussion of analytical solutions to the two most important examples, the parabolic slab and the circular parabolic fiber.<sup>10</sup> The dielectric profile is in both cases given by

$$\epsilon(\rho) = \epsilon_{\text{co}} - \frac{4\rho^2}{k_0^2 r_0^4} \quad (3.69)$$

in which  $\rho$  is the Cartesian coordinate for the parabolic slab and the radial one for the circular parabolic fiber. Here  $\epsilon_{\text{co}}$  is the “core” dielectric constant, the parameter  $r_0$  (more precisely its inverse) is a measure of the guidance strength. For both profiles it will turn out to be the field radius of the fundamental mode. This profile is obviously non-physical since the dielectric profile diverges at infinity; namely,  $\epsilon(\rho) \rightarrow -\infty$  as  $\rho \rightarrow \pm\infty$ . However, provided the guidance is strong enough the optical field of the first few modes will be confined to the regions where the dielectric profile is positive.

In fact, both dielectric profiles form the basis of a Gaussian approximation which replaces the original waveguide by an equivalent parabolic profile in the treatment by the Ritz variational principle. Moreover, the diffraction patterns of the eigenmodes of the parabolic waveguides will turn out to be the Gaussian beams which will play a central role in the analysis of fiber-chip coupling, diffraction gratings and phased arrays.

#### 3.7.1 The Parabolic Slab

As already mentioned, the parabolic slab is described by the dielectric profile

$$\epsilon(x) = \epsilon_{\text{co}} - \frac{4x^2}{k_0^2 r_0^4}$$

and the eigenvalue problem for its TE modes<sup>11</sup> is given by

$$\frac{1}{k_0^2} \left( \frac{d^2}{dx^2} + k_0^2 \epsilon_{\text{co}} - \frac{4x^2}{r_0^4} \right) e_y^{(m)} = \epsilon_m e_y^{(m)}. \quad (3.70)$$

First, we simplify the underlying mathematical problem by the substitutions

$$\begin{aligned} \xi &= \frac{\sqrt{2}x}{r_0} \\ \eta_m &= \frac{k_0^2 r_0^2 (\epsilon_{\text{co}} - \epsilon_m)}{4} - \frac{1}{2} \end{aligned} \quad (3.71)$$

10. The book by Snyder and Love [1] shows more examples of analytical solutions such as the step-index fiber and clad-parabolic profiles.

11. See Equation (3.103).

and obtain the new eigenvalue problem

$$\mathcal{N} |m\rangle = \eta_m |m\rangle \quad (3.72)$$

with the Hermitian operator

$$\mathcal{N} = \frac{1}{2} (\mathcal{Q}_\xi^2 + \xi^2 - \mathcal{E}) \quad (3.73)$$

and the momentum operator

$$\mathcal{Q}_\xi = -i \frac{d}{d\xi}.$$

We have chosen the character  $\mathcal{N}$  here for the Hamiltonian of the eigenvalue problem (3.72) since it will turn out that the operator  $\mathcal{N}$  enumerates the excitation levels.

We will now solve the eigenvalue problem (3.72) by a purely algebraic technique. For this purpose, we introduce two more operators

$$\mathcal{S}_\pm = \frac{1}{\sqrt{2}} (\xi \mp i \mathcal{Q}_\xi) \quad (3.74)$$

which are non-Hermitian but are adjoint to each other; i.e.,

$$\mathcal{S}_\pm = \mathcal{S}_\mp^\dagger. \quad (3.75)$$

These operators are called *staircase operators*. The reason for this designation will become clear soon. By using the staircase operator, we can represent the Hamiltonian by the simple product

$$\mathcal{N} = \mathcal{S}_+ \mathcal{S}_-. \quad (3.76)$$

By using the commutation rule

$$[\mathcal{Q}_\xi, \xi] = \mathcal{E}$$

which is easily proved by partial integration, we can easily derive the two crucial commutation rules

$$\begin{aligned} [\mathcal{S}_+, \mathcal{S}_-] &= \mathcal{E} \\ [\mathcal{S}_\pm^n, \mathcal{N}] &= \mp n \mathcal{S}_\pm^n \end{aligned} \quad (3.77)$$

of the staircase operators. By using these commutation rules, we can show in turn that the Hamiltonian

$$\mathcal{N} |\mathcal{S}_\pm^n m\rangle = (\eta_m \pm n) |\mathcal{S}_\pm^n m\rangle \quad (3.78)$$

satisfies a collection of related eigenvalue problems. Obviously, not only the vector  $|m\rangle$  is an eigenmode of the Hamiltonian  $\mathcal{N}$  but also the vector  $|\mathcal{S}_\pm^n m\rangle$ . Since the scalar product of one vector with itself must always be non-negative, i.e.,

$$0 \leq \langle \mathcal{S}_-^{n+1} m | \mathcal{S}_-^{n+1} m \rangle = \langle \mathcal{S}_-^n m | \mathcal{N} \mathcal{S}_-^n m \rangle = (\eta_m - n) \langle \mathcal{S}_-^n m | \mathcal{S}_-^n m \rangle$$

we can further conclude that all eigenvalues of the Hamiltonian  $\mathcal{N}$  must be non-negative and that the condition

$$\mathcal{S}_- |0\rangle = 0 \quad (3.79)$$

must be satisfied. Using the representation  $\mathcal{N} = \mathcal{S}_+ \mathcal{S}_-$ , we see directly that as a further consequence the eigenvalue of the fundamental state must vanish ( $\eta_0 = 0$ ). The spectrum of the Hamiltonian  $\mathcal{N}$

$$\eta_m = m \quad (3.80)$$

is given by the integers starting with zero. Thus, the operator  $\mathcal{N}$  enumerates the excitation levels. It is often called the enumeration operator.

We can now also understand why the operators  $\mathcal{S}_\pm$  are called staircase operators. In particular, Equation (3.78) shows that the operator

$$\mathcal{S}_+ |m\rangle = \sqrt{m+1} |m+1\rangle \quad (3.81)$$

increases the level of excitation by one; i.e., it ascends one step within the spectrum. The operator

$$\mathcal{S}_- |m\rangle = \sqrt{m} |m-1\rangle \quad (3.82)$$

in contrast, decreases the level of excitation by one; i.e., it descends one step. Obviously, the  $m$ th excited state can be expressed in terms of the fundamental state by

$$|m\rangle = \frac{1}{\sqrt{m!}} \mathcal{S}_+^m |0\rangle. \quad (3.83)$$

Figure 3.7 sums up the results obtained up to now.<sup>12</sup>

In the last step, we return to the analytical representation of Equation (3.79)

$$\left( \xi + \frac{d}{d\xi} \right) e_y^{(0)} = 0$$

and calculate the fundamental mode of the parabolic slab

$$e_y^{(0)} = e^{-\xi^2/2}$$

by separation of the variables. By using the identity

$$\mathcal{S}_+^m e_y^{(0)} = \frac{(-1)^m}{\sqrt{2^m m!}} e^{\xi^2/2} \frac{d^m}{d\xi^m} (e^{-\xi^2/2} e_y^{(0)})$$

12. It should be noted that the whole analysis was carried out using the abstract operators and their commutation rules. It can be applied to any other problem which can be formulated in terms of staircase operators.

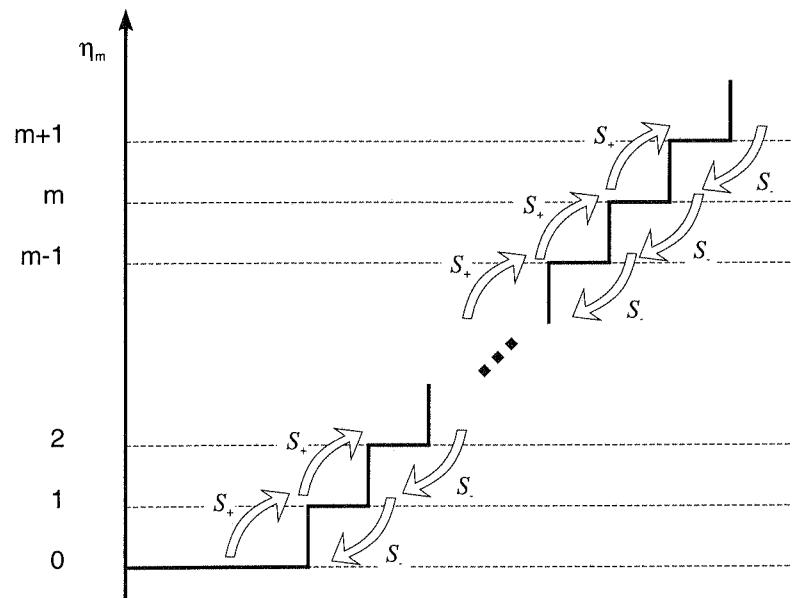


Figure 3.7. Staircase operators.

we can then express all excited modes in terms of the fundamental mode  $e_y^{(0)}$ . As a result of this process, we obtain

$$e_y^{(m)}(x) = \left(2^{m-1/2} m! r_0 \sqrt{\pi}\right)^{-1/2} H_m(\sqrt{2}x/r_0) e^{-x^2/r_0^2} \quad (3.84)$$

with the Hermite polynomials

$$H_m(\xi) = (-1)^m e^{\xi^2} \frac{d^m}{d\xi^m} e^{-\xi^2}. \quad (3.85)$$

The first few representatives of these polynomials are

$$H_0(\xi) = 1$$

$$H_1(\xi) = 2\xi$$

$$H_2(\xi) = 4\xi^2 - 2.$$

Inserting the explicit expressions (3.74) for the staircase operators  $S_{\pm}$  and (3.84) for the eigenmodes of the parabolic slab into Equations (3.81) and (3.82) we obtain two useful recurrence formulas

$$\frac{d H_m(\xi)}{d\xi} = 2m H_{m-1}(\xi) \quad (3.86)$$

and

$$H_{m+1}(\xi) = 2\xi H_m(\xi) - 2m H_{m-1}(\xi) \quad (3.87)$$

for the Hermite polynomials.

We can finally calculate the effective dielectric constants of the dielectric slab

$$\epsilon_m = \epsilon_{co} - \frac{4}{k_0^2 r_0^2} \left(m + \frac{1}{2}\right) \quad (3.88)$$

using Equations (3.71) and (3.80).

### 3.7.2 The Circular Parabolic Fiber

We will now consider an arbitrary Cartesian field component  $\phi^{(lm)}$  of either the electric or magnetic field of a weakly guiding circular parabolic fiber, which is the solution of the eigenvalue problem

$$\frac{1}{k_0^2} (\nabla_t^2 + k_0^2 \epsilon) \phi^{(lm)} = \epsilon_{lm} \phi^{(lm)}$$

as we have shown in Section 3.6. We will now derive two alternative sets of eigenmodes for the circular parabolic fiber by using two different coordinate systems, Cartesian and cylindrical coordinates.

#### Hermite-Gaussian Modes

First, we will derive the eigenmodes of the circular parabolic fiber in Cartesian coordinates. The eigenvalue problem is given by

$$(\mathcal{H}_x + \mathcal{H}_y - \epsilon_{co} \mathcal{E}) \phi_{HG}^{(lm)}(x, y) = \epsilon_{lm}^{(HG)} \phi_{HG}^{(lm)}(x, y) \quad (3.89)$$

with the two reduced Hamiltonians ( $s = x, y$ )

$$\mathcal{H}_s = \frac{1}{k_0^2} \left( \frac{\partial^2}{\partial s^2} + k_0^2 \epsilon_{co} - \frac{4s^2}{r_0^2} \right).$$

The label *HG* stands for Hermite-Gaussian modes. The two reduced Hamiltonians  $\mathcal{H}_x$  and  $\mathcal{H}_y$  describe two identical parabolic slabs treated in the previous section, i.e.,

$$\mathcal{H}_s \phi^{(m)}(s) = \left[ \epsilon_{co} - \frac{4}{k_0^2 r_0^2} \left(m + \frac{1}{2}\right) \right] \phi^{(m)}(s)$$

with the Hermite-Gaussian eigenmodes

$$\phi^{(m)}(s) = \left(2^{m-1/2} m! r_0 \sqrt{\pi}\right)^{-1/2} H_m(\sqrt{2}s/r_0) e^{-s^2/r_0^2}.$$

The full eigenvalue problem (3.89) is separable. We will show in Section 3.9 that the eigenvalues of the full Hamiltonian

$$\epsilon_{lm}^{(\text{HG})} = \epsilon_{\text{co}} - \frac{4}{k_0^2 r_0^2} (m + n + 1) \quad (3.90)$$

are given by the sum over the eigenvalues of the reduced Hamiltonians and that the eigenmodes

$$\phi_{\text{HG}}^{(lm)}(x, y) = \phi^{(l)}(x) \phi^{(m)}(y) \quad (3.91)$$

are given by the product of the eigenmodes of the reduced Hamiltonians.

### Laguerre-Gaussian Modes

We will now derive a second set of eigenmodes by using cylindrical coordinates and introducing the representation

$$\phi_{\text{LG}}^{(lm)}(r, \varphi) = \chi^{(lm)}(r) \begin{cases} \cos(l\varphi) \\ \sin(l\varphi) \end{cases} \quad (3.92)$$

we obtain the ordinary differential equation

$$\left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{l^2}{r^2} - \frac{4r^2}{r_0^4} \right) \chi^{(lm)}(r) = -k_0^2 (\epsilon_{\text{co}} - \epsilon_{lm}^{(\text{LG})}) \chi^{(lm)}(r) \quad (3.93)$$

for the radial coordinate  $r$  in which the label  $LG$  stands for Laguerre-Gaussian modes as will become clear soon. We should keep in mind that the index  $l$  applies to the angular coordinate, the eigenmode  $\chi^{(lm)}$  has  $2l$  zeros for a full circle. The index  $m$ , in contrast, applies to the radial coordinate. We will see that it counts the zeros of an eigenmode along the radial axis.

We will treat this analytical example in a more conventional way;<sup>13</sup> i.e., we will reformulate Equation (3.93) by a series of transformations. Let us start with the standard transformation

$$\chi^{(lm)}(r) = r^l F_{lm}(r)$$

which is used to remove the term  $l/r^2$  from the Hamiltonian. By applying this transformation, we obtain the desired result

$$\left( \frac{d^2}{dr^2} + \frac{2l+1}{r} \frac{d}{dr} - \frac{4r^2}{r_0^4} \right) F_{lm}(r) = -k_0^2 (\epsilon_{\text{co}} - \epsilon_{lm}^{(\text{LG})}) F_{lm}(r).$$

We furthermore expect solutions of the type  $\chi^{(lm)} \propto e^{-r^2/r_0^2}$ . Thus, we first define the dimensionless quantity

$$\xi = \frac{2r^2}{r_0^2}$$

13. It should be noted that we can apply also the methods described in the following to the parabolic slab.

and obtain

$$\left[ \xi \frac{d^2}{d\xi^2} + (l+1) \frac{d}{d\xi} - \frac{\xi}{4} \right] F_{lm}(\xi) = -\frac{k_0^2 r_0^2}{8} (\epsilon_{\text{co}} - \epsilon_{lm}^{(\text{LG})}) F_{lm}(\xi).$$

In the next step, we cut off the term  $e^{-r^2/r_0^2}$  by the transformation

$$F_{lm}(\xi) = L_m^l(\xi) e^{-\xi/2}$$

and arrive at a new eigenvalue problem

$$\left[ \xi \frac{d^2}{d\xi^2} + (l+1-\xi) \frac{d}{d\xi} \right] L_m^l(\xi) = -\eta_{lm} L_m^l(\xi) \quad (3.94)$$

in which the eigenvalue  $\eta_{lm}$  is related to the effective dielectric constant  $\epsilon_{lm}$  via

$$\eta_{lm} = \frac{k_0^2 r_0^2}{8} (\epsilon_{\text{co}} - \epsilon_{lm}^{(\text{LG})}) - \frac{l+1}{2}.$$

In the last step, we will tackle the eigenvalue problem (3.94) by expanding of the solution into a power series

$$L_m^l(\xi) = \xi^a \sum_{n=0}^{\infty} \alpha_n \xi^n \quad (3.95)$$

with  $\alpha_0 \neq 0$ . The factor  $\xi^a$  must be introduced since the leading term  $\xi d^2/d\xi^2$  of the differential Equation (3.94) exhibits a weak singularity at the origin ( $\xi = 0$ ). By inserting the expansion and its first two derivatives

$$\begin{aligned} \frac{d L_m^l}{d\xi} &= (a+n) \sum_{n=0}^{\infty} \alpha_n \xi^{a+n-1} \\ \frac{d^2 L_m^l}{d\xi^2} &= (a+n)(a+n-1) \sum_{n=0}^{\infty} \alpha_n \xi^{a+n-2} \end{aligned}$$

into the eigenvalue problem (3.94), we obtain for each parameter  $l$  and  $m$  a sequence of equations for the powers of  $\xi$  which are used to determine the coefficients  $a$  and  $\alpha_n$ . The first equation (belonging to  $\xi^{a-1}$ ) is given by

$$[a(a-1) + (l+1)a] \alpha_0 = 0.$$

It has two solutions  $a = 0$  and  $a = -l$ . The second one, however, is non-physical since it leads to eigenmodes that are infinite at the origin  $\xi = 0$ . Using  $a = 0$ , we obtain the equation

$$\alpha_{n+1} = \frac{n - \eta_{lm}}{(n+l+1)(n+1)} \alpha_n$$

belonging to the power  $\xi^n$ . The eigenmodes of the guided modes must vanish at infinity; i.e., the power expansion of  $L_m^l(\xi)$  must be finite. We can conclude from this that the eigenvalues

$$\eta_{lm} = m \quad (3.96)$$

must be the integers starting with zero. The eigenfunctions are thus given by the generalized Laguerre polynomials

$$L_m^l(\xi) = \sum_{n=0}^m (-1)^n \binom{m+l}{m-n} \frac{\xi^n}{n!}. \quad (3.97)$$

Explicit expressions for the first representatives are

$$\begin{aligned} L_0^l(\xi) &= 1 \\ L_1^l(\xi) &= l + 1 - \xi \\ L_2^l(\xi) &= \frac{(l+2)(l+1)}{2} - (l+2)\xi + \frac{\xi^2}{2}. \end{aligned}$$

The effective dielectric constants of the circular parabolic fiber are given by

$$\epsilon_{lm}^{(LG)} = \epsilon_{co} - \frac{4}{k_0^2 r_0^2} (l + 2m + 1) \quad (3.98)$$

and its eigenmodes are

$$\chi^{(lm)} = \left[ \frac{2^{l+1} \Gamma(m+1)}{\pi r_0^{2l+2} \Gamma(l+m+1)} \right]^{1/2} r^l L_m^l(2r^2/r_0^2) e^{-r^2/r_0^2}. \quad (3.99)$$

Figure 3.8 shows as an illustration the first few eigenmodes of a weakly guiding circular parabolic fiber.

By using the above material, we can easily find analytic expressions for the transverse electric modes of a circular parabolic fiber. They are (see Section 3.2.3) solutions of the scalar eigenvalue problem

$$\frac{1}{k_0^2} (\Delta_t + k_0^2 \epsilon) e_{\parallel}^{(lm)}(r, \varphi) = \epsilon_{lm} e_{\parallel}^{(lm)}(r, \varphi).$$

The field component  $e_{\parallel}^{(lm)}$ , that is parallel to the isolines of the dielectric profile, is the only non-vanishing electric field component of a TE mode. In the case of circular waveguide structures, it is the angular coordinate  $\varphi$ . Using cylindrical coordinates and again introducing the representation

$$e_{\parallel}^{(lm)}(r, \varphi) \propto \begin{cases} \cos(l\varphi) \\ \sin(l\varphi) \end{cases}$$

we obtain an ordinary differential equation

$$\left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{l^2 + 1}{r^2} - \frac{4r^2}{r_0^4} \right) e_{\parallel}^{(lm)} = -k_0^2 (\epsilon_{co} - \epsilon_{lm}) e_{\parallel}^{(lm)} \quad (3.100)$$

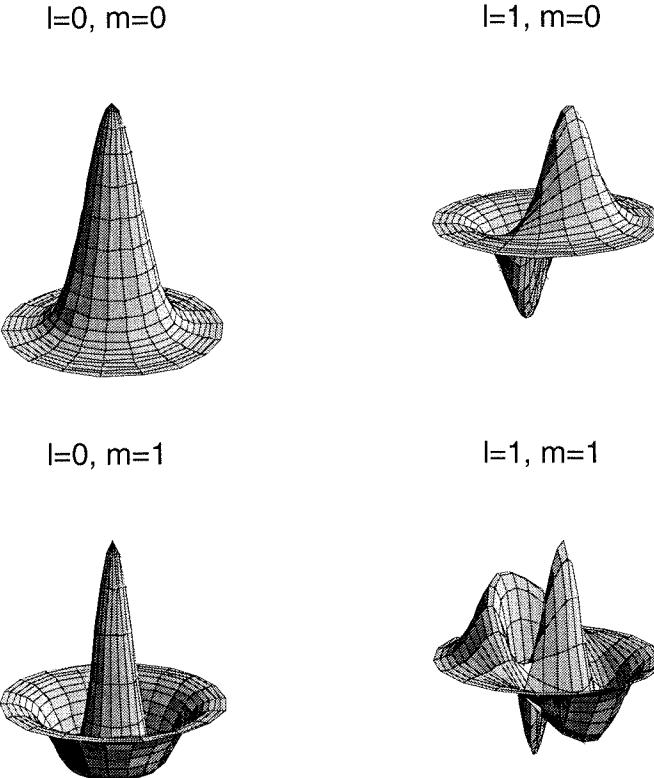


Figure 3.8. The lowest Laguerre-Gaussian eigenmodes  $\chi^{(lm)}$  of a weakly guiding circular parabolic fiber.

for the radial coordinate  $r$ . Using these results and  $\vartheta = \sqrt{l^2 + 1}$  we obtain

$$e_{\parallel}^{(lm)}(r, \varphi) = \chi^{(\vartheta m)}(r) \begin{cases} \cos(l\varphi) \\ \sin(l\varphi) \end{cases} \quad (3.101)$$

for the eigenmodes and

$$\epsilon_{lm} = \epsilon_{co} - \frac{8}{k_0^2 r_0^2} \left( m + \frac{\vartheta + 1}{2} \right) \quad (3.102)$$

for the eigenvalues of the TE modes of the circular parabolic fiber.

### 3.8 SLAB WAVEGUIDES

Dielectric slabs are the simplest optical waveguides. They are good vehicles to study the principal physical effects which form the basis for more complicated waveguide structures. In addition, many waveguide structures in integrated optics – especially the layered semiconductor waveguides – show a more or less slab-like modal behavior.

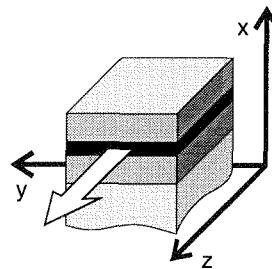


Figure 3.9. Layered medium.

The eigenvalue equations for slab waveguides are formulated using Cartesian coordinates. Here  $y$  and  $x$  are the transverse coordinates parallel and perpendicular to the waveguide layers respectively, and  $z$  is the longitudinal coordinate (see Figure 3.9). Since the slab waveguide remains unchanged by translations along the  $y$ -axis, i.e.,  $\epsilon = \epsilon(x)$ , the eigenmodes will also be invariant with respect to such translations. The eigenvalue Equations (3.11) and (3.12) then decouple and we obtain two equations

$$\frac{1}{k_0^2} \left( \frac{d^2}{dx^2} + k_0^2 \epsilon \right) e_y^{(m)} = \epsilon_m e_y \quad (3.103)$$

$$\frac{1}{k_0^2} \left( \frac{d^2}{dx^2} - \frac{d \ln \epsilon}{dx} \frac{d}{dx} + k_0^2 \epsilon \right) h_y^{(m)} = \epsilon_m h_y^{(m)} \quad (3.104)$$

for the field components parallel to the waveguide layers.

The remaining field components of both the electric and magnetic fields couple only to each other, i.e.,  $e_x \rightleftharpoons e_z$  and  $h_x \rightleftharpoons h_z$ . This is clear from a physical point of view since the direction of the electric field of a  $y$ -polarized plane electric wave remains unchanged if the wave is totally reflected at a layer-to-layer interface. The same argument holds, of course, for the magnetic field of a  $y$ -polarized magnetic wave. The  $x$ - and  $z$ -components are, however, related to the  $y$ -component of the electric field via

$$h_x^{(m)} = i\sqrt{\epsilon_m} e_y^{(m)}$$

$$h_z^{(m)} = -\frac{i}{k_0} \frac{\partial e_y^{(m)}}{\partial x} \quad (3.105)$$

and vice versa via

$$e_x^{(m)} = -\frac{i\sqrt{\epsilon_m}}{\epsilon} h_y^{(m)} \quad (3.106)$$

$$e_z^{(m)} = \frac{i}{k_0 \epsilon} \frac{\partial h_y^{(m)}}{\partial x}$$

We can conclude that *slab waveguides support only transverse electric (TE) and magnetic (TM) modes*.

We will now discuss a quasi-analytic solution of the eigenvalue problem for a multilayer waveguide. This approach makes use of the characteristic matrices of the layer stack (see Section 2.2.4), and the eigenvalues will be shown to be the roots of a transcendental equation. Figure 3.10 shows a multilayer stack located between  $x_- \leq x \leq x_+$  of the  $x$ -axis. It is covered by two different background materials with dielectric constants  $\epsilon_-$  and  $\epsilon_+$  corresponding to the lower and upper half-spaces respectively. We have divided the stack into lower and upper substacks, which are described by the characteristic matrices  $\mathcal{M}_-$  and  $\mathcal{M}_+$  respectively, the matrix elements of the characteristic matrices being  $m_{ij}^{(-)}$  and  $m_{ij}^{(+)}$ . Later on, we will match the amplitudes of the tangential field components at the interface between the two substacks.

Let us briefly recall our discussion of characteristic matrices in Section 2.2 and adapt that notation to our problem. The amplitudes of the tangential field components are now  $e_y$  and  $h_z$  for TE polarization and  $h_y$  and  $e_z$  for TM polarization; i.e., TE modes are s-polarized and TM modes p-polarized.

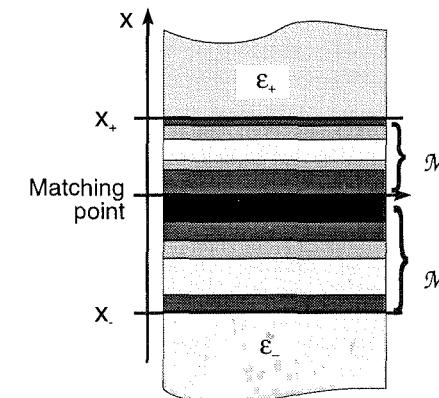


Figure 3.10. Multilayer waveguide.

We will start by calculating of the guided TE modes. An arbitrary layer of thickness  $d$  and dielectric constant  $\epsilon$  which supports a guided mode  $m$  has the equivalent phase thickness  $\delta = k_0 d \sqrt{\epsilon - \epsilon_m}$ , and the characteristic admittance of this layer is  $\eta = \sqrt{\epsilon - \epsilon_m}$  since guided modes can always be normalized. The optical field in the homogeneous regions outside layer stack ( $\psi_-$  for  $x < x_-$  and  $\psi_+$  and  $x > x_+$ ) must vanish at infinity ( $x = \pm\infty$ ); i.e., we obtain the solutions

$$\psi_{\pm} = \exp(\mp k_0 \sqrt{\epsilon_m - \epsilon_{\pm}} (x - x_{\pm}))$$

where  $\psi_+$  stands for both the tangential components of the electric and magnetic fields in the upper half-space and  $\psi_-$  for those in the lower half-space. The magnetic fields in the background regions are related to the electric fields via<sup>14</sup>

$$h_{\pm} = \mp \eta_{\pm} e_{\pm}.$$

The amplitudes of the tangential electric and magnetic fields at the interface can now be calculated starting from the top or bottom side of the multilayer stack. The characteristic matrix of each single layer is given by a matrix of the form

$$\mathcal{M} = \begin{pmatrix} \cos \delta & -\sin \delta / \eta \\ \eta \sin \delta & \cos \delta \end{pmatrix} = \begin{pmatrix} \cosh(i\delta) & -\sinh(i\delta)/(i\eta) \\ -i\eta \sinh(i\delta) & \cosh(i\delta) \end{pmatrix} \quad (3.107)$$

If we choose a unit electric field at the interface, the following four equations express the matching conditions for the remaining three amplitudes  $e_+$ ,  $e_-$  and  $H$ .

$$(\mathcal{M}_-)^{-1} \begin{pmatrix} e_- \\ i\eta_- e_- \\ iH \end{pmatrix} = \begin{pmatrix} 1 \\ e_+ \\ -i\eta_+ e_+ \end{pmatrix} = \mathcal{M}_+ \begin{pmatrix} e_+ \\ -i\eta_+ e_+ \end{pmatrix} \quad (3.108)$$

The elimination of the amplitudes  $e_+$ ,  $e_-$  and  $H$  is straightforward. As the result of this procedure, we obtain one characteristic equation

$$\frac{m_{11}^{(+)} - i\eta_+ m_{12}^{(+)}}{m_{21}^{(+)} - i\eta_+ m_{22}^{(+)}} = -\frac{m_{22}^{(-)} - i\eta_- m_{12}^{(-)}}{m_{21}^{(-)} - i\eta_- m_{11}^{(-)}}, \quad (3.109)$$

whose roots represent the guided TE modes. For the eigenmodes of a lossless layer stack, the characteristic admittances of the background regions are purely imaginary and the characteristic matrices are real. Thus, the roots of the nonlinear equation are real and must be located in the (one-dimensional) interval

$$\max(\epsilon_+, \epsilon_-) \leq \epsilon_m < \max \epsilon(x).$$

For lossy multilayer structures the roots must be found in the complex  $\mathbb{C}$ , and the preceding relation must be applied to the real parts of the eigenvalue and the

14. Subsequently, we always have to choose the signs such that  $\sqrt{\epsilon_m - \epsilon_{\pm}} = +i\eta$ .

dielectric profile. The amplitude  $e_y^{(m)}$  of the electric field inside an arbitrary layer ( $x_0 \leq x < x_0 + d$ ) within the layer stack is given by

$$e_y^{(m)}(x) = e_{y,0} \cos\left(\frac{\delta}{d}(x - x_0)\right) + \frac{h_{z,0}}{\eta} \sin\left(\frac{\delta}{d}(x - x_0)\right). \quad (3.110)$$

For a lossless single-layer slab with a dielectric constant which is embedded in background material having a dielectric constant  $\epsilon_L$ , the expressions simplify significantly. Using  $\eta_+ = \eta_- = \eta_B = -i\eta_m F_m$  with

$$F_m = i\sqrt{\frac{\epsilon_m - \epsilon_B}{\epsilon_L - \epsilon_m}}$$

and keeping in mind that  $\mathcal{M}_- = \mathcal{E}$  and

$$\mathcal{M}_+ = \begin{pmatrix} \cos \delta_m & -\sin \delta_m / \eta_m \\ \eta_m \sin \delta_m & \cos \delta_m \end{pmatrix}$$

we obtain the characteristic equation

$$\tan \delta_m = \frac{2F_m}{1 - F_m^2}.$$

We will recover this result within the framework of a geometrical optical treatment (see Section 3.10). A more detailed analysis of the single-layer slab will be performed in this context.

For calculating of the guided TM modes, we use the alternative characteristic matrices

$$\tilde{\mathcal{M}} = \begin{pmatrix} \cos \delta & -\sin \delta / \zeta \\ \zeta \sin \delta & \cos \delta \end{pmatrix} = \begin{pmatrix} \cosh(i\delta) & -\sinh(i\delta)/(i\zeta) \\ -i\zeta \sinh(i\delta) & \cosh(i\delta) \end{pmatrix} \quad (3.111)$$

which are based on the characteristic impedances  $\zeta = \sqrt{\epsilon - \epsilon_m}/\epsilon$  of the waveguide layers. This is because we intend to formulate the problem in terms of the magnetic field component  $h_y$ . The rest of the procedure runs parallel to the TE case. As a final result, we again obtain the characteristic equation

$$\frac{\tilde{m}_{11}^{(+)} - i\zeta_+ \tilde{m}_{12}^{(+)}}{\tilde{m}_{21}^{(+)} - i\zeta_+ \tilde{m}_{22}^{(+)}} = -\frac{\tilde{m}_{22}^{(-)} - i\zeta_- \tilde{m}_{12}^{(-)}}{\tilde{m}_{21}^{(-)} - i\zeta_- \tilde{m}_{11}^{(-)}}. \quad (3.112)$$

The guided TM modes are found as the roots of this equation.

### 3.9 SEPARABLE EIGENVALUE PROBLEMS

In this section we will discuss eigenvalue problems whose Hamiltonians  $\mathcal{H}$  can be expressed as a sum

$$\mathcal{H} = \mathcal{H}_m + \mathcal{H}_n - \bar{\epsilon}\mathcal{E} \quad (3.113)$$

of two Hamiltonians  $\mathcal{H}_m$  and  $\mathcal{H}_n$  which operate on different coordinate subspaces<sup>15</sup> of the full Hamiltonian  $\mathcal{H}$ . The third term  $\bar{\epsilon}\mathcal{E}$  introduces a reference dielectric constant in order to normalize the reduced eigenvalue problems of  $\mathcal{H}_m$  and  $\mathcal{H}_n$ . We will show first that irrespective of the explicit form of  $\mathcal{H}$  the eigenmodes of the Hamiltonian, i.e., the solutions of the eigenvalue problem

$$\mathcal{H} |mn\rangle = \epsilon_{mn} |mn\rangle, \quad (3.114)$$

are given by the direct product

$$|mn\rangle = |m\rangle |n\rangle \quad (3.115)$$

of the eigenmodes  $|m\rangle$  and  $|n\rangle$  which in turn belong to the two reduced eigenvalue problems

$$\mathcal{H}_m |m\rangle = \epsilon_m |m\rangle \quad (3.116)$$

and

$$\mathcal{H}_n |n\rangle = \epsilon_n |n\rangle. \quad (3.117)$$

The effective dielectric constant of the Hamiltonian  $\mathcal{H}$  is related to the eigenvalues of the reduced eigenvalue problems via

$$\epsilon_{mn} = \epsilon_m + \epsilon_n - \bar{\epsilon}. \quad (3.118)$$

To prove this theorem, we form the scalar product of the full eigenvalue problem (3.114) with an arbitrary adjoint field  $\langle\mu\nu|$ . If the scalar products  $\langle\mu|m\rangle \neq 0$  and  $\langle\nu|n\rangle \neq 0$  do not vanish we obtain

$$\frac{\langle\nu|\mathcal{H}_n n\rangle}{\langle\nu|n\rangle} = \epsilon_{mn} + \bar{\epsilon} - \frac{\langle\mu|\mathcal{H}_m m\rangle}{\langle\mu|m\rangle} \quad (3.119)$$

Obviously, the left-hand side of Equation (3.119) is independent of the coordinate operating in the subspace of the reduced Hamiltonian  $\mathcal{H}_m$ . Thus, the right-hand side of Equation (3.119) must also be independent of this coordinate. Since the eigenvalue is  $\epsilon_{mn}$  is a constant, the right-hand side is independent of the coordinate operating in the subspace of the reduced Hamiltonian  $\mathcal{H}_n$ . We can therefore conclude that both sides of Equation (3.119) must be constant and that the corresponding constant  $\epsilon_n$  must be enumerated only by the quantum number  $n$ . Since Equation (3.119) holds for any adjoint mode  $\langle\nu|$ , we can now cut off the first reduced eigenvalue problem

$$\mathcal{H}_n |n\rangle = \epsilon_n |n\rangle.$$

We still have to solve the remaining equation

$$\epsilon_{mn} + \bar{\epsilon} - \epsilon_n = \frac{\langle\mu|\mathcal{H}_m m\rangle}{\langle\mu|m\rangle}.$$

15. In quantum mechanics, for example, the coordinate subspaces are formed by the coordinates of different particles.

Obviously, the left-hand side of this equation (and hence the right-hand side too) is constant. Since the right-hand side is independent of the quantum number  $n$ , the corresponding constant  $\epsilon_m$  must be enumerated using the quantum number  $m$ . Since the remaining equation again holds for any adjoint mode  $\langle\mu|$ , we obtain the second reduced eigenvalue problem

$$\mathcal{H}_m |m\rangle = \epsilon_m |m\rangle$$

and we know that the eigenvalue of the full Hamiltonian is related to the reduced eigenvalues via Equation (3.118). So far, we have not made any assumptions about the explicit representation of the Hamiltonian  $\mathcal{H}$ .

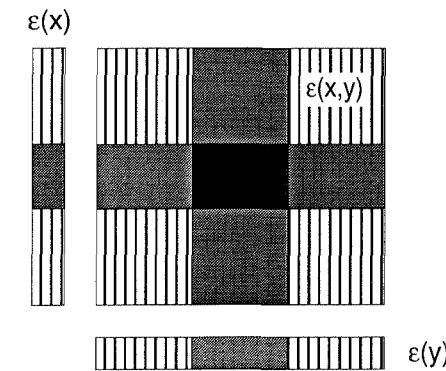


Figure 3.11. Separable step-index profile.

### 3.9.1 Marcatili's Approach

We will now apply these results to a weakly guiding waveguide whose dielectric profile is given by

$$\epsilon(x, y) = \bar{\epsilon} + \Delta\epsilon(x) + \Delta\epsilon(y) \quad (3.120)$$

where  $x$  and  $y$  are the Cartesian coordinates and  $\bar{\epsilon}$  acts as a background dielectric constant. Figure 3.11 shows a step-index profile as an illustrative example. If we either assume that the waveguide is nearly uniform or intend to calculate the TE modes of a weakly guiding slab-like structure, we obtain the following Hamiltonian (see Section 3.6)

$$\mathcal{H} = \mathcal{H}_x + \mathcal{H}_y - \bar{\epsilon}\mathcal{E}$$

with

$$\mathcal{H}_x = \frac{1}{k_0^2} \left\{ \frac{d^2}{dx^2} + k_0^2 [\bar{\epsilon} + \Delta\epsilon(x)] \right\}$$

$$\mathcal{H}_y = \frac{1}{k_0^2} \left\{ \frac{d^2}{dy^2} + k_0^2 [\bar{\epsilon} + \Delta\epsilon(y)] \right\}. \quad (3.121)$$

Each of the two operators  $\mathcal{H}_x$  and  $\mathcal{H}_y$  can be understood as a Hamiltonian describing the TE modes of a slab waveguide with an arbitrary dielectric profile. If we intend to calculate the TM modes of a weakly guiding slab-like structure we have to use

$$\mathcal{H}_x = \frac{1}{k_0^2} \left\{ \frac{d^2}{dx^2} - \frac{d \ln \epsilon}{dx} \frac{d}{dx} + k_0^2 [\bar{\epsilon} + \Delta\epsilon(x)] \right\} \quad (3.122)$$

and  $\mathcal{H}_y$  as given previously. The effective dielectric constants are in both cases given by the Equation (3.118).

Equation (3.120) applies only to few waveguide structures. However, dielectric profiles of that type are often used as a starting point for a perturbation analysis. Zeroth order approximations of this type are called *Marcatili approximations*, although his work (see [2]) aimed at the calculation of hybrid modes for the special waveguide structure shown in Figure 3.11. It should be noted that Marcatili's approach does not work close to cutoff since the optical field then spreads increasingly into the striped regions (see Figure 3.11) which are usually not well described by the dielectric profile (3.120).

### 3.9.2 The Effective Index Method

The effective index method provides an approximate solution of the eigenvalue problem for a weakly guiding waveguide by replacing the dielectric profile  $\epsilon(x, y)$  with a separable profile. Thanks to its accuracy and efficiency, this approximation scheme has become one of the workhorses for the design and modeling of many integrated optical structures.

As in the last section, we will start our discussion with the eigenvalue problem

$$\frac{1}{k_0^2} \left[ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_0^2 \epsilon(x, y) \right] \phi^{(mn)}(x, y) = \epsilon_{mn} \phi^{(mn)}(x, y) \quad (3.123)$$

which applies to the TE modes of a slab-like waveguide structure or to any eigenmodes of a nearly uniform dielectric profile. The two quantum numbers  $m$  and  $n$  again correspond to the number of zeros of the eigenmodes along the  $x$  and  $y$  axes of the coordinate system. Using the representation

$$\phi^{(mn)}(x, y) = \chi^{(m)}(x, y) \eta^{(mn)}(y) \quad (3.124)$$

we obtain the new eigenvalue problem

$$\begin{aligned} & \frac{1}{k_0^2} \left[ \frac{1}{\chi^{(m)}} \frac{\partial^2 \chi^{(m)}}{\partial x^2} + \frac{1}{\eta^{(mn)}} \frac{d^2 \eta^{(mn)}}{dy^2} \right. \\ & \left. + \frac{2}{\chi^{(m)} \eta^{(mn)}} \frac{\partial \chi^{(m)}}{\partial x} \frac{d \eta^{(mn)}}{dy} + \frac{1}{\chi^{(m)}} \frac{\partial^2 \chi^{(m)}}{\partial y^2} \right] = \epsilon_{mn} - \epsilon(x, y). \end{aligned} \quad (3.125)$$

We now choose the factor  $\eta^{(mn)}(y)$  such that the other factor  $\chi^{(m)}(x, y)$  depends only little on the lateral coordinate. The last two terms on the left-hand side of the new eigenvalue problem can then be treated as perturbations within the framework of the Rayleigh-Schrödinger perturbation theory. The unperturbed eigenvalue problem

$$\frac{1}{k_0^2} \left[ \frac{1}{\chi_0^{(m)}} \frac{\partial^2 \chi_0^{(m)}}{\partial x^2} + \epsilon(x, y) \right] = \epsilon_{mn} - \frac{1}{k_0^2 \eta_0^{(mn)}} \frac{d^2 \eta_0^{(mn)}}{dy^2} \quad (3.126)$$

now becomes separable and we easily obtain the two reduced eigenvalue problems

$$\frac{1}{k_0^2} \left[ \frac{\partial^2}{\partial x^2} + k_0^2 \epsilon(x, y) \right] \chi_0^{(m)}(x, y) = \epsilon_m(y) \chi_0^{(m)}(x, y) \quad (3.127)$$

and

$$\frac{1}{k_0^2} \left[ \frac{d^2}{dy^2} + k_0^2 \epsilon_m(y) \right] \eta_0^{(mn)}(y) = \epsilon_{mn} \eta_0^{(mn)}(y) \quad (3.128)$$

If we apply this procedure to the Hamiltonian describing the TM modes of a slab-like waveguide, we have to replace the first reduced eigenvalue problem by

$$\frac{1}{k_0^2} \left[ \frac{\partial^2}{\partial x^2} - \frac{\partial \ln \epsilon(x, y)}{\partial x} \frac{\partial}{\partial x} + k_0^2 \epsilon(x, y) \right] \chi_0^{(m)}(x, y) = \epsilon_m(y) \chi_0^{(m)}(x, y). \quad (3.129)$$

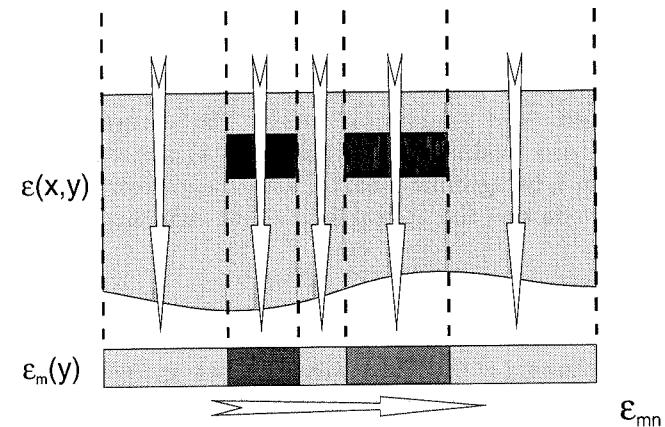


Figure 3.12. Effective index method (EIM).

The zeroth order approximation of the full eigenvalue problem, i.e., the solution of Equation (3.126), is referred to as the *effective index method* (EIM). Figure 3.12 illustrates the solution process for a step-index profile. The process is started by

projecting the dielectric profile onto the  $y$ -axis. The projected dielectric constant at the position  $y = y_0$ , for example, is the effective dielectric constant of a slab waveguide consisting of the waveguide layers  $\epsilon(x, y_0)$  at this location. Depending on the underlying waveguide problem, the eigenvalues<sup>16</sup> are calculated from Equations (3.127) or (3.129). After we have built up the projected dielectric profile in this way, we regard it as a slab waveguide and obtain finally the approximated eigenvalues as solutions of Equation (3.127). The arrows in Figure 3.12 indicate both projection processes.

As a consequence of the projection process, the eigenvalues obtained using the effective index method are independent of the position of the dielectric columns relative to each other. This approximation can yield large errors if we use the effective index method to calculate the coupling coefficients, i.e., the difference between the effective dielectric constants of directional couplers or any other distributed waveguide structure.

Using the first corrections

$$\epsilon_{mn}^{(1)} = \langle mn | (\mathcal{H} - \mathcal{H}_0) mn \rangle$$

of the effective dielectric constants, we can easily check the accuracy of the approximated eigenvalues. This correction should be significantly smaller than the difference between the effective dielectric constant obtained using the effective index method and the background dielectric constant  $\bar{\epsilon}$ .

### 3.10 GEOMETRICAL OPTICS OF THE WAVEGUIDE

*Light is trapped in optical waveguides by total internal reflection.* To clarify this key phrase, in this paragraph we will study the TE modes supported by the symmetric single-layer slab by means of geometrical optics. The waveguide is shown in Figure 3.13. It consists of a guiding layer of thickness  $d_L$  and dielectric constant  $\epsilon_L$ , which is embedded in a background medium with the dielectric constant  $\epsilon_B$ .

Inside the waveguide slab the optical field of the eigenmode  $m$  with the eigenvalue  $\epsilon_m$

$$\psi_m = \frac{1}{2} \left\{ \exp \left( ik_0 [\sqrt{\epsilon_m} z + \sqrt{\epsilon_L - \epsilon_m} x] \right) - \exp \left( ik_0 [\sqrt{\epsilon_m} z - \sqrt{\epsilon_L - \epsilon_m} x] \right) \right\}$$

may be regarded as the sum of two plane waves, known as *Brillouin waves*. Clearly, these waves are in phase only if the phase shift<sup>17</sup> between the direct path  $\overline{IT}$  and the zigzag path  $\overline{IR}$ , which includes reflections at the lower and the upper boundary of the waveguide layer, becomes a multiple of  $2\pi$ . Figure 3.13 shows the Brillouin

16. If the corresponding slab waveguide has no guided modes, the effective dielectric constant is replaced by the value at cutoff.

17. This condition is the integrated optical equivalent to the Bohr-Sommerfeld quantization rules in quasi-classical quantum theory.

wave of the eigenmode  $m$  propagating through the waveguide, where the angle  $\theta_m$  is given by

$$\cos \theta_m = \sqrt{\frac{\epsilon_m}{\epsilon_L}}.$$

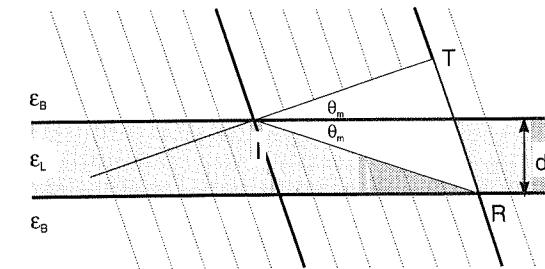


Figure 3.13. Symmetric slab waveguide and the Brillouin waves of the eigenmode  $m$ .

The phase shift caused by the geometrical path difference is given by

$$k_0 (\overline{IR} - \overline{IT}) = k_0 \overline{IR} (1 - \cos(2\theta_m)) = 2\delta_m,$$

where

$$\delta_m = k_0 d_L \sqrt{\epsilon_L - \epsilon_m} \quad (3.130)$$

is the equivalent phase thickness of the waveguide layer supporting the eigenmode  $m$ . For total reflection of a TE-polarized wave (see Equation (2.53)), we obtain an additional phase shift<sup>18</sup>

$$\phi_m = -2 \tan^{-1} F_m \quad (3.131)$$

with the abbreviation

$$F_m = \sqrt{\frac{V^2 - \delta_m^2}{\delta_m^2}} \quad (3.132)$$

and the  $V$  parameter

$$V = k_0 d_L \sqrt{\epsilon_L - \epsilon_B}. \quad (3.133)$$

18. For TM polarization we must only replace  $F_m$  by

$$F_m = \frac{\epsilon_L}{\epsilon_B} \sqrt{\frac{V^2 - \delta_m^2}{\delta_m^2}},$$

since the additional phase shift  $\pi$  for TM polarization does not influence the phase shift for two reflections.

We can conclude from Equation (3.132) that the eigenmode  $m$  of the symmetric slab waveguide is guided, i.e., trapped by total internal reflection, only if the condition

$$\delta_m \geq V \quad (3.134)$$

is satisfied. The cutoff of the eigenmode  $m$  is defined by  $\delta_m = V$ , the eigenvalue at cutoff is  $\epsilon_m = \epsilon_B$ . Unfortunately, there are no such simple rules for arbitrarily shaped waveguides.

Using the results obtained previously, we can now formulate the phase relation for the Brillouin waves of the eigenmode  $m$ . It is

$$\phi_m + \delta_m = \pi m.$$

By taking the tangent of this equation, we arrive at the usual form of the eigenmode equation

$$\tan \delta_m = \frac{2F_m}{1 - F_m^2}. \quad (3.135)$$

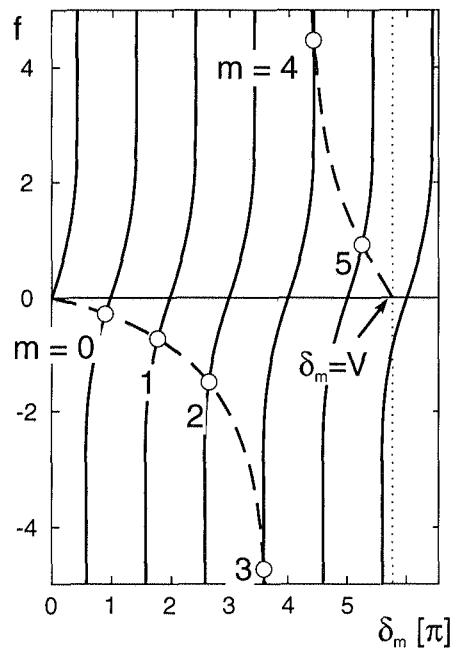


Figure 3.14. Graphical determination of eigenmodes for the symmetric slab (solid –  $\tan \delta_m$ , dashed –  $2F_m/(1 - F_m^2)$ ).

Figure 3.14 shows the graphical solution of Equation (3.135). The solid lines represent the branches of  $\tan \delta_m$ , the dashed lines represent the right-hand side of

Equation (3.135), and the solutions are marked by open circles. The dashed line must end at  $\delta_m = V$  since the right-hand side of Equation (3.135) becomes purely imaginary at this value. The eigenvalues  $\epsilon_m$  can be obtained using

$$\epsilon_m = \epsilon_L - \left( \frac{\delta_m}{k_0 d_L} \right)^2.$$

At the cutoff of the  $m$ th eigenmode ( $F_m = 0$ ) the equivalent phase thickness is  $\delta_m = m\pi$ . Thus, the symmetric single-layer slab waveguide supports  $m+1$  guided modes if the  $V$  parameter satisfies the condition

$$V \geq m\pi. \quad (3.136)$$

It supports at least one guided mode. We can further conclude from Figure 3.14 that the position of the  $m$ th eigenmode is restricted to the interval

$$m\pi \leq \delta_m < (m+1)\pi.$$

For eigenmodes that are far away from cutoff ( $V \gg \delta_m$ ; i.e.,  $F_m \approx V/\delta_m \gg 1$ ), we can approximate the solutions of (3.135) by  $\delta_m \approx V/(V+2)(m+1)\pi$ . By inserting this result into Equation (3.130) we obtain

$$\epsilon_m = \epsilon_L - \left( \frac{(m+1)\pi}{k_0 d_L} \right)^2 \left( 1 - \frac{4}{V} + \dots \right). \quad (3.137)$$

The leading term  $\epsilon_m = \epsilon_L - \alpha \lambda^2/d^2$  exhibits the typical shape of the dispersion curves, i.e., hyperbolic behavior with respect to the geometry and polynomial behavior with respect to the wavelength.

## REFERENCES

- [1] Snyder, A. W., and Love, J. D., *Optical Waveguide Theory*, London, Chapman and Hall, 1983.
- [2] Marcuse, D., *Theory of Dielectric Optical Waveguides*, New York, Academic Press, 1974.
- [3] Collin, R. E., *Field Theory of Guided Waves*, 2 ed., New York, IEEE Press, 1991.
- [4] Kapany, N. S., and Burke, J. J. *Optical Waveguides*, New York, Academic Press, 1972.
- [5] Marcuse, D., *Light Transmission Optics*, New York, Van Nostrand, 1972.
- [6] Lewin, L., *Theory of Waveguides*, London, Newes-Butterworth, 1975.
- [7] Sodha, M. S. and Ghatak, A. K., *Theory of Waveguides*, New York, Plenum Press, 1977.
- [8] Unger, H.-G., *Planar Optical Waveguides and Fibres*, Oxford, Clarendon Press, 1977.
- [9] Hocker, G. B. and Burns, W. K., "Mode Dispersion in Diffused Channel Waveguides by the Effective Index Method", *Applied Optics*, Vol. 16, 1978, pp. 113–118.
- [10] Chilwell, J., and Hodgekinson, I., "Thin-Films Field-Transfer Matrix Theory of Planar Multilayer Waveguides and Reflection from Prism-Loaded Waveguides", *Journal of the Optical Society of America*, Vol. A1, 1984, pp. 742–753.
- [11] Rogge, U., and Pregla, R., "Method of Lines for the Analysis of Dielectric Waveguides", *IEEE Journal of Lightwave Technology*, Vol. 11, 1993, pp. 2015–2020.

## Chapter 4

### Beam Propagation

The spread of the optical field in a half-space is one of the canonical problems of classical optics. The first quantitative results were given by Fraunhofer and later on by Fresnel. Their theories were aimed at calculating the diffraction patterns of various apertures. At the beginning of the twentieth century, Sommerfeld worked out a rigorous theory of diffraction by a metallic half-plane. The beam propagation in a half-space with (nearly) arbitrary inhomogeneities – as in a typical integrated optical chip – can now be simulated numerically. The underlying mathematical procedure is called the *beam propagation method* (BPM), although it covers a couple of algorithms. This type of field propagation was developed in underwater acoustics and seismology before it was adapted to optical problems by Feit and Fleck. The rise of III-V waveguide components which contain high-contrast step-index waveguides led to the application of new algorithms to the BPM problem in order to enhance both the stability and performance of the method. Further developments aimed at the inclusion of reflections and of vectorial waves.

In this chapter we will derive the forward Helmholtz equation which governs the beam propagation. To fit the theory given here into the framework of classical diffraction theory, we will show that the paraxial Helmholtz equation, which is equivalent to Fresnel's diffraction integral is the first term of an expansion of the forward Helmholtz equation. We will study the classical BPM and competitive algorithms based on finite differences (FD-BPM), on a  $z$ -transient variational principle (FE-BPM) and on the eigenmode analysis in comparison. The chapter ends with a short sketch of the vectorial BPM and of treatment of reflections in the framework of the BPM.

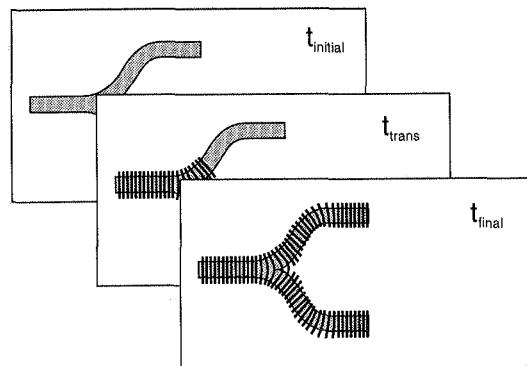
Before we start elaborating the theory, we will deal with the philosophy behind the beam propagation method and its application to the design and modeling of integrated optical components. In the previous chapters we have studied the eigenmode of arbitrary waveguide structures. In most cases, the modal properties allow for the construction of devices such as directional couplers, Bragg and diffraction gratings, whereas only a few components such as optical tapers cannot be con-

structed. In contrast to the eigenmode analysis, the beam propagation method is not a constructive method but a computer simulation. As in the experiment, the BPM yields the response of a given device to an external optical signal. Naturally, an analysis of the results can be used to obtain information about the modal properties of that device. For these reasons, the beam propagation method is well suited to study the following applications:

- layout of devices that are not described by a quantitative theory (e.g., tapers or directional couplers with tapered fan-in),
- influence of technological problems on the operation of components,
- problems with theories and approximations,
- computer experiments to gain some ideas about the behavior of a special device.

## 4.1 FUNDAMENTAL PRINCIPLES

At the beginning of our discussion we would like to get some idea of beam propagation. We shall consider a simple Y-branch as an illustration. As in the previous chapters, the coordinate system is chosen such that the beam propagates mainly in the  $+z$ -direction. The top of Figure 4.1 shows the device at the moment  $t_{\text{initial}}$  immediately before the optical signal arrives at the device. At the time  $t_{\text{trans}}$  (middle of Figure 4.1) the wavefront of the optical signal – i.e., the boundary between light and shadow – is located inside the device. After  $t_{\text{final}}$ , the wavefront has passed the device, and the bottom Figure 4.1 shows the final (time-harmonic) optical field inside the device. The calculation of the optical field according to this procedure –

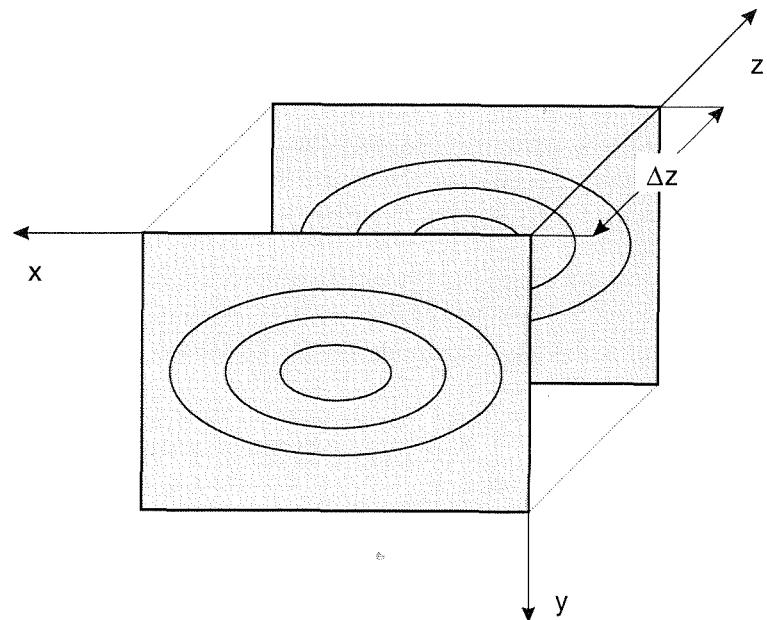


**Figure 4.1.** Three snapshots during the propagation of an optical signal through a Y-branch, before entrance, wave front within the device, after exit.

mathematically speaking the solution of the (hyperbolic) wave equation – is always successful, and for certain problems (e.g., wavefront analysis of nonlinear devices) it is the only way to obtain any results. For most applications in integrated optics, however, only the final state shown in Figure 4.1 is required. The starting oscillations shown in the top and middle of Figure 4.1 are of minor interest. It is then much more efficient to obtain the time-harmonic solution directly. The Helmholtz equations which were derived in Chapter 2 of this book are an appropriate starting point for such an intention. In contrast to the (hyperbolic) wave equation, however, the (elliptic) Helmholtz equation has no well-defined initial value problem. This lack is due to the simultaneous existence of waves propagating forward and backward. The basic idea of the beam propagation method is consequently to consider only the waves traveling forward.

### 4.1.1 Forward Helmholtz Equations

All BPM algorithms described in this chapter run stepwise; i.e., they transport the optical field within one propagation step from the transverse plane at the longitudinal coordinate  $z$  to the transverse plane at  $z + \Delta z$  (see Figure 4.2). We assume



**Figure 4.2.** Operating scheme of BPM algorithms.

furthermore that the dielectric profile remains unchanged during one propagation step; i.e., the dielectric profile

$$\epsilon(\mathbf{r}_t) = \bar{\epsilon} + \Delta\epsilon(\mathbf{r}_t) \quad (4.1)$$

depends only on the transverse coordinates which are designated by the position vector  $\mathbf{r}_t$ . The reference dielectric constant  $\bar{\epsilon}$  introduced in Equation (4.1) will be discussed later on. Under these (very general) conditions we can formulate both vectorial and scalar Helmholtz equations as

$$\frac{\partial^2}{\partial z^2} |\Psi\rangle = -k_0^2 \mathcal{H} |\Psi\rangle \quad (4.2)$$

in which the Hamiltonian  $\mathcal{H}$  with the eigenvalue problem

$$\mathcal{H} |m\rangle = \epsilon_m |m\rangle \quad (4.3)$$

acts only on the transverse coordinates.

The explicit expressions for the vector E-field formulation are  $|\Psi\rangle = |\mathbf{E}_t\rangle$  and

$$\mathcal{H}_{\mathbf{E}} = \frac{1}{k_0^2} [\Delta_t + k_0^2 \epsilon + \nabla_t \cdot \nabla_t (\ln \epsilon)], \quad (4.4)$$

and  $|\Psi\rangle = |\mathbf{H}_t\rangle$  and

$$\mathcal{H}_{\mathbf{H}} = \frac{1}{k_0^2} [\Delta_t + k_0^2 \epsilon + \nabla_t (\ln \epsilon) \times \nabla_t \times], \quad (4.5)$$

for the vector H-field formulation. As pointed out previously, the scalar Helmholtz equations describing the propagation of the optical field in weakly guiding waveguide structures (see Section 3.6) also fit into the framework of Equation (4.2). We obtain  $|\Psi\rangle = |E_y\rangle$  and

$$\mathcal{H}_{\text{TE}} = \frac{1}{k_0^2} (\nabla_t^2 + k_0^2 \epsilon) \quad (4.6)$$

for the TE modes of a slab-like waveguide (or for all modes of a nearly uniform waveguide), and  $|\Psi\rangle = |H_y\rangle$  and

$$\mathcal{H}_{\text{TM}} = \frac{1}{k_0^2} \left( \nabla_t^2 + k_0^2 \epsilon - \frac{\partial(\ln \epsilon)}{\partial x} \frac{\partial}{\partial x} \right) \quad (4.7)$$

for its TM modes.

By taking the square root of Equation (4.2), we arrive at the forward and backward Helmholtz equations:

$$\frac{\partial}{\partial z} |\Psi_{\pm}\rangle = \pm i k_0 \sqrt{\mathcal{H}} |\Psi_{\pm}\rangle, \quad (4.8)$$

i.e., at two differential equations of first order with respect to the longitudinal coordinate  $z$ . We have thus derived two initial value problems which possess the formal solutions<sup>1</sup>

$$|\Psi_{\pm}(z + \Delta z)\rangle = \mathcal{U}_{\text{BPM}}^{(\pm)}(\Delta z) |\Psi_{\pm}(z)\rangle \quad (4.9)$$

with the BPM propagator  $\mathcal{U}_{\text{BPM}}^{(\pm)}(\Delta z)$  given by

$$\mathcal{U}_{\text{BPM}}^{(\pm)}(\Delta z) = \exp(\pm i k_0 \sqrt{\mathcal{H}} \Delta z). \quad (4.10)$$

To obtain a physical understanding of Equation (4.9), we shall consider the very primitive but illustrative example of propagation of an optical field that is constant in the transverse plane ( $\nabla_t \Psi = 0$ ) through a homogeneous medium having a dielectric constant  $\bar{\epsilon}$ . We obtain

$$\mathcal{U}_{\text{BPM}}^{(\pm)} |\Psi_{\pm}\rangle = \exp(\pm i \bar{q} \Delta z) |\Psi_{\pm}\rangle,$$

i.e., a plane wave traveling either forward (+) or backward (-) along the  $z$ -axis.

The general simulation problems which can be tackled using the BPM are assumed to be “similar” to the physical situation described in our example; i.e., we assume in the following that the reference dielectric constant  $\bar{\epsilon}$  can always be chosen such that<sup>2</sup>

$$\|\Delta\epsilon(\mathbf{r}_t) \Psi(\mathbf{r}_t, z)\|_{\max} \ll \|\bar{\epsilon} \Psi(\mathbf{r}_t, z)\|_{\max}.$$

The deviations of the dielectric profile  $\Delta\epsilon(\mathbf{r}_t)$  are small, at least where a significant amount of optical power is concentrated. We will show later on (see Section 4.1.3) how we can find an optimized reference dielectric constant  $\bar{\epsilon}$ . Under these conditions, the Hamiltonian can be written as

$$\mathcal{H} = \bar{\epsilon} [\mathcal{E} + \mathcal{P}(\mathbf{r}_t)] \quad (4.11)$$

in which the operator  $\mathcal{P}$  designates a “small” perturbation of the unity operator  $\mathcal{E}$ . The square root of the Hamiltonian

$$\sqrt{\mathcal{H}} = \sqrt{\bar{\epsilon} (\mathcal{E} + \mathcal{P})}$$

can then be expanded into powers of this perturbation operator  $\mathcal{P}$ . If, incidentally, we switch from the rapidly varying field components  $\Psi$  to their slowly varying amplitudes  $\psi$ , i.e.,

$$\Psi_{\pm}(\mathbf{r}_t, z) = \psi_{\pm}(\mathbf{r}_t, z) e^{\pm i \bar{q} z} \quad (4.12)$$

with the reference propagation constant  $\bar{q} = k_0 \sqrt{\bar{\epsilon}}$ , we obtain two differential equations for the propagating amplitudes

$$\frac{\partial}{\partial z} |\psi_{\pm}\rangle = \pm i \frac{\bar{q}}{2} \left[ \mathcal{P} + 2 \sum_{n=2}^{\infty} \binom{\frac{1}{2}}{n} \mathcal{P}^n \right] |\psi_{\pm}\rangle, \quad (4.13)$$

1. We will see in the next section that commutation rules for the operators have to be taken into account if real solutions are to be derived from Equation (4.8).
2. Here  $\|f(\mathbf{r}_t)\|_{\max} = \max |f(\mathbf{r}_t)|$  is the maximum norm on the transverse plane.

which are of first order with respect to the propagation coordinate  $z$  and of (formally) infinite order with respect to the transverse coordinates.

If we take only the leading terms of the expansion into account, we obtain the paraxial Helmholtz equations

$$\frac{\partial}{\partial z} |\psi_{\pm}\rangle = \pm i \frac{\bar{q}}{2} \mathcal{P} |\psi_{\pm}\rangle. \quad (4.14)$$

We will see later on (see Section 4.3) that the solution of the paraxial Helmholtz equation corresponds to the classical treatment of diffraction, i.e., to the calculation of Fresnel's diffraction integral. Approximations taking more contributions<sup>3</sup> of Equation (4.13) into account are called *wide-angle approximations*, since they improve the results for plane waves traveling at increasing angles to the  $z$ -axis.

In the following, we will restrict ourselves to forward propagation and suppress the indices  $\pm$  indicating the direction of propagation.

#### 4.1.2 Mathematical Assessment

The derivations of the last section were rather formal since we did not worry about the properties of the operators thus derived. We will now make up this lack. A comparison between beam propagation and quantum-mechanical propagation will help us to obtain a better physical understanding of beam propagation. If we compare the forward Helmholtz equation

$$\frac{\partial}{\partial z} |\Psi\rangle = ik_0 \sqrt{\mathcal{H}} |\Psi\rangle \quad (4.15)$$

with the time-dependent Schrödinger equation

$$\frac{\partial}{\partial t} |\Psi\rangle = \frac{i}{\hbar} \mathcal{H}_{QM} |\Psi\rangle \quad (4.16)$$

the similarity between both equations is obvious. We intuitively expected such a correspondence since the forward Helmholtz equation describes the propagation of the optical field with respect to the propagation direction  $z$  and the Schrödinger equation that of a quantum-mechanical state with respect to time  $t$ . However, there are some differences as quantum mechanical propagation is governed by the Hamiltonian  $\mathcal{H}_{QM}$  itself whereas beam propagation is determined by the square root of its Hamiltonian  $\sqrt{\mathcal{H}}$ . A short glance at the derivation of the forward Helmholtz equation in the previous section shows that we obtained the forward Helmholtz equation by restricting ourselves to the waves traveling either in the  $\pm z$ -directions. The time-dependent Schrödinger equation, in contrast, is restricted to the propagation into the "future", i.e., in the  $+t$ -direction.

3. We should note that Equation (4.13) can be replaced by alternative expansions which gain better results for (non-differentiable) step-index profiles.

We know from quantum mechanics that, *provided a linear operator  $\mathcal{H}$  is self-adjoint and the parameter  $t$  is a real number, the propagators*

$$\mathcal{U}(t) = e^{i\mathcal{H}t}$$

*form a unitary group.* We know furthermore that the scalar product is conserved by unitary transformations. In the following, we will analyze the BPM propagator  $\mathcal{U}_{BPM}$  in order to determine whether it is a unitary operator. For this purpose, we refer to the following theorem from the theory of linear operators: *The square root  $\sqrt{\mathcal{H}}$  of an operator is self-adjoint if and only if the operator itself is self-adjoint and non-negative.* Figure 4.3 shows the spectrum of an integrated optical Hamiltonian. It has positive and negative eigenvalues corresponding to the guided and radiation modes ( $\epsilon_m \geq 0$ ) and to the evanescent modes ( $\epsilon_m < 0$ ). Thus, the integrated optical Hamiltonians are *not* non-negative. Hence, their square roots,  $\sqrt{\mathcal{H}}$ , are *not* self-adjoint. Hence, the BPM-propagators  $\mathcal{U}_{BPM}$  are *not* unitary and last not least the optical power is not conserved during propagation. All these phenomena are a consequence of the existence of evanescent modes.

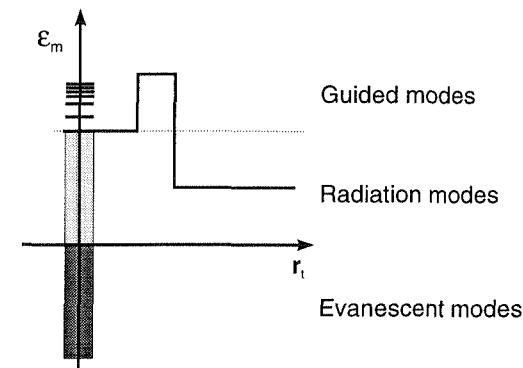


Figure 4.3. Spectrum of an integrated optical Hamiltonian.

It is interesting to note here that the propagator

$$\mathcal{U}_p(z) = e^{i\mathcal{P}\bar{q}z/2} \quad (4.17)$$

describing beam propagation in a paraxial approximation is – in contrast to the true propagator – a unitary operator. In physical terms, *paraxial approximations disregard the influence of evanescent modes.* We can further conclude that the evanescent modes will play a minor role in practical applications of the beam propagation method.

### 4.1.3 Optimized Paraxial Approximations

In Section 4.1.1 we have introduced the reference dielectric constant  $\bar{\epsilon}$ . We will now show how to minimize the error caused by the paraxial approximation by an appropriate choice of this free parameter. For this purpose, we consider the Hamiltonian

$$\mathcal{H}_p = \bar{\epsilon} \left( \mathcal{E} + \frac{\mathcal{P}}{2} \right)^2 \quad (4.18)$$

which leads directly (i.e., without any additional approximation) to the paraxial Helmholtz equation. We expect that the best paraxial approximation is achieved if the mean value

$$\langle \psi | (\mathcal{H} - \mathcal{H}_p) \psi \rangle \stackrel{!}{=} \min \quad (4.19)$$

of the difference between the full Hamiltonian and its paraxial approximation with respect to the initial optical field  $\psi$  is minimized. By calculating the expression on the left-hand side of this equation we obtain

$$\langle \psi | (\mathcal{H} - \mathcal{H}_p) \psi \rangle = \left\langle \psi \left| \frac{\bar{\epsilon}}{4} \mathcal{P} \psi \right. \right\rangle = \left\langle \psi \left| \frac{(\mathcal{H} - \bar{\epsilon})^2}{4\bar{\epsilon}} \psi \right. \right\rangle.$$

The best paraxial approximation must satisfy the equation

$$\frac{\partial}{\partial \bar{\epsilon}} \langle \psi | (\mathcal{H} - \mathcal{H}_p) \psi \rangle = \left\langle \psi \left| \frac{\mathcal{H}^2 - \bar{\epsilon}^2}{4\bar{\epsilon}^2} \psi \right. \right\rangle \stackrel{!}{=} 0.$$

The best reference dielectric constant according to the criterion (4.19) is then given by

$$\bar{\epsilon} = \sqrt{|\langle \psi | \mathcal{H}^2 \psi \rangle|}. \quad (4.20)$$

We see that the reference dielectric constant is closely related to the mean dielectric constant

$$\langle \epsilon \rangle = \langle \psi | \mathcal{H} \psi \rangle$$

of the input beam. In fact, both expressions coincide if the initial beam  $\psi$  is an eigenmode of the full Hamiltonian.<sup>4</sup>

We should note that the derivation given here is universal; i.e., it holds for any paraxial Helmholtz equation irrespective of previous approximations.

---

4. If we replace the criterion (4.19) by

$$\left\langle \psi \left| \sqrt{\mathcal{H} - \mathcal{H}_p} \psi \right. \right\rangle \stackrel{!}{=} \min$$

we arrive at the mean dielectric constant  $\langle \epsilon \rangle$  as a best choice for the reference dielectric constant  $\bar{\epsilon}$ .

### 4.1.4 Boundary Conditions

For the numerical elaboration of the BPM we have to restrict the propagating beam with respect to the transverse coordinates  $r_t$  by a window, i.e., by a finite area  $A$ , in the transverse coordinate plane. To obtain a well-defined boundary value problem, we must formulate additional conditions which have to be satisfied on the boundary  $\partial A$  of the window  $A$ . We will now discuss some options for this choice.

#### Metallic Boundaries

It is assumed that the window is surrounded by perfectly conducting material. We obtain the boundary conditions

$$E_{\parallel}|_{\partial A} = 0 \quad (4.21)$$

$$\frac{\partial E_{\perp}}{\partial n}|_{\partial A} = 0 \quad (4.22)$$

for the electric field and

$$H_{\perp}|_{\partial A} = 0 \quad (4.23)$$

$$\frac{\partial H_{\parallel}}{\partial n}|_{\partial A} = 0 \quad (4.24)$$

for the magnetic field. The two conditions relating to the normal derivative  $\frac{\partial}{\partial n}$  of the optical field are approximations.

If we apply these conditions to a rectangular window, we arrive at

$$E_y|_{x=\text{const}} = 0 \quad (4.25)$$

$$\frac{\partial E_y}{\partial y}|_{y=\text{const}} = 0 \quad (4.26)$$

for TE polarization and

$$\frac{\partial H_y}{\partial x}|_{x=\text{const}} = 0 \quad (4.27)$$

$$H_y|_{y=\text{const}} = 0 \quad (4.28)$$

for TM polarization. Figure 4.4 shows the boundary conditions for both polarizations. For TE polarization, we have to apply the Dirichlet condition  $-f|_{\partial A} = 0$  – as long as the boundary runs parallel to the waveguide layers ( $x = \text{const}$ ). If the boundary runs normal to the waveguide layers ( $y = \text{const}$ ) we must apply the Neumann condition  $\frac{\partial f}{\partial n}|_{\partial A} = 0$ . For TM polarization the situation is reversed.

It is typical for metallic boundaries that an optical field arriving at the boundary is reflected back into the window.

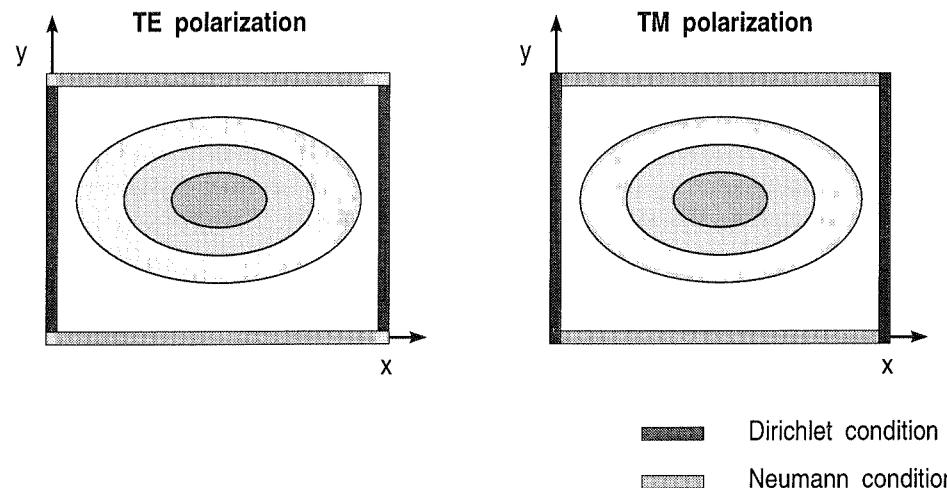


Figure 4.4. Metallic boundaries with TE and TM polarization.

### Periodic Boundaries

The window  $A$  is replaced by a compact manifold with the same surface. Thus, a 1D window is replaced by a circle  $S^1$ , a rectangular 2D window of size  $L_x L_y$  is substituted by a torus  $T^2 = S^1 \times S^1$ . Thus the boundary conditions are

$$\psi(x + mL_x, y) = \psi(x, y + nL_y) = \psi(x, y). \quad (4.29)$$

Obviously, for periodic boundaries all contributions leaving the window on one side return to the window on the opposite side. The most important application of periodic boundaries is the numerical Fourier transformation which is discussed in Section 4.2.

### Absorbing Walls

Absorbing walls are not boundary conditions. Roughly speaking, they are not the wall itself but the paper on it; i.e., they absorb the optical field in the region close to the boundary so that the field and all its derivatives are largely suppressed at the boundary. Absorbing boundaries can be combined with any types of boundary. If they absorb well it no longer matters which boundaries are found behind the absorber.

In addition, a good absorber must not reflect any field back to the window. To realize this property, a smooth increase of absorption must be ensured (see Figure 4.5). This in turn may cause some problems with power conservation in the window for typical BPM runs with several thousand steps. Several different elementary profiles – linear, quadratic or half Gaussian – are known for the absorber. Moreover, digital filters such as the tapered cosine window [1] are used as absorbers.

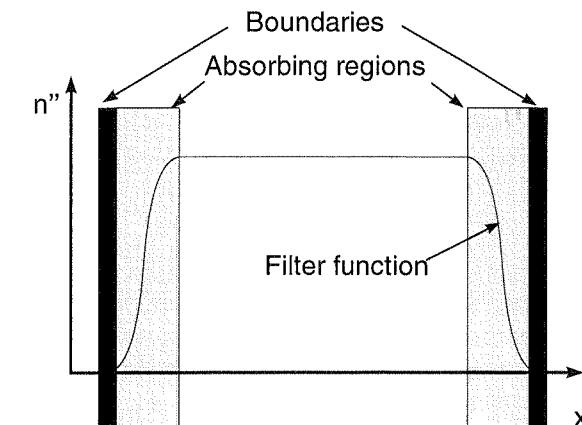


Figure 4.5. Absorbing walls.

The optimal shape of the absorber depends on the angle of illumination. Thus, the absorber itself must be optimized for critical applications.

From a physical point of view, the absorber can be regarded as a region filled with absorbing material. Indeed, the absorber can also be realized by adding absorbing material to the inhomogeneous refractive index profile.

### Transparent Boundaries

The optical field at the boundary is chosen such that the field inside a region close to the boundary represents a plane wave which leaves the computational window. The whole procedure is carried out in two steps. First the wave number of the plane wave is “measured”. In the next propagation step the result is used to calculate the optical field at the boundary. For a 1D window we use the equation

$$\psi(x) = e^{iq_x \Delta x} \psi(x - \Delta x). \quad (4.30)$$

at a point  $x = x_A$  which is not directly adjacent to the boundary to determine the propagation constant  $q_x$ . Using the same equation, we can now determine the optical field at the boundary  $x = x_B$  from the next adjacent point. The extension of this procedure to 2D windows is obvious as long as the entire grid is regular. For many practical problems, the application of transparent boundary conditions results in a good dereflection of the boundary of the computational window.

A rigorous analysis [2] shows that the transparent boundary condition can be formulated as an inhomogeneous Cauchy condition.

## 4.2 CLASSICAL BPM

The classical BPM solves the scalar forward Helmholtz equation (4.8), which holds for the TE modes of any weakly guiding waveguide, by Fourier transformation. This technique requires the Hamiltonian  $\mathcal{H}$  to be split into two parts, i.e.,

$$\mathcal{H} = \mathcal{D} + \mathcal{W}. \quad (4.31)$$

One of these is the diffraction operator  $\mathcal{D}$  which is diagonal in Fourier space. It contains the contributions which are responsible for the diffraction of the input beam. The other one, the operator  $\mathcal{W}$ , is diagonal in position space. It contains all contributions caused by the inhomogeneities of the dielectric profile. A short glance at equations (4.6) and (4.7) shows that this procedure can easily be carried out for TE polarization. For TM polarization, however, the term

$$\frac{\partial \ln \epsilon}{\partial x} \frac{\partial}{\partial x}$$

makes it difficult to proceed in the desired way. And indeed, no classical BPM algorithm is yet known for TM polarization. In consequence, we restrict ourselves for the rest of this section to TE polarization. We then obtain

$$\mathcal{D} = \nabla_t^2 + \bar{q}^2$$

and

$$\mathcal{W} = k_0^2 \Delta \epsilon(\mathbf{r}_t)$$

where  $\bar{q}$  designates again the reference propagation constant (see Section 4.1).

In other words, the *classical BPM* concentrates on the problem of diffraction and regards the inhomogeneities as a perturbation of free space propagation, i.e.,

$$\|k_0^2 \Delta \epsilon(\mathbf{r}_t) \Psi(\mathbf{r}_t, 0)\|_{\max} \ll \|(\nabla_t^2 + \bar{q}^2) \Psi(\mathbf{r}_t, 0)\|_{\max}.$$

The algorithm expands the operator  $\sqrt{\mathcal{H}}$ , as we will see later, into powers of the operator  $\mathcal{W}$ .

### 4.2.1 Diffraction

Before we derive the classical BPM propagator, we will first consider the diffraction of a beam into a homogeneous space, i.e.,  $\mathcal{W} = 0$ . As pointed out before, this is the application for which the classical BPM is designed. The forward Helmholtz equation can then be written as

$$\frac{\partial}{\partial z} |\Psi\rangle = i\sqrt{\nabla_t^2 + \bar{q}^2} |\Psi\rangle. \quad (4.32)$$

Using the relation

$$\sqrt{\nabla_t^2 + \bar{q}^2} = \frac{\nabla_t^2}{\bar{q} + \sqrt{\nabla_t^2 + \bar{q}^2}} + \bar{q}$$

we can reformulate equation (4.32) in terms of the slowly varying amplitude  $\psi(\mathbf{r}_t, z)$

$$\frac{\partial}{\partial z} |\psi\rangle = i \frac{\nabla_t^2}{\bar{q} + \sqrt{\nabla_t^2 + \bar{q}^2}} |\psi\rangle. \quad (4.33)$$

Equation (4.33) can be solved by Fourier transformation with respect to the transverse coordinates  $\mathbf{r}_t$

$$f(\mathbf{q}_t) = \mathcal{F}[f] = \frac{1}{2\pi} \iint_{-\infty}^{+\infty} d^2 r_t e^{-i\mathbf{q}_t \mathbf{r}_t} f(\mathbf{r}_t) \quad (4.34)$$

If we keep in mind<sup>5</sup> that  $\mathcal{F}[(\nabla_t^2)^n f] = (-q_t^2)^n \mathcal{F}[f]$  we can Fourier transform Equation (4.33) by expanding the denominator into a Taylor series, transforming the series term by term and reestablishing the denominator afterwards. In fact, the whole procedure is equivalent to the replacement  $\nabla_t^2 \rightarrow -q_t^2$  in Equation (4.33). We end up with the first-order ordinary differential equation

$$\frac{\partial}{\partial z} |\psi\rangle = -i \frac{q_t^2}{\bar{q} + \sqrt{\bar{q}^2 - q_t^2}} |\psi\rangle. \quad (4.35)$$

The solution of this initial value problem which describes the evolution of the angular spectrum is

$$\psi(\mathbf{q}_t, z + \Delta z) = e^{-i\Delta q_z \Delta z} \psi(\mathbf{q}_t, z). \quad (4.36)$$

The choice  $q_z = q_{z,\text{sc}}$  with

$$\Delta q_{z,\text{sc}} = \frac{q_t^2}{\bar{q} + \sqrt{\bar{q}^2 - q_t^2}}. \quad (4.37)$$

corresponds to the solution of equation (4.35). For the reasons given later, the corresponding propagator is called the *semi-circle* BPM propagator.

So far, we have not made any approximations. Thus, the *classical beam propagation method exactly describes beam propagation in a homogeneous half-space*. If we replace the relative wave number by the leading term of its Taylor expansion, i.e.,

$$\Delta q_{z,p} = \frac{q_t^2}{2\bar{q}}, \quad (4.38)$$

we obtain the *paraxial* BPM propagator. We can derive this propagator directly if we carry out the above procedure for the paraxial Helmholtz equation (4.14).

5. This equation can easily be verified by partial integration.

We will now investigate the physical meaning of the BPM propagators in more detail. First of all, the Fourier transformation (4.34) is an expansion of the initial field into an angular spectrum of plane waves propagating in a medium with a dielectric constant  $\bar{\epsilon}$ . To simplify the further discussion, we assume that the medium is lossless  $\text{Im}(\bar{q}) = 0$  and we consider the BPM propagators for the (rapidly varying) field

$$\Psi(\mathbf{q}_t, z + \Delta z) = e^{iq_z \Delta z} \Psi(\mathbf{q}_t, z), \quad (4.39)$$

where

$$\begin{aligned} q_{z,\text{sc}} &= \sqrt{\bar{q}^2 - q_t^2} \\ q_{z,p} &= \bar{q} - \frac{q_t^2}{2\bar{q}} \end{aligned}$$

are the wave numbers for the semi-circle (sc) and for the paraxial (*p*) case in place of the propagators for the amplitude  $\psi$ . On the basis of these assumptions the BPM propagator merely has to transport the initial phase portrait at  $z$  to the target transverse plane at  $z + \Delta z$ . The amplitude of the plane waves must not be changed during propagation. Figure 4.6 shows the wave number for the semi-circle  $q_{z,\text{sc}}$  and

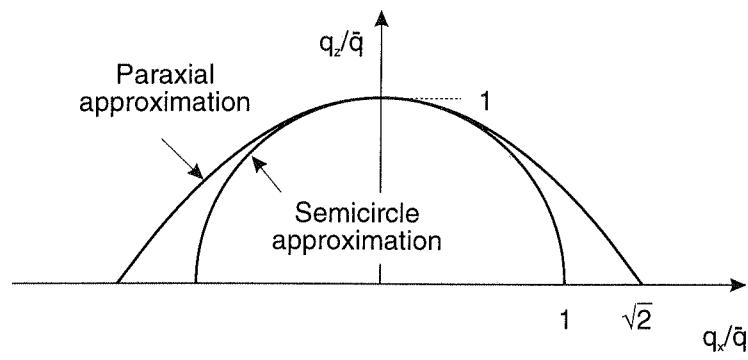


Figure 4.6. The  $z$ -components of the wave numbers for the semi-circle  $q_{z,\text{sc}}$  and for the paraxial  $q_{z,p}$  BPM propagator as a function of the transverse wave number  $q_t$ .

for the paraxial  $q_{z,p}$  propagator as a function of the transverse wave number  $q_t$ . The function  $q_{z,\text{sc}}(q_t)$  forms a semi-circle (therefore “semi-circle” propagator). Physically speaking, all waves of the angular spectrum<sup>6</sup> have the same wave number  $\bar{q}$ . The (approximated) paraxial propagator, in contrast, conserves the wave number only at  $q_t = 0$ . With an increasing transverse wave number  $q_t$ , i.e., with an increasing angle

6. For angles greater than  $\pi/2$ , i.e., for  $q_t^2 > \bar{q}^2$ , the wave number ( $q_{z,\text{sc}}$ ) becomes purely imaginary: the optical field becomes evanescent.

of aperture, the mismatch between the wave number used by the paraxial BPM propagator and the original wave number  $\bar{q}$  increases. Thus, the *paraxial Helmholtz equation is a small-angle approximation of the forward Helmholtz equation*. It holds only for beams having a small angle of aperture.

#### 4.2.2 Derivation of the Propagator

As stated in the introduction to this section, the classical beam propagation method attempts to completely separate all contributions caused by diffraction ( $\nabla_t^2$ ) from those due to inhomogeneities  $\Delta\epsilon$ . It further regards the inhomogeneities as a perturbation of diffraction. Thus, the operator  $\sqrt{\mathcal{H}}$  is expanded into powers of  $\mathcal{W}$ , i.e.,

$$\begin{aligned} \sqrt{\mathcal{H}} &= \sqrt{\mathcal{D} + \mathcal{W}} \\ &= \sqrt{\mathcal{D}} \left( 1 + \frac{1}{2} \mathcal{D}^{-1} \mathcal{W} - \dots \right) \\ &= \sqrt{\mathcal{D}} + \frac{\mathcal{W}}{2\bar{q}} \pm \dots \end{aligned}$$

To avoid many lengthy and complicated calculations<sup>7</sup> we do not solve the forward Helmholtz equation directly but merely elaborate the formal solution (4.9). Inserting this result, we obtain the approximate formal solution

$$|\Psi(z + \Delta z)\rangle = \exp \left( i \left( \sqrt{\mathcal{D}} + \frac{\mathcal{W}}{2\bar{q}} \right) \Delta z \right) |\Psi(z)\rangle \quad (4.40)$$

which is used by the classical BPM. Next, we replace the exponential function containing both operators  $\sqrt{\mathcal{D}}$  and  $\mathcal{W}/2\bar{q}$  by a product of functions containing either  $\sqrt{\mathcal{D}}$  or  $\mathcal{W}/2\bar{q}$ . This can be achieved using the operator expansion

$$\begin{aligned} &\exp \left( i \left( \sqrt{\mathcal{D}} + \frac{\mathcal{W}}{2\bar{q}} \right) \Delta z \right) \\ &= \exp \left( i \sqrt{\mathcal{D}} \frac{\Delta z}{2} \right) \exp \left( i \frac{\mathcal{W}}{2\bar{q}} \Delta z \right) \exp \left( i \sqrt{\mathcal{D}} \frac{\Delta z}{2} \right) \\ &\quad + i \frac{\Delta z^3}{24} \left( \left[ \sqrt{\mathcal{D}}, \left[ \sqrt{\mathcal{D}}, \frac{\mathcal{W}}{2\bar{q}} \right] \right] + \left[ \frac{\mathcal{W}}{2\bar{q}}, \left[ \frac{\mathcal{W}}{2\bar{q}}, \sqrt{\mathcal{D}} \right] \right] \right) \pm \dots, \quad (4.41) \end{aligned}$$

which is well known from quantum mechanics. We now use the first term on the right-hand side of this equation in place of the exponential function in Equation (4.40). Thus,  $\sqrt{\mathcal{D}}$  and  $\mathcal{W}/2\bar{q}$  are now separated. The deviations starting with the second term will be discussed later.

7. For the interested reader, a direct derivation of the BPM propagator is found in [3].

For our further elaboration we benefit from the fact that the operators describing the diffraction are diagonal in Fourier space,

$$\left\langle \mathbf{q}_t \left| \exp \left( i\sqrt{\mathcal{D}} \frac{\Delta z}{2} \right) \tilde{\mathbf{q}}_t \right. \right\rangle = \exp \left( iq_{z,\text{sc}} \frac{\Delta z}{2} \right) \delta(\mathbf{q}_t - \tilde{\mathbf{q}}_t), \quad (4.42)$$

and the operator describing the distortion of the phase front by the inhomogeneities is diagonal in position space,

$$\left\langle \mathbf{r}_t \left| \exp \left( i \frac{\mathcal{W}}{2\bar{q}} \Delta z \right) \tilde{\mathbf{r}}_t \right. \right\rangle = \exp \left( i \frac{\Delta\epsilon(\mathbf{r}_t)}{2\bar{\epsilon}} \bar{q} \Delta z \right) \delta(\tilde{\mathbf{r}}_t - \mathbf{r}_t). \quad (4.43)$$

The quantities in these equations were introduced in the previous paragraph.

In the following we make extensive use of Parseval's theorem, which states that the unity operator

$$\mathcal{E} = \sum_j |\phi_j\rangle \langle \phi_j| \quad (4.44)$$

can be represented by any complete set of functions. An important application of this theorem is the Fourier transformation where the plane waves

$$\langle \mathbf{q}_t | \mathbf{r}_t \rangle = \frac{1}{2\pi} e^{-i\mathbf{q}_t \cdot \mathbf{r}_t}$$

form the complete set of functions. We obtain

$$\begin{aligned} f(\mathbf{q}_t) &= \langle \mathbf{q}_t | f \rangle \\ &= \iint_{-\infty}^{+\infty} d^2 r_t \langle \mathbf{q}_t | \mathbf{r}_t \rangle \langle \mathbf{r}_t | f \rangle \\ &= \frac{1}{2\pi} \iint_{-\infty}^{+\infty} d^2 r_t \exp(-i\mathbf{q}_t \cdot \mathbf{r}_t) f(\mathbf{r}_t) \\ &= \mathcal{F}[f], \end{aligned}$$

i.e., the usual formula (compare (4.34)) for the Fourier transformation, and we also get some idea of this technique. We now elaborate the formal solution by inserting appropriate expansions of the unity operator between each two terms of the operator:

$$\begin{aligned} \Psi(\mathbf{q}_t, z + \Delta z) \\ &= \langle \mathbf{q}_t | \Psi(z + \Delta z) \rangle \\ &= \left\langle \mathbf{q}_t \left| \exp \left( i\sqrt{\mathcal{D}} \frac{\Delta z}{2} \right) \exp \left( i \frac{\mathcal{W}}{2\bar{q}} \Delta z \right) \exp \left( i\sqrt{\mathcal{D}} \frac{\Delta z}{2} \right) \Psi(z) \right. \right\rangle \end{aligned}$$

$$\begin{aligned} &= \exp \left( i\sqrt{\bar{q}^2 - q_t^2} \frac{\Delta z}{2} \right) \iint_{-\infty}^{+\infty} d^2 \tilde{q}_t \left\langle \mathbf{q}_t \left| \exp \left( i \frac{\mathcal{W}}{2\bar{q}} \Delta z \right) \tilde{\mathbf{q}}_t \right. \right\rangle \\ &\quad \times \exp \left( i\sqrt{\bar{q}^2 - \tilde{q}_t^2} \frac{\Delta z}{2} \right) \Psi(\tilde{\mathbf{q}}_t, z) \\ &= \exp \left( iq_z \frac{\Delta z}{2} \right) \\ &\quad \times \mathcal{F} \left[ \exp \left( i \frac{\Delta\epsilon(\mathbf{r}_t)}{2\bar{\epsilon}} \bar{q} \Delta z \right) \mathcal{F}^{-1} \left[ \exp \left( i\tilde{q}_z \frac{\Delta z}{2} \right) \Psi(\tilde{\mathbf{q}}_t, z) \right] \right] \end{aligned} \quad (4.45)$$

Although we have left out some intermediate steps for this short derivation, the general procedure becomes clear. Naturally, we may also replace the semi-circle wave number  $q_{z,\text{sc}}$  by the wave number of the paraxial operator  $q_{z,p}$ .

Since the operator  $\bar{q}$  – precisely  $(\bar{q} \mathcal{E})$  – commutes with both operators  $\sqrt{\mathcal{D}}$  and  $\mathcal{W}/2\bar{q}$  we may also write

$$\begin{aligned} \psi(\mathbf{q}_t, z + \Delta z) &= \exp \left( -i\Delta q_z \frac{\Delta z}{2} \right) \\ &\quad \times \mathcal{F} \left[ \exp \left( i \frac{\Delta\epsilon(\mathbf{r}_t)}{2\bar{\epsilon}} \bar{q} \Delta z \right) \mathcal{F}^{-1} \left[ \exp \left( -i\Delta \tilde{q}_z \frac{\Delta z}{2} \right) \psi(\tilde{\mathbf{q}}_t, z) \right] \right], \end{aligned} \quad (4.46)$$

i.e., the corresponding equation for the slowly varying amplitudes.

### 4.2.3 Interpretation of the Classical BPM

In the previous paragraph, we derived the propagator of the classical using rather formal operator techniques. We shall now turn our interest to the physical interpretation of the classical BPM and to a discussion of its limits. Figure 4.7 shows the propagation scheme of the classical BPM. It is a graphical representation of Equations (4.45) and (4.46). The propagation starts at  $z = 0$ . The first propagation step comprises the following tasks:

- The input field is transformed into its angular spectrum:

$$\Psi(\mathbf{r}_t, 0) \rightarrow \Psi(\mathbf{q}_t, 0)$$

- The angular spectrum is propagated by one half-step:

$$\Psi(\mathbf{q}_t, 0) \rightarrow \Psi_{\text{old}}(\mathbf{q}_t, \frac{\Delta z}{2})$$

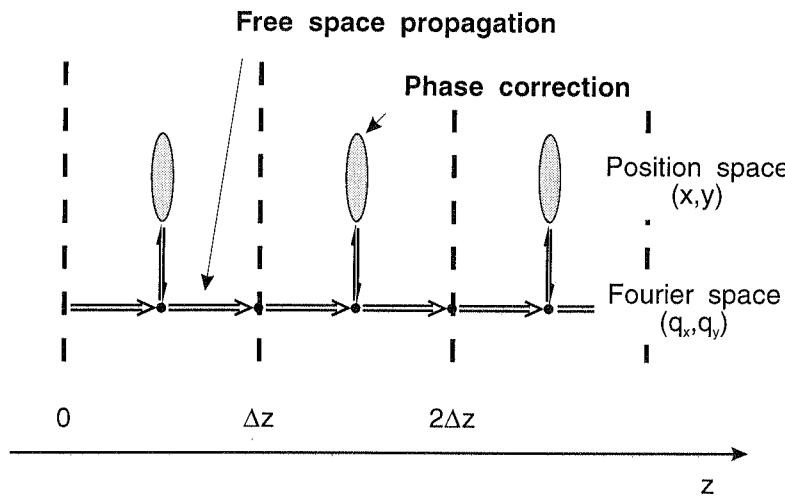


Figure 4.7. Propagation scheme of the classical beam propagation method.

- The angular spectrum is transformed back into position space:

$$\Psi_{\text{old}}(q_t, \frac{\Delta z}{2}) \rightarrow \Psi_{\text{old}}(r_t, \frac{\Delta z}{2})$$

- The phase portrait is corrected:

$$\Psi_{\text{old}}(r_t, \frac{\Delta z}{2}) \rightarrow \Psi_{\text{new}}(r_t, \frac{\Delta z}{2})$$

- The optical field is transformed into its angular spectrum:

$$\Psi_{\text{new}}(r_t, \frac{\Delta z}{2}) \rightarrow \Psi_{\text{new}}(q_t, \frac{\Delta z}{2})$$

- The angular spectrum is propagated by one half-step:

$$\Psi_{\text{new}}(q_t, \frac{\Delta z}{2}) \rightarrow \Psi(q_t, \Delta z)$$

- The angular spectrum is transformed back into position space:

$$\Psi(q_t, \Delta z) \rightarrow \Psi(r_t, \Delta z).$$

For each new propagation step, all these actions are repeated. Appendix B shows for further illustration a simple software program.

Thus, the propagation of an optical beam in an inhomogeneous medium is replaced by a sequence of free space propagation steps and phase corrections. In drastic terms, the input beam knows nothing about the inhomogeneities located in the interval  $[0, \Delta z]$  until it arrives at  $\Delta z/2$ . Then the entire information about the inhomogeneities in the interval is replaced by a single phase correction. Afterwards, the beam propagates the second half-step without being influenced by any inhomogeneity.

We intuitively suspect such an algorithm to be sensitive with respect to the step size  $\Delta z$ . To obtain a more formal understanding of this effect, we will discuss the operator expansion (see Equation (4.41))

$$\begin{aligned} & \exp \left( i \left( \sqrt{\mathcal{D}} + \frac{\mathcal{W}}{2\bar{q}} \right) \Delta z \right) \\ = & \exp \left( i \sqrt{\mathcal{D}} \frac{\Delta z}{2} \right) \exp \left( i \frac{\mathcal{W}}{2\bar{q}} \Delta z \right) \exp \left( i \sqrt{\mathcal{D}} \frac{\Delta z}{2} \right) \\ & + i \frac{\Delta z^3}{24} \left( \left[ \sqrt{\mathcal{D}}, \left[ \sqrt{\mathcal{D}}, \frac{\mathcal{W}}{2\bar{q}} \right] \right] + \left[ \frac{\mathcal{W}}{2\bar{q}}, \left[ \frac{\mathcal{W}}{2\bar{q}}, \sqrt{\mathcal{D}} \right] \right] \right) \pm \dots, \end{aligned}$$

in more detail. We recognize that both terms of the most significant deviation include the commutator  $\left[ \sqrt{\mathcal{D}}, \frac{\mathcal{W}}{2\bar{q}} \right]$ . The leading term of this commutator is

$$\left\langle r_t \left| \left[ \nabla_t^2, k_0^2 \frac{\Delta \epsilon}{\epsilon} \right] \tilde{r}_t \right. \right\rangle = k_0^2 \nabla_t^2 \left( \frac{\Delta \epsilon}{\epsilon} \right) \delta(\tilde{r}_t - r_t).$$

We recognize that the derivatives of the dielectric function, which in fact change the light path during propagation, are also responsible for the mismatch between the two representations of the exponential function. Furthermore, we recognize that for a given structure the perturbation term increases with  $\Delta z$ . In other words; *The smoothness of the refractive index profile determines the maximum step size of the classical BPM and, in consequence, the performance of the whole algorithm.*

As expected from this assessment, the classical BPM does a good job for diffused refractive index profiles which are typical for devices realized by ion exchange in glass or by ion implantation in LiNbO<sub>3</sub>, but it turns out to be inadequate for step-index profiles typical of devices in the III-V material systems InGaAsP or GaAlAs.

#### 4.2.4 Numerical Fourier Transformation

In the previous paragraphs of Section 4.2 we derived the classical beam propagation method. We illuminated the physical ideas behind this algorithm and pointed out its limits. We will now turn our interest to the realization of the algorithm on a computer.

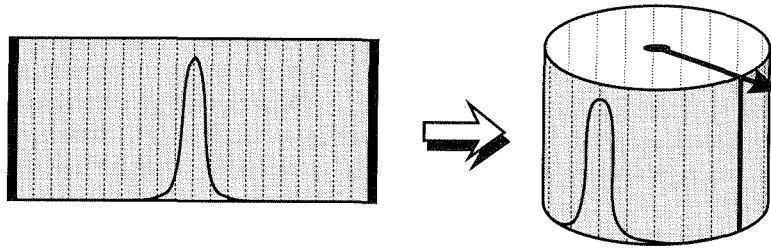


Figure 4.8. Numerical Fourier transformation.

We start with a short examination of the numerical Fourier transformation of a function  $f(x)$ , i.e.,

$$f(q) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx e^{-iqx} f(x),$$

and its inverse, i.e.,

$$f(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dq e^{iqx} f(q).$$

For the numerical treatment, we map a window of each function  $f(x)$  and  $f(q)$  onto a circle  $S^1$ . The two circles with perimeters  $L_x$  and  $L_q$  are divided by  $N$  equidistant points (see Figure 4.8). In terms of another terminology, we apply periodic boundary conditions to the windows of  $f(x)$  and  $f(q)$ . To avoid confusion between the entire Fourier transformation operating on the real axis and its numerical approximation, we indicate the approximated functions by capital letters, i.e.,  $F(x)$  and  $F(q)$ . We then may represent both functions

$$F(x_l) = \frac{1}{\sqrt{N}} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}-1} F(q_j) e^{iq_j x_l}$$

and

$$F(q_j) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} F(x_l) e^{-iq_j x_l}$$

by a finite Fourier series. The coefficients  $q_j$  and  $x_l$  are subject to the following discussion.

Let us imagine that we would like to transform a function  $f(x)$  into Fourier space. We then choose the coordinate system in position space and the size of the window  $[0, L_x]$  so that the function  $f(x)$  is “small” outside the window. Obviously, we must guarantee that the same condition is satisfied in Fourier space, i.e., that  $f(q)$  is “small” outside the window  $[-L_q/2, L_q/2]$ . If the window is discretized by  $N$  points, we obtain  $x_l = l \Delta x$  with  $\Delta x = L_x/N$ . Furthermore, we know that

$q_j = j \Delta q$  with  $\Delta q = L_q/N = 2\pi/L_x$ . Thus, the window in Fourier space and that in position space are interdependent. We obtain

$$L_q = \frac{2\pi}{\Delta x} \quad (4.47)$$

and

$$\Delta q = \frac{2\pi}{L_x}, \quad (4.48)$$

i.e., the size of the window in Fourier space is determined by the discretization of the window in position space and vice versa. By a hand-waving argument, we may state that for the numerical Fourier transformation the original function (in position space) must be well described by its discretization. We will discuss the consequences of this theorem (sampling theorem) on the classical BPM later. A much more detailed treatment of this topic is found in [1].

The numerical 2D Fourier transformation is defined in the same way. The function  $f(x, y)$  is mapped onto the torus  $T^2 = S^1 \times S^1$ . All the above results apply to both coordinates  $x$  and  $y$ , independently.

The fast Fourier transformation (FFT) which is defined for the special choice  $N = 2^\tau$  is an effective algorithm for carrying out the numerical Fourier transformation. Many of the oscillating terms which determine the time complexity of the algorithm then coincide. The number of independent terms is reduced from  $2^{2\tau}$  to  $C(\tau) = \tau 2^{\tau-1}$ . Figure 4.9 shows the function  $\gamma(\tau) = C(\tau)/C(\tau - 1)$  which describes the CPU time of a Fourier transformation compared to one with half as many points. We recognize that the time complexity is quadratic ( $\gamma = 4$ ) for the Fourier transformation with four points, and that with increasing number of points it approaches an increasingly linear time behavior ( $\gamma = 2$ ).

What are the consequences of these results for the classical beam propagation method? First of all, the classical BPM operates on a rigid grid which often causes

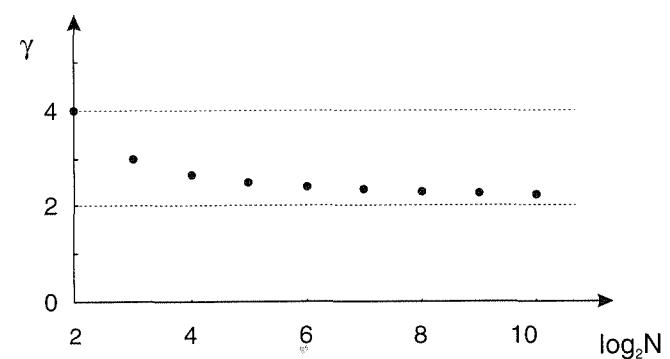


Figure 4.9. Increase of CPU time  $\gamma(\tau)$  for a Fourier transformation as a function of  $\tau = \log_2 N$  if the number of discretization points  $N$  is doubled.

a large numerical overhead. We discussed this point in more detail in Section 4.1. In addition, the sampling theorem means that the classical BPM is sensitive to the “non-smoothness” of the optical field. Since abrupt changes of the optical field and of the refractive index profile are usually interdependent the treatment of step refractive index profiles by the classical BPM results in a significant increase of CPU time (in comparison to the treatment of diffused waveguide structures). As an illustrative (and at the same time realistic) example we shall consider an etched rib waveguide device which requires an increase of the resolution by a factor of 4 for each coordinate  $x, y$  and a parallel decrease in step size by a factor of 4 in comparison with the same diffused waveguide device. In that case, the simulation of the etched structure will need approximately 100-fold more CPU time. This feature of the classical BPM makes it difficult to study true 3D waveguide structures where the influence of the chip surface on the operation of a device is typically the central point of interest. For that reason, the full 3D version of the classical BPM is rarely used, although the algorithm is easy to implement.

Second, due to the mapping of the windows on the circles, i.e., on compact manifolds, the classical BPM exhibits strange behavior if the optical field arrives at the boundary. Figure 4.10 shows the free space propagation of a Gaussian beam.

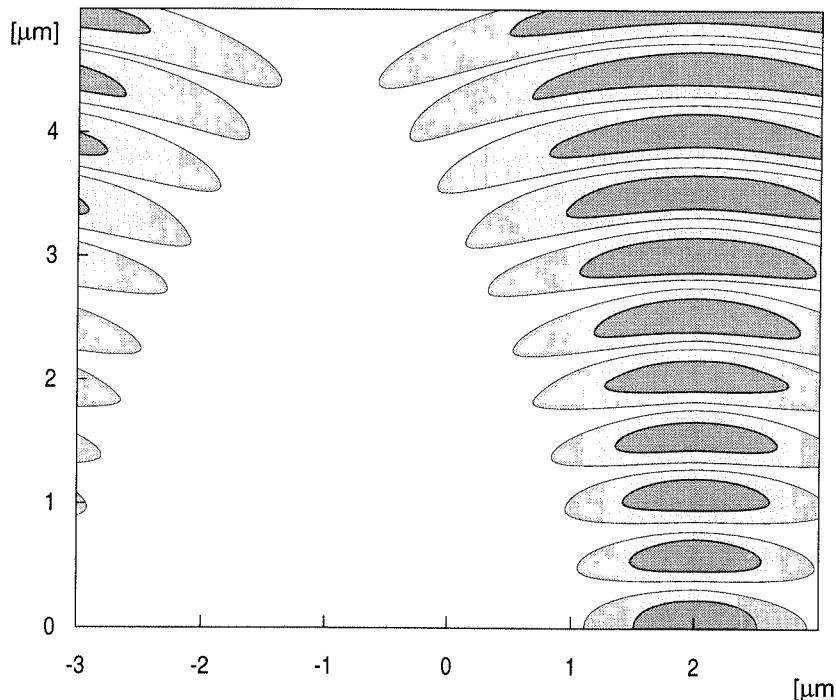


Figure 4.10. Influence of periodic boundary conditions on the behavior of the classical BPM.

After a few steps, the optical field leaves the operation window on the right-hand side and immediately returns to the operation window on the left-hand side. Looking at Figure 4.8 we can easily understand this behavior. This effect is, of course, damped out by absorbing walls. However, we should keep it in mind, especially if we find interferences close to the boundary of the BPM window.

#### 4.2.5 Optimized BPM Propagators

For the numerical evaluation of the forward Helmholtz equation, we also replaced the continuous function  $f(x)$  by a discrete function  $f(x_i)$  and the continuous Fourier transformation by a discrete Fourier series. To ensure a consistent treatment, we must replace the (continuous) derivatives by finite differences.<sup>8</sup> If we use the central difference approximation for the second derivatives, we obtain

$$\begin{aligned} \mathcal{F}[f^{(CD)}(x)] &= \mathcal{F}\left[\frac{f(x + \Delta x) - 2f(x) + f(x - \Delta x)}{\Delta x^2}\right] \\ &= q^2 \left[ \frac{\exp(-iq\Delta x) - 2 + \exp(iq\Delta x)}{(q\Delta x)^2} \right] f(q) \\ &= -q^2 \operatorname{sinc}^2\left(\frac{q\Delta x}{2}\right) f(q). \end{aligned} \quad (4.49)$$

If we compare this result with the standard expression  $\mathcal{F}[f''(x)] = -q^2 f(q)$  for the second derivative, we see that we have to replace

$$q_t^2 \rightarrow \tilde{q}_t^2 = q_x^2 \operatorname{sinc}^2\left(\frac{q_x \Delta x}{2}\right) + q_y^2 \operatorname{sinc}^2\left(\frac{q_y \Delta y}{2}\right).$$

in all propagators of the classical BPM to avoid numerical viscosity. Thus, we obtain

$$\Delta \tilde{q}_{z,\text{sc}} = \frac{\tilde{q}_t^2}{\sqrt{\bar{q}^2 - \tilde{q}_t^2} + \bar{q}} \quad (4.50)$$

$$\Delta \tilde{q}_{z,\text{p}} = \frac{\tilde{q}_t^2}{2\bar{q}} \quad (4.51)$$

for the wave numbers occurring in the propagators of the amplitude and

$$\tilde{q}_{z,\text{sc}} = \sqrt{\bar{q}^2 - \tilde{q}_t^2} \quad (4.52)$$

$$\tilde{q}_{z,\text{p}} = \bar{q} - \frac{\tilde{q}_t^2}{2\bar{q}} \quad (4.53)$$

for those of the optical field.

8. If we leave out this step, the large-angle waves are not treated correctly; i.e., they are attenuated (positive numerical viscosity) or amplified (negative numerical viscosity) [4].

### 4.3 DIFFRACTION INTEGRALS

So far we have studied the forward Helmholtz equation. In this section, we will show that solving the paraxial Helmholtz equation is equivalent to calculating the Fresnel diffraction integral. We will derive then the Fraunhofer diffraction integral, which is a far-field approximation of Fresnel's diffraction integral, and discuss its properties.

For later use we first define the convolution

$$g \otimes f = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} d\xi g(x - \xi) f(\xi) \quad (4.54)$$

of two position space functions  $f(x)$  and  $g(x)$ . The Fourier transform of a convolution then is given by

$$\begin{aligned} \mathcal{F}[g \otimes f] &= \frac{1}{2\pi} \iint_{-\infty}^{+\infty} dx d\xi e^{iqx} g(x - \xi) f(\xi) \\ &= \frac{1}{(2\pi)^2} \iint_{-\infty}^{+\infty} dx d\xi \iint_{-\infty}^{+\infty} d\tilde{q} d\hat{q} e^{iqx - i\tilde{q}(x - \xi) - i\hat{q}\xi} g(\tilde{q}) f(\hat{q}) \\ &= g(q) f(q), \end{aligned} \quad (4.55)$$

i.e., by the product of the Fourier transforms  $g(q)$  and  $f(q)$ .

Since the evolution into  $z$ -direction according to the paraxial Helmholtz equation (4.36), i.e.,

$$\psi(\mathbf{q}_t, z) = \exp\left(-i\frac{q_t^2}{2\bar{q}} z\right) \psi(\mathbf{q}_t, 0),$$

is a product of functions in Fourier space, it may be regarded as the Fourier transform of a convolution. Using the inverse Fourier transform<sup>9</sup>

$$\begin{aligned} \mathcal{F}^{-1}\left[\exp\left(-iaq^2\right)\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dq \exp\left(iqx - iaq^2\right) \\ &= \sqrt{\frac{1}{4\pi a}} \exp\left(i\frac{x^2}{4a} - i\frac{\pi}{4}\right) \end{aligned}$$

9. This integral can be carried out by completing the square and utilizing

$$\int_{-\infty}^{\infty} d\xi e^{-i\xi^2} = \sqrt{\pi} \exp\left(-i\frac{\pi}{4}\right).$$

we obtain the Fresnel diffraction integral

$$\psi(\mathbf{r}_t, z) = -i \frac{\bar{q}}{2\pi z} \iint_{-\infty}^{+\infty} d^2\tilde{r}_t \exp\left(i\frac{\bar{q}}{2z} (\tilde{\mathbf{r}}_t - \mathbf{r}_t)^2\right) \psi(\tilde{\mathbf{r}}_t, 0). \quad (4.56)$$

The Fraunhofer diffraction integral<sup>10</sup> is a far-field approximation ( $\tilde{r}_t \ll r_t$ ) of the Fresnel diffraction integral, i.e.,

$$\psi(\mathbf{r}_t, z) = -i \frac{\bar{q}}{z} \exp\left(i\frac{\bar{q}r_t^2}{2z}\right) \psi(\bar{q}\mathbf{r}_t/z, 0). \quad (4.57)$$

This means that *the far-field  $\psi(\mathbf{r}_t, z)$  forms an image of the Fourier transform of the near-field  $\psi(\mathbf{q}_t, 0)$  in position space*.

The first term in equation (4.57) describes the reduction of the local amplitude due to the spread of energy. We can easily understand that term if we keep in mind that  $|\psi|^2 \propto 1/z^2$  and that the surface  $O$  of a sphere around  $z = 0$  grows with  $O \propto z^2$ . The second term is due to the curved phase fronts of the beam in the far-field. Regarding the near-field as a point source located at  $z = 0$ , we expect spherical phase fronts. From the Fraunhofer diffraction integral for the rapidly varying optical field, however, we obtain

$$z(\mathbf{r}_t) = z_0 - \frac{r_t^2}{2z_0} + \frac{r_t^2 \Delta z}{2z_0^2} - \dots,$$

i.e., phase fronts of nearly parabolic shape. The curvature  $k = 1/z_0$  at the optical axis agrees with that of the expected sphere. With increasing angle, the mismatch between the paraboloid and the expected sphere will increase. This mismatch is an artifact of the approximations on the way from the Helmholtz equation to the Fraunhofer diffraction integral.

### 4.4 IMPROVED BPM ALGORITHMS

In Section 4.2 we derived the classical beam propagation method. This operator technique for the propagator  $\exp(i\sqrt{H}\Delta z)$  is intended to split the beam propagation in an inhomogeneous medium into steps of free space propagation followed by phase corrections to account for the inhomogeneities. We have seen that this treatment makes the classical BPM sensitive to the step size  $\Delta z$ . We will now examine some modern BPM algorithms which overcome this problem. All these algorithms share the property of avoiding the split-operator technique.

10. The Fraunhofer diffraction integral can also be derived by using Huygen's principle.

#### 4.4.1 Eigenmode Based BPM

This algorithm utilizes the local eigenmodes at the actual cross section to evaluate this particular BPM step.<sup>11</sup> Let us consider an eigenmode  $j$  having a propagation constant  $q_j$  and an amplitude  $\psi_j$ . The eigenmode equation corresponding to the scalar Helmholtz equation then can be written as

$$\mathcal{H}\psi_j = q_j^2\psi_j.$$

Since all eigenmodes – guided *and* radiation modes – form a complete spectrum we can represent the entire field at  $z_0$  by

$$\Psi(x, y, z_0) = \sum_j a_j(z_0) e^{iq_j z_0} \psi_j(x, y) \quad (4.58)$$

in terms of the eigenmodes. In the numerical treatment, the radiation modes are naturally replaced by the quasi radiation modes of a large finite cavity. In Equation (4.58) the complex quantity  $a_j(z_0)$  stands for the relative strength and phase of the eigenmode  $j$  within the entire field  $\Psi_{\pm}(z_0)$ . It is calculated by decomposition of the entire field

$$a_j(z_0) = \langle \psi_j | \Psi(z_0) \rangle e^{-iq_j z_0}. \quad (4.59)$$

The optical field after one propagation step is easily calculated by

$$\begin{aligned} |\Psi(z_0 + \Delta z)\rangle &= e^{i\sqrt{\mathcal{P}}t} |\Psi(z_0)\rangle \\ &= \sum_j a_j e^{iq_j(z_0+\Delta z)} |\psi_j\rangle. \end{aligned} \quad (4.60)$$

So far, we have analyzed a one-step beam propagation method. A multi-step BPM proceeds as shown in Figure 4.11. At the beginning of each propagation step (at  $z_0$ ), the eigenmodes of the corresponding cross section are calculated. Next, the entire field must be expanded into the complete set of eigenmodes. The eigenmodes are then propagated to  $z = z_0 + \Delta z$  and the optical field is reassembled.

We have not yet made any assumptions either about the underlying Hamiltonian  $\mathcal{H}$  or the algorithm used for the calculation of the eigenmodes. Thus, the eigenmode-based BPM can be applied to beam propagation according to the scalar Helmholtz equation for both TE and TM polarization. It is also completely independent of the refractive index contrast. The eigenmode-based beam propagation method has two more important advantages. First, it solves the forward Helmholtz equation itself and not one of its approximate forms. So, the algorithm can tackle wide-angle problems. Second, the propagation step size is limited by the longitudinal variations of the refractive index profile and not by the algorithm itself. At a brief glance, therefore, the eigenmode-based BPM seems to be the best way

11. We shall keep in mind here that the refractive index profile is assumed to be constant over one propagation step.

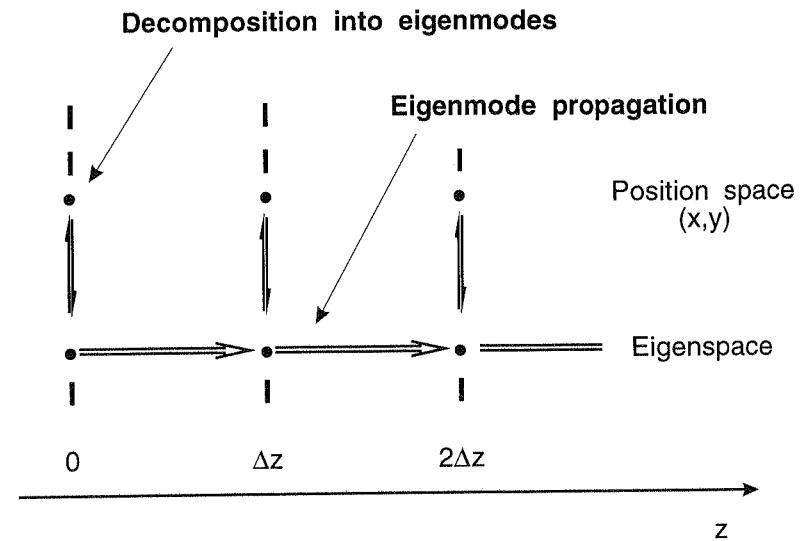


Figure 4.11. Eigenmode based beam propagation method.

of solving beam propagation. However, the applied strategy results in numerical overhead. This becomes clear if we compare the eigenmode-based BPM with the most similar algorithm, namely, the classical BPM. Both methods are based on the expansion of the entire field to a complete set of functions. The classical BPM utilizes the (problem-independent) plane waves, whereas the eigenmode based BPM utilizes the problem-adapted set of eigenmodes. The modal analysis required for guided and radiation modes which has to be performed after each propagation step represents an extra task of the eigenmode based BPM compared with the classical algorithm (see Figure 4.11). The decomposition of the entire field, which is a fast Fourier transformation (FFT) for the classical BPM reduces the performance of the eigenmode based algorithm even further. For these reasons, the eigenmode based BPM is rarely used despite its attractive features.

#### 4.4.2 Finite Differences: FD-BPM

Finite difference algorithms are easy to derive and implement. We will concentrate here on algorithms for the paraxial Helmholtz equations

$$\frac{\partial \psi}{\partial z} = i \frac{\bar{\mathcal{P}}}{2} \mathcal{P} \psi.$$

The extension to wide-angle approximations is discussed in the paper by Hadley [5]. These algorithms usually operate on a rigid grid; i.e., the spacing between two grid

points along the  $x$ -axis

$$\Delta x = x_{m+1} - x_m,$$

is constant. The derivatives along this axis are represented by the central differences

$$f_m = f(x_m)$$

$$f'_m = \frac{f_{m+1} - f_{m-1}}{2\Delta x}$$

$$f''_m = \frac{f_{m+1} - 2f_m + f_{m-1}}{\Delta x^2}.$$

In general, finite difference algorithms are divided into two classes, explicit and implicit algorithms. The explicit FD algorithms are based on the direct solution

$$\psi(z + \Delta z) = \exp\left(i\frac{\bar{q}\Delta z}{2}\mathcal{P}\right)\psi(z), \quad (4.61)$$

and the simplest algorithm of this type yields

$$\psi_m(z + \Delta z) = \sum_n \left(\mathcal{E} + i\frac{\bar{q}\Delta z}{2}\mathcal{P}\right)_{mn} \psi_n(z). \quad (4.62)$$

Thus, we obtain the solution of the paraxial Helmholtz equation by merely performing a matrix multiplication (with a band matrix). The application of more sophisticated explicit finite-difference schemes is found in [6]. Apart from their lack of stability with respect to increasing step size explicit FD algorithms have another severe disadvantage for the application to beam propagation: *Even if the operator  $\mathcal{P}$  is unitary, the approximated explicit FD operators no longer have this characteristic,*

$$\mathcal{E} - i\frac{\bar{q}\Delta z}{2}\mathcal{P} \neq \left(\mathcal{E} + i\frac{\bar{q}\Delta z}{2}\mathcal{P}\right)^{-1},$$

i.e., the application of explicit finite-difference algorithms results in numerical losses during propagation (see Section 4.1).

Implicit FD algorithms match the solution of the entire differential equation at  $z + \Delta z/2$ . We then obtain

$$\exp\left(-i\frac{\bar{q}\Delta z}{4}\mathcal{P}(z + \Delta z)\right)\psi(z + \Delta z) = \exp\left(i\frac{\bar{q}\Delta z}{4}\mathcal{P}(z)\right)\psi(z). \quad (4.63)$$

The simplest algorithm of this type is the Crank-Nicolson scheme

$$\sum_n C_{mn}^-(z + \Delta z)\psi_n(z + \Delta z) = \sum_p C_{mp}^+(z)\psi_p(z) \quad (4.64)$$

with

$$C_{mn}^\pm(z) = \left(\mathcal{E} \pm i\frac{\bar{q}\Delta z}{4}\mathcal{P}(z)\right)_{mn}.$$

At first glance, we recognize that we must now solve a system of linear equations to obtain the solution  $\psi(z + \Delta z)$  of the paraxial Helmholtz equation. This is a disadvantage compared to the explicit FD schemes discussed previously. However, implicit FD schemes have two bonus points that prevail over the drawback mentioned earlier. First, implicit FD schemes are stable for any step size. The second (and more important) argument is that symmetric implicit FD schemes do not yield numerical losses during propagation. For a proof of this statement, we assume a longitudinal invariant  $\mathcal{P}(z + \Delta z) = \mathcal{P}(z)$  unitary operator  $\mathcal{P}^\dagger = \mathcal{P}$ . We then obtain

$$\begin{aligned} \left[ \frac{\sum_{n=0}^N (-i\bar{q}\Delta z\mathcal{P}/4)^n}{\sum_{m=0}^N (i\bar{q}\Delta z\mathcal{P}/4)^m} \right]^\dagger &= \frac{\sum_{n=0}^N (i\bar{q}\Delta z\mathcal{P}/4)^n}{\sum_{m=0}^N (-i\bar{q}\Delta z\mathcal{P}/4)^m} \\ &= \left[ \frac{\sum_{n=0}^N (-i\bar{q}\Delta z\mathcal{P}/4)^n}{\sum_{m=0}^N (i\bar{q}\Delta z\mathcal{P}/4)^m} \right]^{-1}, \end{aligned}$$

i.e., the numerical operator is also unitary. The denominator was again expanded into a Taylor series treated term by term and combined into the new denominator.

The following examples will show explicit expressions for the matrix elements of an implicit finite-difference scheme for the scalar forward Helmholtz equation for both two and three dimensions and for both TE and TM polarization.

## 2D FD-BPM

For the sake of simplicity, we exclude the two boundary points  $L$  and  $R$  from the ordinary enumeration of the grid (see Figure 4.12). Using the equations (4.6) and (4.7) and the central differences defined previously, we obtain

$$C_{mn}^\pm(z) = (1 + d_m(z))\delta_{m,n+1} + c_m^\pm(z)\delta_{m,n} + (1 - d_m(z))\delta_{m,n-1}, \quad (4.65)$$

where

$$\delta_{m,n} = \begin{cases} 1 & \text{for } m = n \\ 0 & \text{otherwise} \end{cases} \quad (4.66)$$

is Kronecker's  $\delta$ . The other abbreviations used in (4.66) equation are

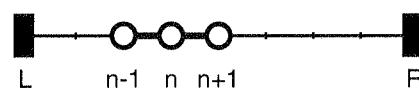
$$c_m^\pm(z) = k_0^2 \Delta \epsilon_m(z) \Delta x^2 - 2 \pm i \frac{4\bar{q}\Delta x^2}{\Delta z},$$

and

$$d_m(z) = \begin{cases} 0 & \text{for TE polarization} \\ \frac{\Delta \epsilon_{m+1}(z) - \Delta \epsilon_{m-1}(z)}{2(\epsilon + \Delta \epsilon_m(z))} & \text{for TM polarization.} \end{cases}$$

Thus, the matrix  $(C_{mn}^\pm)$  for the two-dimensional FD-BPM is a tridiagonal band matrix. Figure 4.12 shows the stencil of the interdependent points. The Dirichlet

a) 2D FD-BPM



b) 3D FD-BPM

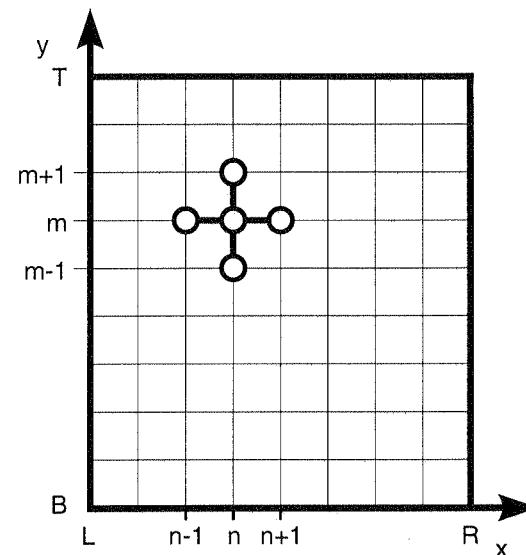


Figure 4.12. Enumeration of points and stencil for the 2D FD-BPM (a) and 3D FD-BPM (b).

boundary condition  $\psi_L = \psi_R = 0$  is the natural boundary condition of this algorithm; i.e., it is satisfied without any changes. If the first two points of the window are interdependent (e.g., for the Neumann boundary condition), i.e.,

$$\begin{aligned}\psi_L &= a_L \psi_1 \\ \psi_R &= a_R \psi_N,\end{aligned}$$

the two matrix elements

$$\begin{aligned}C_{11}^{\pm}(z) &= c_1^{\pm}(z) + a_L [1 - d_m(z)] \\ C_{NN}^{\pm}(z) &= c_N^{\pm}(z) + a_R [1 + d_m(z)]\end{aligned}$$

must be replaced.

### 3D FD-BPM

Since we do not want to go into details of enumerating grid points, we will introduce here doublet indices  $(n, m)$  to represent a specific point on the  $xy$ -plane. The first index of these tuples belongs to the  $x$ -coordinate, and the second one to the  $y$ -coordinate. The boundary itself is again excluded from enumeration. We refer to it by the indices  $L$  (left),  $R$  (right),  $T$  (top) and  $B$  (bottom). We then obtain

$$\begin{aligned}C_{(m,n)(o,p)}^{\pm}(z) &= \Delta x^2 \delta_{(m,n)(o,p+1)} \\ &+ [1 + d_{(m,n)}(z)] \Delta y^2 \delta_{(m,n)(o+1,p)} + c_{(m,n)}^{\pm}(z) \delta_{(m,n)(o,p)} \\ &+ [1 - d_{(m,n)}(z)] \Delta y^2 \delta_{(m,n)(o-1,p)} + \Delta x^2 \delta_{(m,n)(o,p-1)}\end{aligned}\quad (4.67)$$

with

$$c_{(m,n)}^{\pm}(z) = k_0^2 \Delta \epsilon_{(m,n)}(z) \Delta x^2 \Delta y^2 - 2(\Delta x^2 + \Delta y^2) \pm i \frac{4\bar{q} \Delta x^2 \Delta y^2}{\Delta z}$$

and

$$d_{(m,n)}(z) = \begin{cases} 0 & \text{for TE polarization} \\ \frac{\Delta \epsilon_{(m+1,n)}(z) - \Delta \epsilon_{(m-1,n)}(z)}{2(\bar{\epsilon} + \Delta \epsilon_{(m,n)}(z))} & \text{for TM polarization.} \end{cases}$$

The remarks on the boundary conditions made in the previous paragraph also apply to the three-dimensional case.

### 4.4.3 Finite Elements: FE-BPM

This algorithm considers the paraxial Helmholtz equation as the Euler-Lagrange equation of a  $z$ -transient variational principle (ZTV) and treats the variational principle by a finite-element procedure.

For the derivation of the algorithm, we first consider the functional

$$W(\psi) = \int_A dx dy \{T(x, y, z, \psi, \psi_x, \psi_y) - i\zeta \psi \psi_z\} \quad (4.68)$$

where we used the more compact notations  $\psi_x = \partial \psi / \partial x$  for the derivatives of the optical field  $\psi$ .  $A$  is the window of operation in the  $xy$ -plane of the transverse coordinates. This functional becomes stationary ( $\delta W = 0$ ) in the  $xy$ -plane if the Euler-Lagrange equation

$$\frac{\partial}{\partial x} \frac{\partial T}{\partial \psi_x} + \frac{\partial}{\partial y} \frac{\partial T}{\partial \psi_y} - \frac{\partial T}{\partial \psi} + i\zeta \psi_z = 0 \quad (4.69)$$

holds on an area  $A$  together with the boundary condition

$$\left[ \frac{\partial T}{\partial \psi_x} n_x + \frac{\partial T}{\partial \psi_y} n_y \right]_{\partial A} = 0.$$

To obtain the paraxial Helmholtz equations, we have to choose

$$T = \frac{1}{2} [\chi(x, y) (\psi_x^2 + \psi_y^2) - \gamma(x, y) \psi_z^2], \quad (4.70)$$

which in turn corresponds to the Euler-Lagrange equation

$$\nabla_t (\chi \nabla_t \psi) + \gamma \psi + i\zeta \frac{\partial \psi}{\partial z} = 0 \quad (4.71)$$

with the Neumann boundary condition

$$[\nabla_t \psi \cdot \mathbf{n}]_{\partial A} = 0.$$

For the arbitrary parameter functions  $\chi, \gamma$  and  $\zeta$  occurring in equation (4.70), we choose

$$\begin{aligned} \chi &= 1 \\ \gamma &= k_0^2 \Delta \epsilon(x, y) \\ \zeta &= 2\bar{q} \end{aligned} \quad (4.72)$$

to obtain the paraxial Helmholtz equation for TE polarization and

$$\begin{aligned} \chi &= \frac{1}{\bar{\epsilon} + \Delta \epsilon(x, y)} \\ \gamma &= \frac{k_0^2 \Delta \epsilon(x, y)}{\bar{\epsilon} + \Delta \epsilon(x, y)} \\ \zeta &= \frac{2\bar{q}}{\bar{\epsilon} + \Delta \epsilon(x, y)} \end{aligned} \quad (4.73)$$

to obtain this equation for TM polarization where we used the alternative paraxial scalar Helmholtz equation

$$\left( \nabla \frac{1}{\epsilon} \nabla + k_0^2 \right) H_y = 0$$

for TM polarization.

For the subsequent treatment, we substitute the unknown function  $\psi$  by a test function  $\psi_{\text{test}}$  that has a finite number of parameters  $p_n$  (Rayleigh-Ritz procedure). The variational principle is then replaced by a minimization of the test function with respect to all its parameters, i.e.,

$$\delta W = 0 \rightarrow \frac{\partial W}{\partial p_n} = 0.$$

Following the usual finite-element procedure, we choose a number of nodes  $\psi_n$  in the area  $A$  and regard the values of the optical field at all these nodes as the parameters of the test function  $\psi_{\text{test}}$ . We connect the nodes and then divide the area into meshes. If these meshes are triangular, the solution is planar inside each mesh. The finite element expansion is performed (according to Galerkin) mesh by mesh, i.e.,

$$\psi_{\text{test}} = \sum_{m,n} b_{mn}(x, y) \psi_n(z), \quad (4.74)$$

where the index  $m$  runs over all meshes and the index  $n$  runs over all nodes. The matrix  $(b_{mn})$  is sparse, since all matrix elements which do not connect a mesh with its surrounding nodes will vanish. In addition, each matrix element  $b_{mn}(x, y)$  vanishes outside the area of the mesh  $m$ . Up to now, the only difference between this and the usual finite-element procedure is that the optical field at the nodes depends on the propagation coordinate  $z$ .

However, the minimization of the functional (4.68) using these test functions does not lead to a set of linear equations here, but results in

$$\frac{\partial W}{\partial \psi_n} = \sum_{n'} \left[ B_{nn'}(z) \psi_{n'} - i C_{nn'}(z) \frac{\partial \psi_{n'}}{\partial z} \right] = 0, \quad (4.75)$$

i.e., a system of ordinary differential equations which describes the evolution of the optical field at the nodes  $\psi_k(z)$  with respect to the propagation coordinate  $z$ . For the reasons stated earlier, the two matrices

$$B_{nn'} = \sum_m \left\{ \chi_m \iint_{A_m} dx dy \left[ \frac{\partial b_{mn}}{\partial x} \frac{\partial b_{mn'}}{\partial x} + \frac{\partial b_{mn}}{\partial y} \frac{\partial b_{mn'}}{\partial y} \right] - \gamma_m \iint_{A_m} dx dy b_{mn} b_{mn'} \right\} \quad (4.76)$$

and

$$C_{nn'} = \sum_m \zeta_m \iint_{A_m} dx dy b_{mn} b_{mn'} \quad (4.77)$$

occurring in Equation (4.75) are sparse. The abbreviations  $\chi_m$ ,  $\gamma_m$  and  $\zeta_m$  denote the value of the functions  $\chi$ ,  $\gamma$  and  $\zeta$  inside the mesh  $m$  where they are assumed constant. Finally, we solve the system of differential Equations (4.75) using an implicit finite difference scheme, i.e., by solving the equations

$$\sum_{n'} \left( C_{nn'} + i B_{nn'} \frac{\Delta z}{2} \right) \psi_{n'} \Big|_{z+\Delta z} = \sum_{n'} \left( C_{nn'} - i B_{nn'} \frac{\Delta z}{2} \right) \psi_{n'} \Big|_z. \quad (4.78)$$

A detailed description of this procedure is found in the previous section.

Highly sophisticated finite-element algorithms [7] operate on adaptive meshes. Although their description is beyond the limits of our book we will make a few comments about their features. The essential figure of merit for the performance of such a algorithm is a formulation which allows to add grid points to the mesh

without recalculating all matrix elements. In addition, an efficient local error estimator is required to control the numerical discretization of the cross section and the longitudinal step size. The control of numerical errors is known to be an advantage for itself especially for “error tolerant” algorithms such as the FE-BPM, which do not indicate erroneous results by their noisy shape. Much more important, however, are the advantages for the simulation of devices with longitudinally varying cross sections (e.g., tapers).

#### 4.5 VECTORIAL BEAM PROPAGATION METHOD

So far, we have restricted ourselves to BPM algorithms which deal with the scalar forward Helmholtz equation. We will now use the vector  $\mathbf{H}$ -field formulation since the transverse magnetic (in contrast to the electric) field components are continuous at a transverse dielectric interface. As shown in Section 4.1.1 the paraxial approximation of the forward Helmholtz equation is then given by

$$\frac{\partial \mathbf{h}_t}{\partial z} = i \frac{\bar{q}}{2} \mathcal{P}_{\mathbf{H}} \mathbf{h}_t \quad (4.79)$$

in which  $\bar{q} = k_0 \sqrt{\epsilon}$  designates the reference wave number. The perturbation operator  $\mathcal{P}_{\mathbf{H}}$  is given by

$$\mathcal{P}_{\mathbf{H}} = \frac{1}{\bar{q}^2} [\Delta_t + k_0^2 \Delta \epsilon + \nabla_t (\ln \epsilon) \times \nabla_t \times]. \quad (4.80)$$

As for the scalar forward Helmholtz equations, the paraxial approximation can be optimized according to the method described in Section 4.1.3. The formulation of the vectorial BPM in terms of the transverse field components avoids non-physical solutions such as magnetic monopoles ( $\nabla \cdot \mathbf{H} \neq 0$ ). If required, the longitudinal field components can easily be calculated a posteriori using Maxwell’s equation

$$\nabla \cdot \mathbf{H} = 0.$$

We then obtain

$$H_z(z + \Delta z) = H_z(z) + \int_z^{z+\Delta z} dz \nabla_t \cdot \mathbf{H}_t. \quad (4.81)$$

Equation (4.79) can be treated easily using one of the finite difference approaches shown in Section 4.4.2. We should note here that the stencil of the vectorial equation will comprise ten grid points instead of five for the scalar forward Helmholtz equation (see Figure 4.13).

#### 4.6 BIDIRECTIONAL BEAM PROPAGATION

The beam propagation method as described so far separates the waves traveling forward from those running backward. On the basis of this strategy, we could

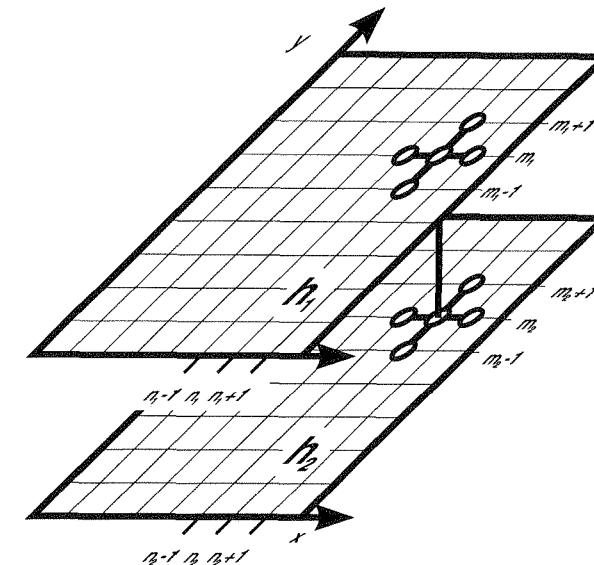


Figure 4.13. Finite difference stencil for the vectorial BPM.

construct two independent initial value problems for the waves traveling in the  $\pm z$ -direction. Obviously, this attempt will fail if abrupt changes of the refractive index profile along the direction of propagation were to act as the source of reflections. We will now discuss extensions of the BPM which allow us to handle reflections.

The general strategy of such algorithms is shown in Figure 4.14. The arrows in this diagram stand for the partial waves within the integrated optical device traveling into either the  $+z$ - or  $-z$ -direction. Their length accounts for the optical power stored in the corresponding partial wave. The incident field given by the lower left arrow is propagated (without taking reflections into account) to the first interface at  $z = z_1$ . At the interface, the input field  $\Psi^i(\mathbf{r}_t, z)$  is divided into a reflected ( $r$ ) and a transmitted field ( $t$ ) such that

$$\Psi^t(\mathbf{r}_t, z) = \Psi^i(\mathbf{r}_t, z) + \Psi^r(\mathbf{r}_t, z).$$

The transmitted part of the field then travels in the  $+z$ -direction until the next interface occurs, and the reflected part is propagated backwards to the plane of incidence. At the next interface, the same scenario takes place, and so on. The whole procedure is stopped if the optical power stored in the new partial waves falls below a predefined cutoff. The total transmitted field is given by the superposition of all partial waves traveling in the  $+z$ -direction, and the total reflected field is found by the superposition of all reflected contributions. The superposition of partial waves corresponds to Airy’s summation (see Section 2.2.3). Roughly speaking, the bidirectional BPM is realized by the conjunction of a unidirectional BPM and an algorithm

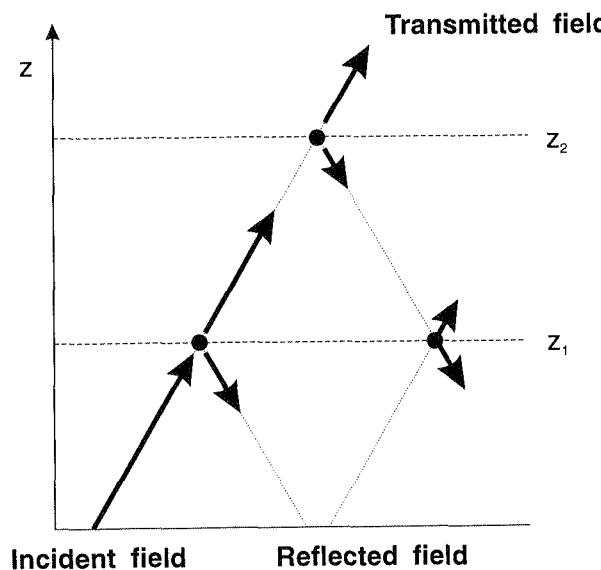


Figure 4.14. Strategy of the bidirectional beam propagation method.

for the treatment of reflections at the internal interfaces. The bidirectional BPM can thus be combined with any of the BPM algorithms described previously.

We will now derive the operator governing the reflection of a beam at a single dielectric interface. For this purpose, we will consider the reflection operator<sup>12</sup>

$$\mathcal{R} = \frac{\sqrt{\mathcal{H}_{\text{in}}} - \sqrt{\mathcal{H}_{\text{out}}}}{\sqrt{\mathcal{H}_{\text{in}}} + \sqrt{\mathcal{H}_{\text{out}}}} \quad (4.82)$$

describing the reflection coefficient of a TE polarized beam in a weakly guiding waveguide at the interface between two media “in” and “out”.

The reflection operator is a straightforward generalization of the Fresnel formula for plane waves (see Equation (2.46)). Using the wave numbers  $q_l(\mathbf{r}_t) = k_0 \sqrt{\epsilon_l(\mathbf{r}_t)}$  and  $\bar{q}_l = k_0 \sqrt{\bar{\epsilon}_l}$  with the index  $l = \text{in}, \text{out}$  we obtain

$$k_0 \sqrt{\mathcal{H}_l} = \sqrt{\nabla_t^2 + q_l(\mathbf{r}_t)^2}$$

12. Here and in the following, operator fractions must be read as

$$\frac{\mathcal{A}}{\mathcal{B}} = \mathcal{A}\mathcal{B}^{-1}.$$

$$\approx \sqrt{\mathcal{D}_l} + q_l(\mathbf{r}_t) - \bar{q}_l$$

for the square root of the Hamiltonian where  $\mathcal{D}_l = \sqrt{\nabla_t^2 + \bar{q}_l^2}$  is the diffraction operator acting in a uniform medium with a dielectric constant  $\bar{\epsilon}$ . If we approximate further

$$\begin{aligned} & k_0 \left( \sqrt{\mathcal{H}_{\text{in}}} \pm \sqrt{\mathcal{H}_{\text{out}}} \right) \\ &= \left( \sqrt{\mathcal{D}_{\text{in}}} \pm \sqrt{\mathcal{D}_{\text{out}}} \right) \left( \mathcal{E} + \frac{q_{\text{in}}(\mathbf{r}_t) \pm q_{\text{out}}(\mathbf{r}_t) - (\bar{q}_{\text{in}} \pm \bar{q}_{\text{out}})}{\sqrt{\mathcal{D}_{\text{in}}} \pm \sqrt{\mathcal{D}_{\text{out}}}} \right) \\ &\approx \frac{\left( \sqrt{\mathcal{D}_{\text{in}}} \pm \sqrt{\mathcal{D}_{\text{out}}} \right) \left( q_{\text{in}}(\mathbf{r}_t) \pm q_{\text{out}}(\mathbf{r}_t) \right)}{\bar{q}_{\text{in}} \pm \bar{q}_{\text{out}}} \end{aligned}$$

we arrive at a more convenient formulation

$$\mathcal{R} \approx \frac{\left( \sqrt{\mathcal{D}_{\text{in}}} - \sqrt{\mathcal{D}_{\text{out}}} \right) \left( q_{\text{in}}(\mathbf{r}_t) - q_{\text{out}}(\mathbf{r}_t) \right) \left( \bar{q}_{\text{in}} + \bar{q}_{\text{out}} \right)}{\left( \sqrt{\mathcal{D}_{\text{in}}} + \sqrt{\mathcal{D}_{\text{out}}} \right) \left( q_{\text{in}}(\mathbf{r}_t) + q_{\text{out}}(\mathbf{r}_t) \right) \left( \bar{q}_{\text{in}} - \bar{q}_{\text{out}} \right)} \quad (4.83)$$

of the reflection operator. If we also neglect any contributions caused by the non-commutativity of the operators, we can divide the reflection operator into two operators. One

$$r(\mathbf{q}_t, z) = \frac{\sqrt{\bar{q}_{\text{in}}^2 - q_t^2} - \sqrt{\bar{q}_{\text{out}}^2 - q_t^2}}{\sqrt{\bar{q}_{\text{in}}^2 - q_t^2} + \sqrt{\bar{q}_{\text{out}}^2 - q_t^2}} \quad (4.84)$$

comprises all diffraction terms. It is diagonal in the Fourier space. The second one

$$r_I(\mathbf{r}_t, z) = \frac{q_{\text{in}}(\mathbf{r}_t) - q_{\text{out}}(\mathbf{r}_t)}{q_{\text{in}}(\mathbf{r}_t) + q_{\text{out}}(\mathbf{r}_t)} \frac{\bar{q}_{\text{in}} + \bar{q}_{\text{out}}}{\bar{q}_{\text{in}} - \bar{q}_{\text{out}}} \quad (4.85)$$

contains the local wave numbers. It is diagonal in the position space. Using this representation we can easily calculate the reflected field

$$\Psi^r(\mathbf{r}_t, z) = r_I(\mathbf{r}_t, z) \mathcal{F}^{-1} [r(\mathbf{q}_t, z) \mathcal{F} [\Psi^i(\mathbf{r}_t, z)]] \quad (4.86)$$

by a sequence of Fourier transformations and multiplications. A detailed description of an analogous procedure within the framework of the classical BPM is found in Section 4.2.

## 4.7 APPLICATION OF THE BPM

We have so far discussed the philosophy of the beam propagation method and the advantages and disadvantages of some particular algorithms which are suited for beam propagation. We will now turn to some topics dealing with the application of the beam propagation method and with the appropriate choice of algorithms.

### 4.7.1 3D/2D Beam Propagation Method

With a few exceptions during the derivation of the BPM algorithms we did not distinguish between the (full) 3D-BPM and restricted versions which operate on one fewer dimension, i.e., on the  $xz$ -plane. However, as regards the corresponding requirements on computer hardware, this choice is of some importance. To obtain some idea of magnitudes involved in our discussion let us consider a 2D-BPM showing a time complexity  $\mathcal{O}_{2D}$  in comparison to an equivalent 3D-BPM with a time complexity  $\mathcal{O}_{3D}$ . If we assume that both algorithms operate on a regular grating, i.e., all algorithms discussed here except the FE-BPM, we obtain the relation

$$\frac{\mathcal{O}_{3D}}{\mathcal{O}_{2D}} \geq n_y \quad (4.87)$$

The equality in Equation (4.87) corresponds to a linear time complexity which is not attained for any of the algorithms discussed here. This means that *the CPU-time of a 3D-BPM simulation is at least that of  $n_y$  2D-BPM simulations of the same problem*. Thus, it is strictly recommended to use the 2D-BPM whenever possible. Furthermore, it becomes clear why algorithms operating on an adaptive mesh are such promising candidates for the 3D-BPM even if they are highly sophisticated.

From a purely theoretical point of view, the 2D-BPM can be applied only to beam propagation in slab waveguide structures provided that the beam is homogeneous in one transverse direction. However, no interesting applications of this type are known. For practical applications the 2D-BPM is used in another sense. Analogously to the effective index method (EIM) which was introduced as a method to calculate the eigenmodes of weakly guiding structures in Section 3.9.2, the 2D-BPM operates on a 2D effective dielectric profile which is a projection of the entire 3D device. We usually carry out this projection using the effective index method (EIM), i.e., utilizing the effective index of the stack of layers under a given point. Sometimes, it may be more convenient to use a predefined refractive index profile and to adapt some parameters of the function so that the field radius, the  $V$  parameter or other interesting physical quantity is adapted to the full three-dimensional problem. At this point of our discussion it looks as if the usage of the 2D-BPM and that of the effective index method were governed by the same conditions as regards the refractive index profile. Unfortunately, the usage of the 2D-BPM is more restricted, since radiation into two dimensions greatly differs from that into three dimensions. However, this restriction is not severe, since most of the devices of interest for integrated optics are not radiation driven. At the end of this brief discussion we should keep in mind that

- the 2D-BPM can be applied to diffused and other weakly guiding structures as long as a low level of radiation is guaranteed,
- the 3D-BPM must be used especially for devices whose operation is influenced by the chip surface and/or by radiation,

- the vectorial BPM must be used for devices whose operation is essentially driven by the chip surface.

### 4.7.2 Benchmark Results

In this chapter we presented various algorithms which allow us to simulate the propagation of beams in a half-space. At its end we will briefly examine the results of a benchmark test which was initiated at the ECOC workshop on Optical Waveguide Modelling, Numerical Simulation and Theory in Teupitz (Germany) in 1992 [8]. Driven by the current availability of software, the tests were restricted to 2D-BPM algorithms, i.e., to beam propagation starting from a one-dimensional cross section. The following performance parameters of a BPM algorithm should be assessed from the user's point of view.

#### • Accuracy

Since the beam propagation method is continuously evolving into an increasingly realistic design tool for integrated optical circuits, the accuracy of BPM algorithms is of great importance. To provide an absolute assessment of the accuracy of the algorithms, benchmark problems should have an analytical or at least a quasi analytical solution.

#### • Robustness

BPM algorithms should be stable and reliable. New algorithms are usually tested for these qualities by propagating an eigenmode along a straight waveguide and verifying the stability of the field distribution. Obviously, this test represents a necessary but by no means a sufficient condition for stability. The robustness can be much better judged by trying them on a sequence of increasingly critical benchmark problems.

#### • Universality

A BPM software package should be able to tackle a large variety of physical problems irrespective of the underlying geometry and refractive index profile. The smoothing of the original refractive index profile in order to simplify the numerical handling often applied within "classical" BPM strictly speaking violates the requirement of universality.

#### • Effectivity

High space and/or time complexity also represents a potential drawback of BPM algorithms, especially if the software operates close to the limits of the available computer hardware.

In the following, we will examine a part of the "tilted waveguide test", one of the benchmark problems formulated in [8] in more detail. Let us first discuss

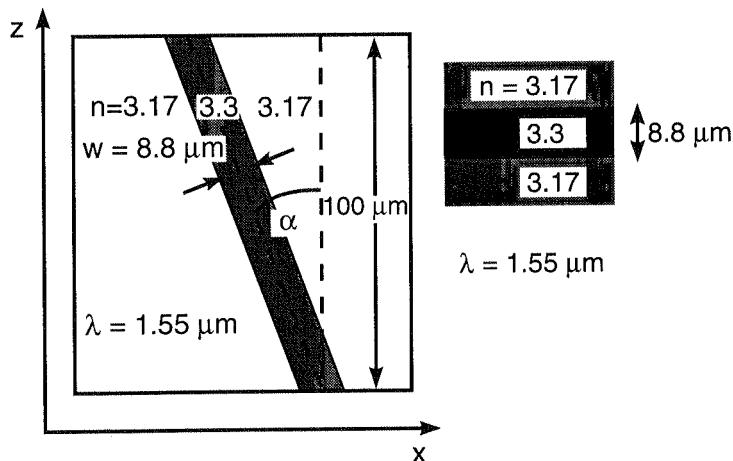


Figure 4.15. Geometry and refractive index profiles for the “tilted waveguide test”.

the underlying waveguide structure (see Figure 4.15). It consists of a slab multimode step-index waveguide,  $8.8 \mu\text{m}$  in thickness, exhibiting a refractive index step of  $\Delta n = 0.13$ . To demonstrate the physical relevance of the test structure, InP was chosen as the background material. At a wavelength of  $\lambda = 1.55 \mu\text{m}$  this waveguide supports 11 guided modes for both TE- and TM-polarization. The benchmark problem consists of a sequence of simulations in which the fundamental, fifth or tenth order mode is to be propagated through waveguides having tilt angles between  $0^\circ$  and  $20^\circ$  with respect to the  $z$ -axis of the computational window. For each of the examples, the numerical power transfer, i.e., the unexpected power transfer to any other guided or radiation modes, was calculated. Obviously, the “tilted waveguide test” has the advantage of being rather simple and of possessing an exact solution; namely, no power transfer to any other mode.

This benchmark problem was intended to assess the following features of a BPM algorithm by means of a physical example:

- transport of “radiation like” tenth order guided modes,
- wide angle behavior,
- influence of transverse and longitudinal discretization.

Unfortunately, it turned out that the test was insufficiently critical to characterize the wide angle behavior; i.e., paraxial algorithms could master the propagation through the most tilted waveguide by choosing an optimized reference dielectric constant (see Section 4.1.3).

Table 4.1 lists the BPM programs which have been tested by the “tilted waveguide” benchmark test. The second row shows the underlying algorithms. The abbreviations used within this row are

**Table 4.1**  
Description of the BPM programs and longitudinal and transverse discretization required for the tilted waveguide test.

<b>Program</b>	<b>Algorithm</b>	<b>TE<sub>0</sub> / <math>\alpha=0^\circ</math></b>		<b>TE<sub>10</sub> / <math>\alpha=20^\circ</math></b>	
		$N_z$	$\langle N_x \rangle$	$N_z$	$\langle N_x \rangle$
FTBPM	FFT	100	256	1000	256
Praha-FT	FFT	100	512	2500	4096
LETI-FD	FD	200	1024	200	1024
TUD-FD	CN-FD	1000	1100	1000	1100
FD0BPM	CN-FD	100	1701	1000	1701
UP-FD	CN-FD	250	401	2000	1201
UK-FD	CN-FD	1000	1401	1000	1401
TUB-SE	SE-FD	100	1401	1000	1401
UJ-SE	SE-FD	1000	1024	1000	1024
HHI-FE/FD	FE/CN-FD	500	1024	2000	1024
AMIGO	FE/FE	21	42	1429	1311
SNL-Pade	WA-FD	1000	1501	3421	6001
FD2BPM	WA-CN-FD	100	512	1000	2048

FFT - classical BPM (see Section 4.2)

FD - explicit finite differences (see Section 4.4.2)

CN-FD - implicit finite differences (see Section 4.4.2)

SE-FD - higher order explicit finite differences (see Section 4.4.2)

FE/FD - finite elements/finite differences (see Section 4.4.3)

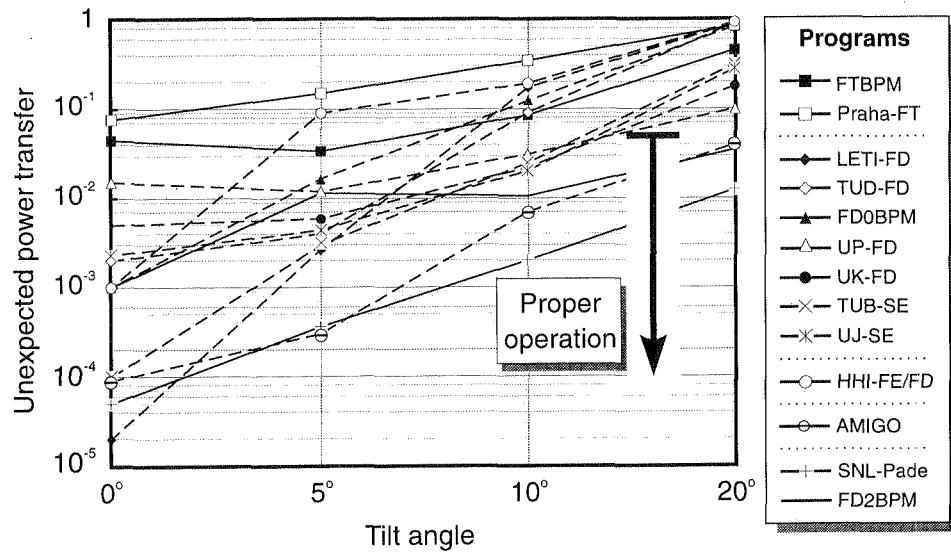
FE/FE - adaptive finite elements (see Section 4.4.3)

WA-FD - wide angle finite differences

WA-CN-FD - wide angle implicit finite differences

Table 4.1 also shows the mean number of transverse discretization points  $\langle N_x \rangle$  and the number of propagation steps  $N_z$  in order to give a rough idea of the effort that was spent to obtain the benchmark results.

All BPM programs mastered the propagation of the fundamental mode along the untilted waveguide. For the fundamental mode propagating through the untilted waveguide ( $\alpha = 0^\circ$ ), the unexpected power transfer, i.e., the relative power



**Figure 4.16.** Unexpected power transfer to any other guided or radiation modes vs. tilt angle for a tilted waveguide excited by a tenth order guided mode ( $TE_{10}$ ). Solid lines: Classical BPM and wide angle approximations. Dashed lines: Paraxial approximations.

transfer to any other guided or radiation mode caused by unexpected instability, was  $< -30$  dB for all BPM programs. In contrast, the propagation of the tenth order mode resulted for most of the algorithms in severe problems which got worse with increasing tilt angle. We see that only three candidates were able to propagate the eigenmode for  $\alpha = 20^\circ$  with an acceptable accuracy (unexpected power transfer  $< -15$  dB), good results (power transfer  $< -20$  dB) could be obtained only once. Most software packages failed to master this benchmark test (power transfer  $> -3$  dB). Figure 4.16 illustrates the evolution of the numerical power transfer with an increasing tilt angle if the waveguide is stimulated by the tenth order mode. The results, which are based on the paraxial Helmholtz equation, are indicated by dashed lines, the wide angle approximations and the classical BPM are indicated by solid lines. As expected, we observe an increase of the unexpected power transfer with growing tilt angle.<sup>13</sup> We see furthermore that the classical BPM does not offer any advantage over the paraxial algorithms. The wide angle approximations, however, seem to provide significant advantages with comparison to the algorithms based on the paraxial Helmholtz equation. The adaptive finite element program AMIGO delivered excellent results despite being based on a paraxial algorithm due to the use of an optimized reference dielectric constant  $\bar{\epsilon}$  (see Section 4.1.3).

13. The scattering of the results for significant power transfers ( $> -10$  dB) indicates the breakdown of the corresponding BPM algorithms.

## REFERENCES

- [1] Papoulis, A., *Systems and Transforms with Application in Optics*, New York, McGraw-Hill, 1968.
- [2] Schmidt, F., and Deufhard, P., "Discrete Transparent Boundary Conditions for Fresnel's Equation", *Proc. Integrated Photonics Research Conference (IPR)*, Vol. 3, 1994, pp. 45–47.
- [3] Van Roey, J., van der Donk, J., and Lagasse, P. E., "Beam Propagation: Analysis and Assessment", *Journal of the Optical Society of America*, Vol. 71, 1981, pp. 803–810.
- [4] Clearbourn, J. F., *Fundamentals of Geophysical Data Processing*, New York, McGraw-Hill, 1976.
- [5] Hadley, G. R., "Multistep Method for Wide Angle Beam Propagation", *Optics Letters*, Vol. 17, 1992, pp. 1743–1745.
- [6] Splett, A., Majd, M. and Petermann, K., "A Novel Beam Propagation Method for Large Refractive Index Steps and Large Propagation Distances", *IEEE Photonics Technology Letters*, Vol. 3, 1991, pp. 466–468.
- [7] Schmidt, F., "An Adaptive Approach to the Numerical Solution of Fresnel's Wave Equation", *IEEE Journal of Lightwave Technology*, Vol. 11, 1993, pp. 1425–1434.
- [8] Nolting, H.-P., and März, R., "Results of Benchmark Tests for Different Numerical BPM Algorithms", in *Linear and Nonlinear Optics*, Proceedings of the SPIE and EOS, Vol. 2212, 1994.
- [9] Feit, M. D. and Fleck, J. A., Jr., "Light Propagation in Graded-Index Optical Fibres", *Applied Optics*, Vol. 17, 1978, pp. 3990–3998.
- [10] Thylen, L., and Yevick, D., "Beam Propagation Method in Anisotropic Media", *Applied Optics*, Vol. 21, 1982, pp. 2751–2754.
- [11] Koch, T. B., Davies, J. B., Fernandez, F. A., and März, R., "Computation of Wave Propagation in Integrated Optical Devices using z-Transient Variational Principle", *IEEE Transactions on Magnetics*, Vol. 27, 1991, pp. 3876–3879.
- [12] Yevick, D., and Hermansson, B., "Efficient Beam Propagation Techniques", *IEEE Journal of Quantum Electronics*, Vol. 26, 1990, pp. 109–112.
- [13] Yevick, D., Bardyszewski, W., Hermansson, B., and Glasner, M., "Split-Operator Electric Field Reflection Techniques", *IEEE Photonics Technology Letters*, Vol. 3, 1991, pp. 527–529.
- [14] Pregla, R., "MoL BPM – Method of Lines Based Beam Propagation Method", in *Methods for Modeling and Simulation of Optical Guided-Wave Devices*, Amsterdam, Elsevier, 1995.
- [15] KaczmarSKI, P., and Lagasse, P. E., "Bidirectional Beam Propagation Method", *Electronics Letters*, Vol. 24, 1988, pp. 675–676.
- [16] Hadley, G. R., "Transparent Boundary Condition for Beam Propagation", *Optics Letters*, Vol. 16, 1991, pp. 624–626.
- [17] Huang, W., Chenglin X., Chu S.-T., and Chaudhuri, S. K., "The Finite-Difference Vector Beam Propagation Method: Analysis and Assessment", *IEEE Journal of Lightwave Technology*, Vol. 10, 1992, pp. 295–305.

## Chapter 5

### Mode Conversion

Although integrated optics aims at handling light by means of waveguides, one of its crucial problems is to provide efficient interfaces between waveguides, i.e., to transport the optical power efficiently from one waveguide to another. The most obvious and prominent example of such an interface is the fiber-to-chip coupling, but there are also a couple of internal interfaces such as laser-to-waveguide or waveguide-to-waveguide interfaces and on-chip optical systems such as geodesic lenses or focusing spectrographs.

This chapter covers the three most important techniques of interfacing: butt coupling, imaging by optical systems and adiabatic mode conversion by optical tapers. It starts with a discussion of the Hermite-Gaussian beams which act as a model for beams from a waveguide or fiber endface. On the basis of this model, we will derive analytical expressions for the coupling loss caused by different types of misalignment. We will continue our discussion with an outline of Gaussian optics, i.e., with the theory of imaging Hermite-Gaussian beams by optical systems. This chapter will be concluded by a discussion of optical tapers. Within the framework of this discussion we will establish the local normal mode theory and derive some fundamental remarks on tapering from it.

#### 5.1 HERMITE-GAUSSIAN BEAMS

In this section we will study Hermite-Gaussian beams, i.e., the propagation of Hermite-Gaussians in a homogeneous half-space. We will do so for two reasons. Firstly, Hermite-Gaussians are the eigenmodes of the parabolic waveguides (see Section 3.7). This means they can act as equivalent waveguides to describe beams emitted from the endface of an integrated optical waveguide. Secondly, it will become apparent in this section that Hermite-Gaussian beams offer a closed solution of the forward Helmholtz equation (see Section 4.2), i.e., of the paraxial beam propagation problem.

### 5.1.1 Evolution of Hermite-Gaussian Beams

Within the framework of the paraxial approximation, the evolution of an Hermite-Gaussian beam is given by

$$\phi^{(m)}(x, z) = N_m(z) H_m \left( \frac{\sqrt{2}x}{r_0(z)} \right) \exp \left( i \frac{k_0 n}{2\zeta} x^2 \right) \quad (5.1)$$

with the normalization constant

$$N_m(z) = \left( 2^{m-1/2} r_0(z) m! \sqrt{\pi} \right)^{-1/2} e^{-i(2m+1)\varphi(z)/2}.$$

The physical parameters characterizing the evolution of the Hermite-Gaussian beam are the complex beam parameter

$$\zeta = z - iz_c$$

and the confocal length

$$z_c = \frac{k_0 n r_0^2}{2},$$

the phase shift

$$\varphi(z) = \tan^{-1} \left( \frac{z}{z_c} \right)$$

and the local beam radius

$$r_0(z) = r_0 \sqrt{1 + \tan^2 \varphi(z)}.$$

Relating all coordinates to the beam radius  $r_0$  in the beam waist we see that *the relative complex beam parameter  $\zeta/z_c$  completely determines the evolution of the Hermite-Gaussian beam*. Obviously, this statement also holds for two-dimensional Hermite-Gaussian beams

$$\phi^{(mn)}(x, y, z) = \phi^{(m)}(x, z) \phi^{(n)}(y, z) \quad (5.2)$$

provided they are described by a single relative complex beam parameter  $\zeta/z_c$ , i.e., if they exhibit rotational symmetry.

For the derivation of Equation (5.1) we use the technique of generating functions. A generating function of a set of polynomials is defined by a power series over the underlying polynomials. It is often (and also in our case) much easier to obtain results by treating the generating function rather than the underlying polynomials. The desired expressions for the polynomials are then derived by term-by-term identification of the final power series.

The generating function of the Hermite polynomials is given by

$$F(t, \xi) = e^{-t^2+2t\xi} = \sum_{m=0}^{\infty} \frac{t^m}{m!} H_m(\xi). \quad (5.3)$$

For a proof of Equation (5.3) we regard the polynomials  $H_m(\xi)$  as unknown functions. We differentiate Equation (5.3) with respect to the variable  $\xi$  and obtain

$$\sum_{m=1}^{\infty} \frac{t^m}{m!} \left( \frac{\partial^2 H_m(\xi)}{\partial \xi^2} - 2m H_{m-1}(\xi) \right) = 0. \quad (5.4)$$

By differentiating the same equation with respect to the second variable  $t$  and inserting the result (5.4) we obtain

$$\sum_{m=1}^{\infty} \frac{t^m}{m!} [H_{m+1}(\xi) - 2\xi H_m(\xi) + 2m H_{m-1}(\xi)] = 0. \quad (5.5)$$

We can conclude that the (currently unknown) functions  $H_m(\xi)$  must satisfy the recurrence formulas (3.86) and (3.87) of the Hermite polynomials. The direct calculation of the first two representatives of the series on the right-hand side of Equation (5.3) yields  $H_0(\xi) = 1$  and  $H_1(\xi) = 2\xi$ , i.e., the first two Hermite polynomials.

In the following, we will use the generating function  $F(t, \xi)$  to derive the fundamental properties – Fourier transform and Fresnel’s diffraction integral – of Hermite-Gaussian beams. Within the framework of these calculations we will often switch to the dimensionless coordinates

$$\xi = \frac{\sqrt{2}x}{r_0} \quad (5.6)$$

in position space and

$$\eta = \frac{q_x r_0}{\sqrt{2}} \quad (5.7)$$

in Fourier space. In our further discussion,  $\phi^{(m)}(x, z)$  designates a Hermite-Gaussian beam. To avoid clumsy notation we suppress the variable  $z$  at the beam waist ( $z = 0$ ), i.e.,

$$\phi^{(m)}(x, 0) = \phi^{(m)}(x) = N_m H_m(\sqrt{2}x/r_0) e^{-x^2/r_0^2}. \quad (5.8)$$

The corresponding notation is applied in Fourier space.

At the beginning of our discussion we normalize the Hermite-Gaussian functions. By inserting the expression corresponding to the normalization integral we obtain

$$\begin{aligned} & \sum_{m,n=0}^{\infty} \frac{s^m t^n}{m! n!} \int_{-\infty}^{\infty} dx \phi^{(m)}(x) \phi^{(n)}(x) \\ &= N_m N_n \frac{r_0}{\sqrt{2}} \int_{-\infty}^{\infty} d\xi e^{-s^2-t^2+2\xi(s+t)-\xi^2} \\ &= N_m N_n r_0 \sqrt{\frac{\pi}{2}} e^{2st} \\ &= \sum_{m,n=0}^{\infty} \frac{s^m t^n}{m! n!} \delta_{mn} N_m^2 2^{m-1/2} r_0 m! \sqrt{\pi}. \end{aligned}$$

We see that the normalization constant is given by

$$N_m = \left( 2^{m-1/2} r_0 m! \sqrt{\pi} \right)^{-1/2}.$$

Incidentally, we have rederived the orthogonality relations

$$\int_{-\infty}^{\infty} dx \phi^{(m)}(x) \phi^{(n)}(x) = \delta_{mn}$$

of the Hermite-Gaussian functions.

In the next step, we derive an expression for the Fourier transform of a Hermite-Gaussian function. We obtain

$$\begin{aligned} & \sum_{m=0}^{\infty} \frac{t^m}{m!} \phi^{(m)}(q_x) \\ &= \sum_{m=0}^{\infty} \frac{t^m}{m!} \mathcal{F}[\phi^{(m)}(x)] \\ &= N_m \frac{r_0}{\sqrt{4\pi}} \int_{-\infty}^{\infty} d\xi e^{-t^2 + 2\xi t - \xi^2/2 - i\eta\xi} \\ &= N_m \frac{r_0}{\sqrt{2}} e^{t^2 - 2i\eta t - \eta^2/2} \\ &= \sum_{m=0}^{\infty} \frac{t^m}{m!} (-i)^m N_m \frac{r_0}{\sqrt{2}} H_m(q_x r_0 / \sqrt{2}) e^{-(q_x r_0 / 2)^2}, \end{aligned}$$

and we see that the Fourier transform of a Hermite-Gaussian function

$$\phi^{(m)}(q_x) = (-i)^m N_m \frac{r_0}{\sqrt{2}} H_m\left(\frac{q_x r_0}{\sqrt{2}}\right) e^{-(q_x r_0 / 2)^2} \quad (5.9)$$

is again a Hermite-Gaussian function.

In the last step we derive an analytic expression for the spread of a Hermite-Gaussian beam in a homogeneous half-space with a refractive index  $n$ . We have shown in Section 4.2 that the evolution of the amplitude is given within the framework of the paraxial Helmholtz equation by

$$\phi^{(m)}(x, z) = \mathcal{F}^{-1}\left[\exp\left(-i \frac{q_x^2}{2k_0 n} z\right) \phi^{(m)}(q_x, 0)\right].$$

We further saw in Section 4.2 that this solution corresponds to calculating the Fresnel diffraction integral. The evolution of the amplitude is then given by

$$\sum_{m=0}^{\infty} \frac{t^m}{m!} \phi^{(m)}(x, z)$$

$$\begin{aligned} &= \sum_{m=0}^{\infty} \frac{t^m}{m!} \mathcal{F}^{-1}\left[\exp\left(-i \frac{q_x^2}{2k_0 n} z\right) \phi^{(m)}(q_x, 0)\right] \\ &= N_m \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dt e^{t^2 - 2i\eta t - \eta^2(1+i\tan\varphi(z))/2 - i\eta\xi} \\ &= N_m e^{-i\varphi(z)/2} \frac{r_0}{r_0(z)} \exp\left(t^2 - 2 \frac{(t - \xi/2)^2}{1 + i\tan\varphi(z)}\right) \\ &= \sum_{m=0}^{\infty} \frac{t^m}{m!} N_m(z) H_m\left(\frac{\sqrt{2}x}{r_0(z)}\right) \exp\left(i \frac{k_0 n}{2\zeta} x^2\right). \end{aligned}$$

The evolution of the Hermite-Gaussian beam in a homogeneous half-space is obviously described by Equation (5.1).

### 5.1.2 Physical Properties of the Hermite-Gaussian Beam

For a brief discussion of the physical properties of the Hermite-Gaussian beams we separate in Equation (5.1) the terms describing the intensity distribution from those describing the phase portrait by using

$$\frac{1}{\zeta} = \frac{1}{R(z)} + i \frac{2}{k_0 n r_0(z)^2}. \quad (5.10)$$

In the far-field region, the parameter

$$R(z) = \frac{|\zeta|^2}{z} \quad (5.11)$$

will turn out to be the radius of curvature of the phase fronts at the optical axis. Inserting Equation (5.10) into (5.1) we obtain

$$\begin{aligned} \phi^{(m)}(x, z) &= |N_m(z)| H_m\left(\frac{\sqrt{2}x}{r_0(z)}\right) \exp\left(-\frac{x^2}{r_0(z)^2}\right) \\ &\times \exp\left(i \frac{k_0 n}{2R(z)} x^2 - (2m+1)\varphi(z)/2\right). \end{aligned} \quad (5.12)$$

By using Equation (5.12) we see that the local beam radius – in precise terms the  $1/e$ -radius of the envelope term  $\exp(-x^2/r_0(z)^2)$  – is given by  $r_0(z)$ . In the far-field region, the beam contour

$$r_0(z) \approx z \tan \theta \quad (5.13)$$

looks like a cone with a diffraction angle

$$\theta = \tan^{-1}\left(\frac{2}{k_0 n r_0}\right). \quad (5.14)$$

We see that the diffraction angle increases with decreasing spot size. This behavior is typical for near-fields of any shape, and Equation (5.14) shows the simple relationship between both quantities for the Hermite-Gaussian beam. The point where the beam conus of Equation (5.13) has the same width as the near-field can be regarded as the transition from near-field to far-field. This point of transition is described by the confocal length  $z_c$ .

To calculate the phase fronts of the beam, we must consider the optical field (instead of its amplitude). The loci of constant phase are obviously the solutions of the implicit equation

$$k_0 n z + \frac{k_0 n x^2}{2R(z)} - \frac{(2m+1)\varphi(z)}{2} = k_0 n z_0 - \frac{(2m+1)\varphi(z_0)}{2}.$$

where the constant on the right-hand side has been chosen such that  $z(0) = z_0$ . By differentiating this implicit equation twice with respect to the variable  $x$ , we obtain at the optical axis ( $x = 0$ )

$$\left. \frac{d^2 z}{dx^2} \right|_{x=0} = -\frac{1}{R(z_0)} \frac{1}{1 - (2m+1)r_0 \tan \theta / |4\zeta(z_0)|}. \quad (5.15)$$

For small  $m$ , i.e., for low-order Hermite-Gaussian beams, the second term on the right-hand side of Equation (5.15) plays a minor role. Thus, the parameter  $R(z_0)$  describes the local radius of curvature of the phase fronts at  $z = z_0$  as a good approximation. At the optical axis, the Hermite-Gaussian beam can then be regarded as a spherical wave which was emitted at

$$z_0 - R(z_0) \approx -\frac{r_0^2}{z_0 \tan \theta}. \quad (5.16)$$

We see that the effective source point of the Hermite-Gaussian beam, i.e., the source point of the local spherical wave, moves during beam propagation. At the beam waist ( $z_0 \rightarrow 0$ ) it is located at infinity ( $z_0 - R(z_0) \rightarrow -\infty$ ). In the far-field ( $z \rightarrow \infty$ ) the effective source point is located at the beam waist ( $R(z_0) = z_0$ ).

Figure 5.1 summarizes the results of our brief discussion.

## 5.2 COUPLING OF GAUSSIAN BEAMS

To calculate the coupling losses between integrated optical chips and fibers we must usually match the eigenmodes of different types of waveguide. The most obvious example of this kind is butt coupling, where the eigenmodes are matched without any optical system. Other examples are fiber-to-chip coupling via lenses or lensed fibers, on-chip waveguide-to-waveguide interfaces and the matching of diffracted fields inside a planar spectrograph (see Chapter 8). We will concentrate in this section on calculating of coupling losses due to mode mismatch. For the

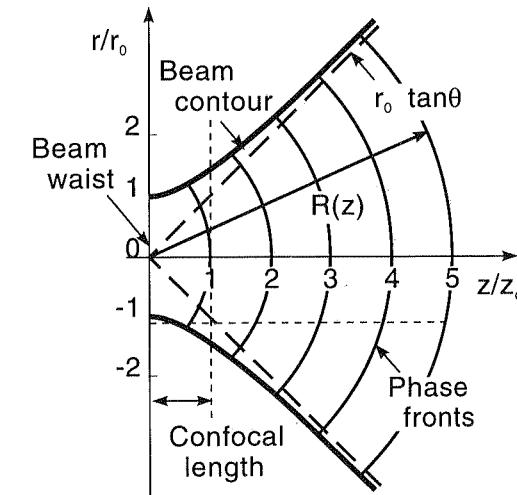


Figure 5.1. Physical properties of the Hermite-Gaussian beam.

sake of clarity we will formulate the alignment problem using the language of fiber-to-chip coupling. The application of the results to the other cases just listed is straightforward.

In our further discussion, we will assume that the eigenmodes of both fiber and chip exhibit a Gaussian shape. Secondly, we will assume an aberration-free optical system which maps a magnified image  $\Phi_i$  of the chip endface onto a plane in front of the fiber endface. The spot magnification factor of the optical system is

$$m = r_0^{(i)} / r_0^{(c)},$$

where the parameters  $r_0^{(c)}$  and  $r_0^{(i)}$  designate the field radii ( $1/e$ -radii of the optical field) of the eigenmode  $\Phi_c$  at the chip endface and of its image in front of the fiber endface. For a lossless fiber-to-chip coupling, the image  $\Phi_i$  of the chip near-field and the near-field  $\Phi_f$  must match with respect to field radius and position, and the optical axes of both beams must coincide. In physical terms, *we will discuss the coupling loss caused by mismatch of the fundamental modes of two parabolic waveguides*. Although this model suffers from a considerable number of restrictions it is a workhorse for the treatment of misalignments in integrated optics.

The coupling loss<sup>1</sup> caused by the misalignment of the two beams is then given by

$$\gamma = \gamma_x + \gamma_y \quad (5.17)$$

where each of the contributions

$$\gamma_s = -10 \log |\langle \Phi(\mu_s r_0, \Delta\alpha_s, \Delta s, \Delta z) | \Phi(r_0, 0, 0, 0) \rangle|^2 \quad (5.18)$$

<sup>1</sup> The coupling loss is given in decibels (dB) related to the optical power.

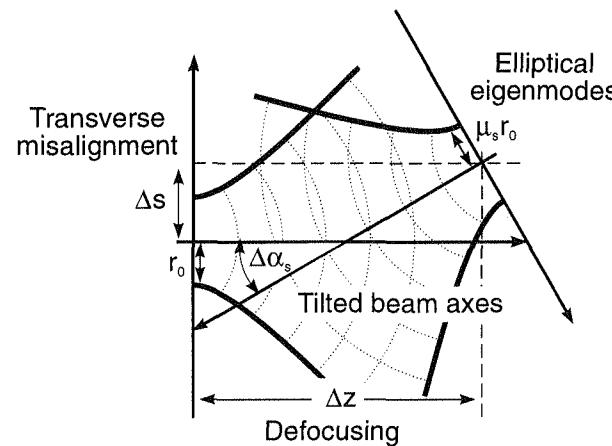


Figure 5.2. Different types of misalignment.

designates the coupling loss with respect to one of the Cartesian coordinate axes ( $s = x, y$ ). The function  $\Phi(\mu_s r_0, \Delta\alpha, \Delta s, \Delta z)$  describes the (rapidly varying) field of a Gaussian beam. The parameter

$$\mu_s = \frac{m}{m_{\text{req}}^{(s)}}$$

designates the ratio between the spot magnification  $m$  of the optical system and the spot magnification  $m_{\text{req}}^{(s)}$  which is required to map the half-axes  $s$  of both Gaussian beams onto each other. Obviously, the best mapping of a set of half-axes  $s$  is obtained for  $\mu_s = 1$ .

Figure 5.2 shows the configuration for a misaligned coupling. The intermediate image and the endface of a chip or fiber are misaligned by a shift  $\Delta r = (\Delta x, \Delta y, \Delta z)$  and the optical axes of both beams are tilted by an angle  $\Delta\alpha$ . We can and will interpret Equation (5.18) in two different ways. On the one hand we can regard Figure 5.2 as a diagram showing the region in front of the optical fiber. This means that all physical parameters ( $r_0, \Delta z, \mu_s, \Delta s$  and  $\Delta\alpha_s$ ) in Figure 5.2 refer to the fiber endface.<sup>2</sup> On the other hand we can calculate the overlap integral in front of the chip, i.e., on the “other side” of the optical system. Figure 5.2 then shows the region in front of the chip and  $r_0, \Delta z, \mu_s, \Delta s$  and  $\Delta\alpha_s$  are the parameters describing the spots on the chip side of the optical system. From a physical point of view, the first case applies to the tolerances for the adjustment of a fiber to a prefabricated subassembly consisting of chip and optical systems. The second case, in contrast, applies to the adjustment of a chip to a subunit consisting of a fiber and the optical system. Since fiber-to-chip coupling often requires a considerable

spot magnification  $-m = 2 \dots 6$  are typical values for the coupling of laser diodes – the alignment tolerances for both cases can differ significantly.

Since the calculation of the overlap integral (5.18) is somewhat lengthy, we will restrict ourselves to a brief sketch of some intermediate steps. First, we shall rotate our coordinate system such that the intermediate image becomes parallel to the  $xy$ -plane; i.e., we match the two beams

$$\gamma_s = -10 \log |\langle \Phi(\mu_s r_0, 0, \Delta s, \Delta z) | \Phi(r_0, -\Delta\alpha_s, 0, 0) \rangle|^2$$

in a tilted reference plane. This procedure has the advantage that we can neglect the diffraction of the beam emitted from the fiber and thus obtain a simplified overlap integral which has the standard form

$$\int_{-\infty}^{\infty} dx e^{-ax^2+2ibx} = \frac{1}{\sqrt{a\pi}} e^{-b^2/a}.$$

If we calculate the overlap integral, we see that the coupling loss has three contributions

$$\gamma_s = \gamma_s^{(\tau)} + \gamma_s^{(\Delta r)} + \gamma_s^{(\Delta\alpha)} \quad (5.19)$$

where the first

$$\gamma_s^{(\tau)} = 10 \log \left( \frac{\sqrt{1 + \tau_s^2}}{2\tau_s \Gamma_s \cos \Delta\alpha_s} \right) \quad (5.20)$$

designates the coupling loss caused by defocusing and mismatch of the field radii of the eigenmodes. The second contribution in Equation (5.19)

$$\gamma_s^{(\Delta r)} = 10 \log(e) \left[ \frac{(\Gamma_s \Delta s \cos \Delta\alpha_s)^2}{\sqrt{2} r_0} \right]^2 \quad (5.21)$$

stands for coupling losses due to a transverse misalignment, and the third contribution

$$\gamma_s^{(\Delta\alpha)} = 20 \log(e) \left( 1 - \Gamma_s^2 \right) \frac{\cos^2 \Delta\alpha_s}{\tan^2 \theta} \quad (5.22)$$

describes the coupling loss caused by the tilt of the two optical axes with respect to each other. The other two parameters occurring in Equation (5.19) are

$$\tau_s = \mu_s \cos \Delta\alpha_s$$

and

$$\Gamma_s = \sqrt{\frac{1 + \tau_s^2}{(1 + \tau_s^2)^2 + (\Delta z \cos^2 \Delta\alpha_s / z_c)^2}}$$

where  $z_c = k_0 n r_0^2 / 2$  is the confocal length and  $\theta = \tan^{-1}(2/k_0 n r_0)$  is the diffraction angle. Both parameters were introduced in the previous section.

To obtain a physical insight into the underlying magnitudes we will discuss some special configurations.

2. This was our original interpretation of the overlap integral in Figure 5.2.

### Elliptical Eigenmodes

A large class of integrated optical waveguides supports elliptical eigenmodes. It is impossible to transform such a mode into a circular one by a rotational symmetric imaging system. In consequence, coupling losses will occur. For perfect alignment, i.e., for  $\Delta r = 0$  and  $\Delta\alpha_x = \Delta\alpha_y = 0$ , the contributions  $\gamma_s^{(\Delta\alpha)}$  and  $\gamma_s^{(\Delta r)}$  vanish and we obtain

$$\gamma_s = 10 \log \left( \frac{1 + \mu_s^2}{2\mu_s} \right) \quad (5.23)$$

for the coupling loss per coordinate axis ( $s = x, y$ )

By differentiating Equation (5.17) with respect to a mean magnification factor  $\mu$  of the imaging system<sup>3</sup> we can easily show that the best fiber-to-chip coupling can be obtained if the imaging system magnifies the spot such that

$$\mu_x \mu_y = 1, \quad (5.24)$$

i.e., the imaging system coupling an elliptical mode to a single-mode fiber must be designed such that it couples a circular spot whose field radius is the geometric mean of the half-axes matched perfectly to the fiber. The minimum coupling loss is given by

$$\gamma_{\text{opt}} = 10 \log \left( \frac{(\mu_x/\mu_y + 1)^2}{4\mu_x/\mu_y} \right). \quad (5.25)$$

Figure 5.3 shows for the minimal coupling loss vs. the ratio  $\mu_x/\mu_y$  of the half-axes of an elliptical chip near-field. We see that the coupling losses caused by the ellipticity of the chip near-fields are moderate. For a ratio  $\mu_x/\mu_y = 2$  of the half-axes for example the minimal coupling loss is only 0.5 dB.

### Defocusing

We will now study the case of defocusing, i.e., of the shift  $\Delta z$  of the image along the optical axis. As in the previous paragraph we choose all other parameters ( $\Delta x = \Delta y = 0$ ,  $\Delta\alpha_x = \Delta\alpha_y = 0$ ,  $\mu_x = \mu_y = 1$ ) such that the only mismatch is due to defocusing ( $\Delta z \neq 0$ ). The coupling loss is then given by<sup>4</sup>

$$\gamma_{\text{def}} = 10 \log \left( 1 + \left( \frac{\Delta z}{2z_c} \right)^2 \right). \quad (5.26)$$

Figure 5.4 shows the coupling loss as a function of the relative defocusing  $z/z_c$ .

The confocal length  $z_c$  of a single-mode fiber driven in the near-infrared region is  $\geq 50 \mu\text{m}$ ; i.e., the tolerances against defocusing on the fiber side are rather high. The situation can be quite different on the chip side of the imaging system. For example, the confocal parameter for an eigenmode emitted by a semiconductor laser or an InGaAsP waveguide can be  $< 1 \mu\text{m}$ .

3. This means that we substitute  $\mu_x = \mu \tilde{\mu}_x$  and  $\mu_y = \mu \tilde{\mu}_y$  for the moment.

4. The coupling loss per coordinate axis is half the overall coupling loss  $\gamma_s = \gamma/2$ .

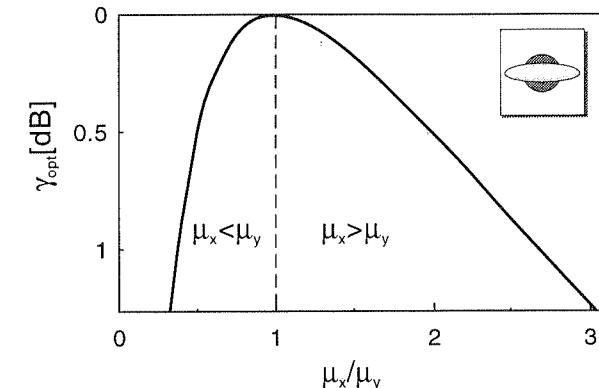


Figure 5.3. Minimal coupling loss vs. ratio  $\mu_x/\mu_y$  of half-axes for the coupling of an elliptical near-field to a circular fiber.

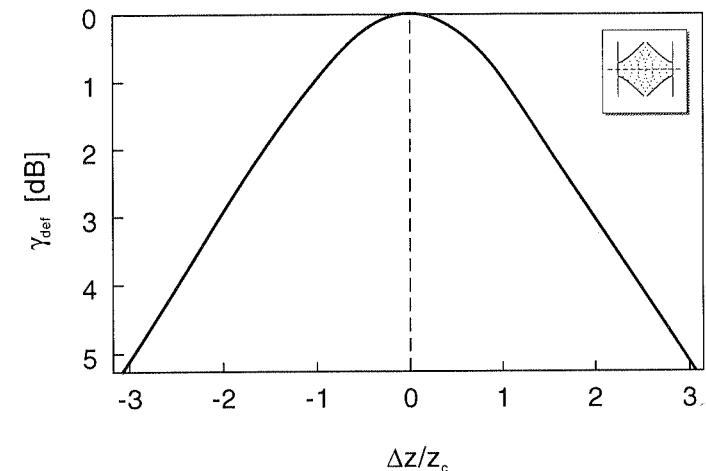


Figure 5.4. Coupling loss vs. relative defocusing.

### Transverse Misalignment

Another interesting case is that of the transverse misalignment of two beams, i.e., misalignment of the beams in the plane which is perpendicular to the optical axes of the two beams. If we assume perfect matching of the other parameters ( $\Delta z = 0$ ,  $\Delta\alpha_x = \Delta\alpha_y = 0$ ,  $\mu_x = \mu_y = 1$ ) we find

$$\gamma_{\text{tm}} = 10 \log(e) \frac{\Delta x^2 + \Delta y^2}{r_0^2}. \quad (5.27)$$

We see that the transverse shifts  $\Delta x$  and  $\Delta y$  of the intermediate image are related to the field radius  $r_0$  of the eigenmode at the endface. Figure 5.5 shows the coupling as a function of the relative transverse shift  $\rho = \sqrt{\Delta x^2 + \Delta y^2}/r_0$ . The parabolic shape of the coupling loss with increasing mismatch can also be observed in planar spectrographs provided that both fan-in and fan-out are formed by single-mode waveguides (see Chapter 8).

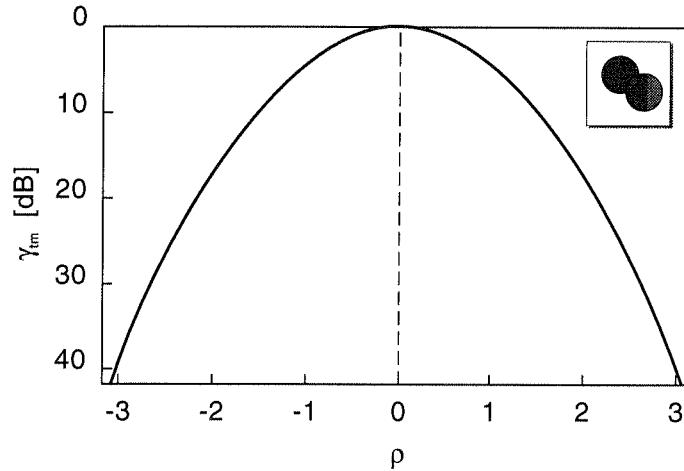


Figure 5.5. Coupling loss vs. relative transverse misalignment.

If, for example, a coupling loss of 1 dB due to transverse misalignment can be tolerated, then the intermediate image must be located inside a circular window which is defined by the field radius of the eigenmode of the fiber, i.e.,

$$\rho \leq \frac{1}{\sqrt{10 \log(e)}} \approx \frac{1}{2}.$$

For a single-mode fiber in the near-infrared region, the radius of this circle is approximately  $2.5 \mu\text{m}$ . On the chip side of the imaging system the situation can again be quite different since the field radii of the eigenmodes of semiconductor waveguides can be below  $1 \mu\text{m}$ .

### Tilted Beam Axes

The last case we will study is that of two beams with tilted axes. As in previous examples we assume optimum values for the other parameters, i.e.,  $\Delta r = 0$  and  $\mu_x = \mu_y = 1$ . The coupling loss is then given by

$$\gamma = \frac{10 \log(e)}{\tan^2 \theta} \left( \frac{\sin^2 \Delta \alpha_x}{2 - \sin^2 \Delta \alpha_x} + \frac{\sin^2 \Delta \alpha_y}{2 - \sin^2 \Delta \alpha_y} \right) \quad (5.28)$$

where

$$\theta = \tan^{-1} \left( \frac{2}{k_0 n r_0} \right)$$

is the diffraction angle. We see that the tolerance toward a tilt of the axes of the beams diminishes with a decreasing diffraction angle  $\theta$ . Thus, for example, a coupling loss of 1 dB due to angular misalignment can be tolerated as long as the optical axis of the intermediate image is located inside a angular window which is defined by the diffraction angle of the eigenmode of the fiber, i.e.,

$$\sqrt{\Delta \alpha_x^2 + \Delta \alpha_y^2} \leq \frac{\theta}{\sqrt{10 \log(e)/2}} \approx \frac{\theta}{\sqrt{2}}$$

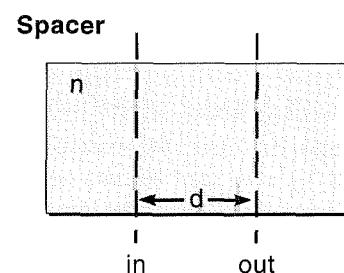
For a standard single-mode fiber we obtain  $\theta \approx 5^\circ$ , i.e., angular tolerances of approximately  $3.5^\circ$ .

## 5.3 GAUSSIAN OPTICS

In this section we will study the transformation of Gaussian beams by optical systems. We will take advantage of the fact that the Gaussian beam looks like a bundle of rays emitted from a position in complex space and from this derive a concise description of its transformation rules based on the  $ABCD$  matrices known from classical ray optics. The end of this section is formed by an analysis of spherical and chromatic aberrations.

### 5.3.1 Transformation Rules for the Beam Parameter

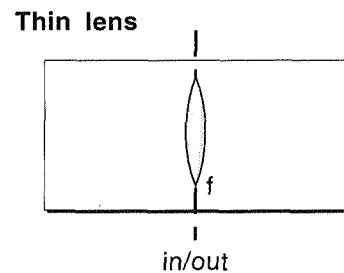
We saw in Section 5.1 that the evolution of Hermite-Gaussian beams is determined by a single parameter, the complex beam parameter  $\zeta$ . We will first derive the transformation rules of the two most elementary optical systems, the spacer, the dielectric interface and the thin lens (see Figure 5.6).



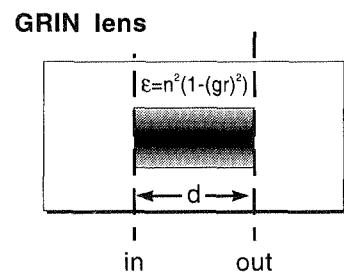
$$\begin{pmatrix} 1 & d \\ 0 & 1 \end{pmatrix}$$



$$\begin{pmatrix} 1 & 0 \\ 0 & n_{\text{in}}/n_{\text{out}} \end{pmatrix}$$



$$\begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix}$$



$$\begin{pmatrix} \cos(gd) & \sin(gd)/ng \\ -ng \sin(gd) & \cos(gd) \end{pmatrix}$$

### Spacer

The spacer is a layer of thickness  $d$  which consists of the same material as the background surrounding the spacer. The beam parameter on the incoming side of the spacer is then given by

$$\zeta_{\text{in}} = z_{\text{in}} + iz_c$$

where  $z_c = k_0 nr_0^2/2$  designates the confocal parameter of the Hermite-Gaussian beam. The beam parameter on the output side of the spacer (again in the same material) is given by

$$\zeta_{\text{out}} = z_{\text{in}} + d + iz_c,$$

i.e., for a spacer the beam parameter is transformed by

$$\zeta_{\text{out}} = \zeta_{\text{in}} + d \quad (5.29)$$

### Dielectric Interface

At the dielectric interface, the refractive index changes from  $n_{\text{in}}$  to  $n_{\text{out}}$ . The optical field defined by Equation (5.1) is obviously continuous at the interface if

$$\frac{\zeta_{\text{out}}}{n_{\text{out}}} = \frac{\zeta_{\text{in}}}{n_{\text{in}}} \quad (5.30)$$

is satisfied. We should note that the local beam radius  $r_0(z)$  and the phase shift  $\varphi(z)$  are not affected by this transformation.

### Thin Lens

The transformation rule describing the action of a thin lens of focal length  $f$  is given by the generalized "lens formula"

$$\frac{1}{\zeta_{\text{out}}} = \frac{1}{\zeta_{\text{in}}} - \frac{1}{f} \quad (5.31)$$

which relates the positions of object and image to the focal length of the lens. We will now prove that Equation (5.31) does indeed describe the transformation by a thin lens.

For this purpose let us consider the representation

$$\frac{1}{\zeta} = \frac{1}{R(z)} + i \frac{2}{k_0 n r_0(z)^2}$$

for the inverse beam parameter introduced in Section 5.1. We see directly that the local beam radius  $r_0(z)$  is not affected by the transformation (5.31). Using the representation (compare Equation (5.12))

$$\begin{aligned} \phi^{(m)}(x, z) &= |N_m(z)| H_m \left( \frac{\sqrt{2}x}{r_0(z)} \right) \exp \left( -\frac{x^2}{r_0(z)^2} \right) \\ &\times \exp \left( i \frac{k_0 n}{2R(z)} x^2 - (2m+1)\varphi(z)/2 \right) \end{aligned}$$

Figure 5.6.  $ABCD$  matrices of some elementary optical systems.

of the Hermite-Gaussian beam we see that only the phase portrait is affected by the transformation (5.31). Thus, the operator describing the transformation by a thin lens is unitary; i.e., the optical power is conserved. Since the focal length of the lens is a real number we can replace Equation (5.31) by

$$\frac{1}{R_{\text{out}}} = \frac{1}{R_{\text{in}}} - \frac{1}{f}$$

We saw in Section 5.1 that in the far-field limit the parameters  $R_{\text{in}}$  and  $R_{\text{out}}$  stand for the distances between the beam waists of the Hermite-Gaussian beams and the actual position; i.e., Equation (5.31), can be regarded as the lens formula of the Hermite-Gaussian beams.

The transformation rules for the elementary optical systems treated so far can be regarded as special cases of a bilinear mapping

$$\zeta_{\text{out}} = \frac{A\zeta_{\text{in}} + B}{C\zeta_{\text{in}} + D} \quad (5.32)$$

where the coefficients  $A$ ,  $B$ ,  $C$  and  $D$  describe the underlying optical system. We can easily show that Equation (5.32) and the set of linear equations

$$\begin{pmatrix} \alpha_{\text{out}}\zeta_{\text{out}} \\ \alpha_{\text{out}} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} \alpha_{\text{in}}\zeta_{\text{in}} \\ \alpha_{\text{in}} \end{pmatrix} \quad (5.33)$$

with the arbitrary parameters  $\alpha_{\text{in}} \neq 0$  and  $\alpha_{\text{out}} \neq 0$  have the same solution.

The  $ABCD$  matrices of the elementary optical systems treated so far are listed in Figure 5.6.

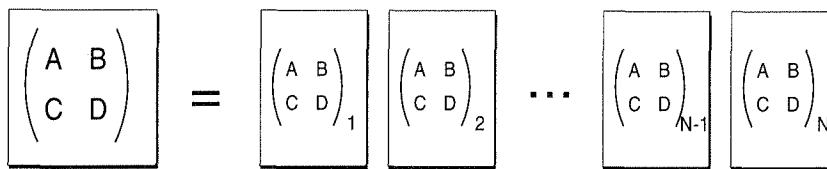


Figure 5.7. Compound optical system.

Using the matrix representation (5.33) we see that a compound optical system consisting of  $N$  subsystems  $S_i$  (see Figure 5.7) is described by the stacked  $ABCD$  matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix}_N \cdots \begin{pmatrix} A & B \\ C & D \end{pmatrix}_1 \quad (5.34)$$

As an illustrative example we will now derive the  $ABCD$  matrix of a graded-index (GRIN) lens. A GRIN lens is a rod of length  $d$  whose cross section is described by

the dielectric profile

$$\epsilon(r) = n^2 \left[ 1 - (gr)^2 + \sum_{l=2}^{\infty} h_{2l} (gr)^{2l} \right]. \quad (5.35)$$

In the following we will use the classical BPM in its paraxial approximation to treat a parabolic GRIN lens ( $h_{2l} = 0$ ). Since we know from our discussions in Section 4.2 that the classical BPM is unstable for large propagation steps, we decompose the GRIN rod into a sequence of thin plates each having a thickness  $\Delta d$ . Due to the split-operator technique of the classical BPM, the  $ABCD$  matrix of a single plate is given by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{\Delta d} = \begin{pmatrix} 1 & \Delta d/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f & 1 \end{pmatrix} \begin{pmatrix} 1 & \Delta d/2 \\ 0 & 1 \end{pmatrix} \quad (5.36)$$

in which the first and third terms on the right-hand side describe the free-space propagation over half the thickness of the plate, whereas the second term on the right-hand side accounts for the phase correction of the plate. Using the paraxial propagator of the classical BPM (compare Equation (4.38)) and the Equation (5.31) describing the action of a thin lens, we see that the focal length corresponding to the phase correction of a single plate is given by

$$f = \frac{1}{g^2 \Delta d}.$$

For an infinitely thin plate we can regard the terms occurring in Equation (5.36) as the first terms of four Taylor series and replace the  $ABCD$  matrix by

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{\Delta d} \approx \begin{pmatrix} \cos(g\Delta d) & \sin(g\Delta d)/g \\ -g \sin(g\Delta d) & \cos(g\Delta d) \end{pmatrix}.$$

We should note that the determinant of the  $ABCD$  matrix is unaffected by this approximation. Since the  $ABCD$  matrix describing a thin plate of the GRIN rod is unimodular, the full rod is described by replacing  $g\Delta d \rightarrow gd$  (see Section 2.2.4). If we additionally stack two dielectric interfaces accounting for the transition from the GRIN lens to free space, we obtain

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{\text{GRIN lens}} = \begin{pmatrix} \cos(gd) & \sin(gd)/(ng) \\ -ng \sin(gd) & \cos(gd) \end{pmatrix} \quad (5.37)$$

for the  $ABCD$  matrix of the GRIN lens embedded in free space. A precise analysis [1] of the rays traveling through the GRIN lens would show that the dielectric profile

$$\epsilon(r) = n^2 \left( 1 - (gr)^2 + \frac{2}{3}(gr)^4 - \frac{17}{45}(gr)^6 + \dots \right) \quad (5.38)$$

yields an aberration-free operation of the GRIN lens.

### 5.3.2 Imaging by Optical Systems

We have shown in the previous section that a Hermite-Gaussian beam traveling through a compound optical system is transformed by stacking the  $ABCD$  matrices of the optical subsystems. We will now relate the  $ABCD$  matrices for the Hermite-Gaussian beam to the corresponding matrices for geometrical optics and then derive the fundamental laws of imaging.

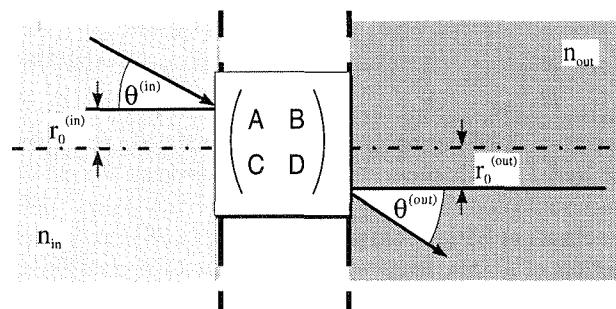


Figure 5.8. Action of an  $ABCD$  matrix.

For this purpose, we shall restrict ourselves to the far-field region which is characterized by

$$z \approx \frac{r_0(z)}{\tan \theta} \gg z_c$$

and neglect the influence of the confocal parameter  $z_c$  on the beam parameter  $\zeta$ . Choosing  $\alpha_s = \tan \theta_s$  ( $s = \text{in}, \text{out}$ ) we can replace Equation (5.33) by its geometrical optical equivalent

$$\begin{pmatrix} r_0^{(\text{out})} \\ \tan \theta^{(\text{out})} \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} r_0^{(\text{in})} \\ \tan \theta^{(\text{in})} \end{pmatrix}. \quad (5.39)$$

Figure 5.8 illustrates the action of an  $ABCD$  matrix. Let us now briefly discuss the basics of imaging:

#### **ABCD Matrix of an Imaging Optical System**

A necessary condition for the formation of an optical image (see Figure 5.9) is that all rays emitted from an arbitrary point in the object plane intersect the image plane at a single point. Let us consider a ray starting in the object plane at a point which is  $r_0^{(\text{in})}$  distant from the optical axis, and let us assume that it travels in the direction  $\theta^{(\text{in})}$  (with respect to the optical axis). The image of the point will then be found

$$r_0^{(\text{out})} = Ar_0^{(\text{in})} + B \tan \theta^{(\text{in})}$$

distant from the optical axis. The angle of intersection is

$$\tan \theta^{(\text{out})} = Cr_0^{(\text{in})} + D \tan \theta^{(\text{in})}.$$

The condition that all rays emitted from one point – irrespective of the starting angle  $\theta^{(\text{in})}$  – intersect the image plane at a single point requires that the coefficient  $B = 0$  of the  $ABCD$  matrix vanish. We see furthermore that the coefficient  $A$  stands for the spot magnification factor

$$m = \frac{r_0^{(\text{out})}}{r_0^{(\text{in})}}. \quad (5.40)$$

Using the second equation describing the imaging it becomes clear that the coefficient  $D$  describes the angular magnification

$$m_\theta = \frac{\tan \theta^{(\text{out})}}{\tan \theta^{(\text{in})}}. \quad (5.41)$$

of a bundle of rays emitted at the optical axis ( $r_0^{(\text{in})} = 0$ ). Using the rays starting parallel to the optical axis ( $\theta^{(\text{in})} = 0$ ) we see that the coefficient  $C$  of the  $ABCD$  matrix must be given by  $C = -1/f_{\text{out}}$ , i.e., by the inverse of the focal length on the output side of the optical system.

To summarize, we can say that an optical system which maps the input plane  $I$  to the output plane  $O$  is described by the  $ABCD$  matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{I \rightarrow O} = \begin{pmatrix} m & 0 \\ -1/f_{\text{out}} & m_\theta \end{pmatrix} \quad (5.42)$$

where  $m$  denotes the spot magnification,  $m_\theta$  is the angular magnification and  $f_{\text{out}}$  the focal length on the output side of the optical system. The angular magnification  $m_\theta$  of an optical system described by a unimodular  $ABCD$  matrix is given by the inverse spot magnification, i.e.,  $m_\theta = 1/m$ .

#### **General Lens-like Optical Systems**

The thick lens (see Figure 5.9) is the most general lens-like optical system; i.e., each optical system described by an  $ABCD$  matrix with  $C \neq 0$  can be replaced by a single thick lens whose principal planes are located at

$$s_{\text{in}} = \frac{A - 1}{C} \quad (5.43)$$

and

$$s_{\text{out}} = \frac{D - (AD - BC)}{C} \quad (5.44)$$

distant from the input and output sides of the optical system. The focal lengths of the thick lens are given by

$$f_{\text{in}} = -\frac{AD - BC}{C} \quad (5.45)$$

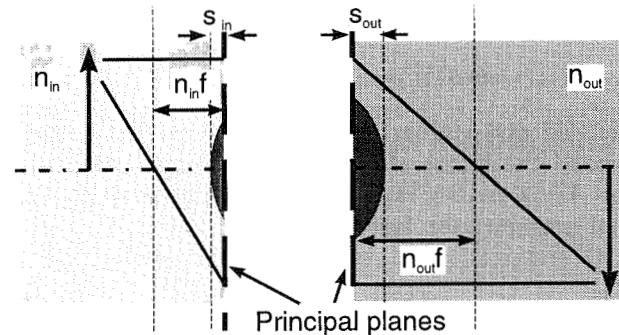


Figure 5.9. Imaging by a thick lens.

and

$$f_{\text{out}} = -\frac{1}{C}. \quad (5.46)$$

For an optical system described by a unimodular  $ABCD$  matrix  $AD - BC = 1$ , the focal length on the input and output sides will coincide, i.e.,  $f_{\text{in}} = f_{\text{out}}$ . In any other case the optical system has different materials with refractive indices  $n_{\text{in}}$  and  $n_{\text{out}}$  on its input and output sides. The focal lengths can then be given

$$\begin{aligned} f_{\text{in}} &= n_{\text{in}}f \\ f_{\text{out}} &= n_{\text{out}}f \end{aligned} \quad (5.47)$$

in terms of the focal length  $f$  of a thick lens embedded in free space.

For a proof of the first part of this theorem, we satisfy the four conditions

$$\left( \begin{array}{cc} 1 & -s_{\text{out}} \\ 0 & 1 \end{array} \right) \left( \begin{array}{cc} A & B \\ C & D \end{array} \right) \left( \begin{array}{cc} 1 & -s_{\text{in}} \\ 0 & 1 \end{array} \right) \stackrel{!}{=} \left( \begin{array}{cc} 1 & 0 \\ -1/f_{\text{out}} & f_{\text{in}}/f_{\text{out}} \end{array} \right).$$

The formulas for the focal lengths can be derived from

$$\left( \begin{array}{cc} 1 & 0 \\ 0 & 1/n_{\text{out}} \end{array} \right) \left( \begin{array}{cc} 1 & 0 \\ -1/f & 1 \end{array} \right) \left( \begin{array}{cc} 1 & 0 \\ 0 & n_{\text{in}} \end{array} \right) \stackrel{!}{=} \left( \begin{array}{cc} 1 & 0 \\ -1/f_{\text{out}} & f_{\text{in}}/f_{\text{out}} \end{array} \right).$$

Incidentally, the  $ABCD$  matrix describing the transport of a beam from the input focus of a thick lens to its output focus is given by

$$\left( \begin{array}{cc} A & B \\ C & D \end{array} \right)_{F_{\text{in}} \rightarrow F_{\text{out}}} = \left( \begin{array}{cc} 0 & f_{\text{in}} \\ -1/f_{\text{out}} & 0 \end{array} \right). \quad (5.48)$$

### Imaging of Compound Optical Systems

The spot magnification of a compound optical system consisting of  $N$  optical subsystems is given by the product

$$m = \prod_{i=1}^N m^{(i)} \quad (5.49)$$

of the spot magnifications  $m^{(i)}$  of the subsystems, and the angular magnification of the overall optical system is given by the product

$$m_\theta = \prod_{i=1}^N m_\theta^{(i)} \quad (5.50)$$

of the angular magnifications  $m^{(i)}$  of the subsystems.

We prove this theorem for a two-component optical system and obtain

$$\left( \begin{array}{cc} m^{(1)} & 0 \\ -1/f_{\text{out}}^{(1)} & m_\theta^{(1)} \end{array} \right) \left( \begin{array}{cc} m^{(2)} & 0 \\ -1/f_{\text{out}}^{(2)} & m_\theta^{(2)} \end{array} \right) \stackrel{!}{=} \left( \begin{array}{cc} m^{(1)}m^{(2)} & 0 \\ C & m_\theta^{(1)}m_\theta^{(2)} \end{array} \right).$$

The extension to  $N$ -component optical systems is straightforward.

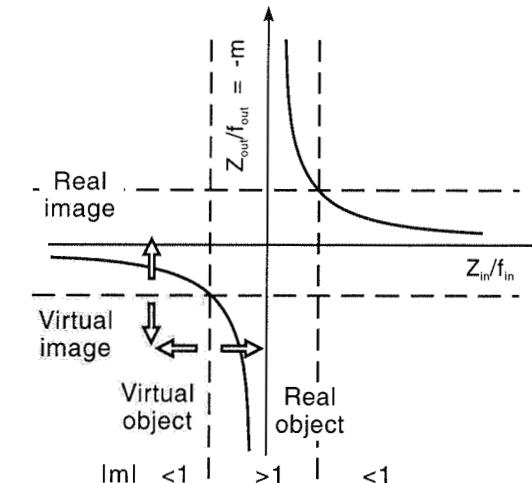


Figure 5.10. Newton's law of imaging.

### Newton's Law of Imaging

The imaging properties of any lens-like optical system are governed by Newton's law of imaging

$$Z_{\text{in}}Z_{\text{out}} = f_{\text{in}}f_{\text{out}} \quad (5.51)$$

where

$$Z_s = d_s - f_s \quad (5.52)$$

with  $s = \text{in}, \text{out}$  designate the distances between the object (in), image (out) and the corresponding focal points. The spot magnification is given by

$$m = -\frac{Z_{\text{out}}}{f_{\text{out}}}, \quad (5.53)$$

and the angular magnification of the mapping by

$$m_\theta = -\frac{Z_{\text{in}}}{f_{\text{out}}}. \quad (5.54)$$

Figure 5.10 shows a diagram illustrating Newton's law of imaging. The shaded regions indicate virtual objects and images.

We have already shown that the single thick lens is the most general lens-like optical system. Thus, we need only satisfy the four conditions

$$\begin{pmatrix} 1 & f_{\text{out}} + Z_{\text{out}} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1/f_{\text{out}} & f_{\text{in}}/f_{\text{out}} \end{pmatrix} \begin{pmatrix} 1 & f_{\text{in}} + Z_{\text{in}} \\ 0 & 1 \end{pmatrix} \stackrel{!}{=} \begin{pmatrix} m & 0 \\ -1/f_{\text{out}} & m_\theta \end{pmatrix}.$$

to prove Newton's law of imaging.

### 5.3.3 Spherical Aberrations

Up to now we have assumed a perfect optical system which converts Hermite-Gaussian beams back into Hermite-Gaussian beams. Real optical systems, however, exhibit aberrations; i.e., the shape of the near-field is distorted by the optical system. In this section we will study the most significant axial aberration, namely, spherical aberrations.

We saw in Section 5.2 that the action of a thin circular lens embedded in free space is described by the unitary operator

$$\mathcal{U}_{\text{thin lens}} = \exp \left( -i \frac{k_0}{2f} r^2 \right) \quad (5.55)$$

where  $f$  designates the focal length of the lens. The action of a real lens exhibiting spherical aberrations must then be given by the operator

$$\mathcal{U}_{\text{real lens}} = \exp \left( -i \frac{k_0}{2f} r^2 \left( 1 - h_{\text{SA}} \frac{r^2}{f^2} \right) \right) \quad (5.56)$$

where the aberration parameter  $h_{\text{SA}}$  stands for the strength of the spherical aberrations. Before we start to examine the coupling loss caused by spherical aberrations,

let us make some brief comments on the mathematical structure of the lens operator  $\mathcal{U}_{\text{real lens}}$ . We saw in Section 4.1 that the beam propagation in lossless media as a solution of the paraxial Helmholtz equation is described by unitary operators. This feature is caused by neglecting the evanescent modes, which keeps any optical path retraceable. Hence, the lens operator  $\mathcal{U}_{\text{real lens}}$  describing a lossless lens must be also unitary, i.e., the aberration parameter  $h_{\text{SA}}$  must be a real quantity for lossless lenses.<sup>5</sup>

For calculating the coupling loss, we shall restrict ourselves to Gaussian beams, i.e., to the fundamental Hermite-Gaussian beams. We shall formulate the whole section again by using the terminology of fiber-to-chip coupling where  $\Phi_c$  stands for the optical field of a more divergent beam emitted from the chip endface and  $\Phi_f$  for the field of a less divergent beam emitted from the fiber endface. The coupling loss (in decibels related to the optical power) is then given by

$$\gamma = -10 \log |\langle \Phi_f | \mathcal{U}_{\text{real lens}} \Phi_c \rangle|^2 \quad (5.57)$$

where the overlap integral is examined at the principal plane on the fiber side of the lens. As pointed out before, we assume a perfectly aligned optical system, a setup without transverse misalignment and tilted beam axes. We can then divide the coupling loss

$$\gamma = \gamma_x^{(\tau)} + \gamma_y^{(\tau)} + \gamma_{\text{SA}} \quad (5.58)$$

into three contributions. The first two terms are

$$\gamma_s^{(\tau)} = 5 \log \left( \left( \frac{\mu_s}{2} + \frac{1}{2\mu_s} \right)^2 + (\Delta z/z_c)^2 \right) \quad (5.59)$$

with the index  $s = x, y$ . The other parameters are the ratios  $\mu_s$  between the required and actual spot magnification and the confocal length  $z_c$  of the near-field. Equation (5.59) accounts for defocusing and improper magnification of the optical systems. Such contributions would also occur for an aberration-free optical system (see also Section 5.2). The third term

$$\gamma_{\text{SA}} = -10 \log \left| \frac{1}{\pi \sqrt{\eta_x \eta_y}} \iint_{-\infty}^{+\infty} dx dy e^{-ix^2/\eta_x - iy^2/\eta_y + i(x^2+y^2)^2} \right|^2 \quad (5.60)$$

stands for the extra losses caused by the spherical aberrations. Obviously, this contribution depends only on two complex parameters  $\eta_x$  and  $\eta_y$  which are defined by

$$\frac{1}{\eta_s} = \left( \frac{1}{\zeta_{cs}} - \frac{1}{\zeta_f} + \frac{1}{f} \right) \sqrt{\frac{k_0 n f^3}{2 h_{\text{SA}}}} \quad (5.61)$$

<sup>5</sup> The parameters of spherical aberrations obtained by a ray tracing analysis depend on the input beam; i.e., they lead as an artefact to non-linear lens operators. The linear operators calculated in the next section refer to a source which is located at the focus of the lens.

with  $s = x, y$ . The beam parameters  $\zeta_{cx}$  and  $\zeta_{cy}$  belong to the elliptical Gaussian beam emitted from the chip endface, the parameter  $\zeta_f$  describes the beam emitted from the fiber endface. We see immediately that the parameters  $\eta_x$  and  $\eta_y$  vanish with vanishing spherical aberrations  $h_{SA} \rightarrow 0$ . In general, the overlap integral (5.60) describing the coupling losses due to spherical aberrations must be examined numerically.

However, for the coupling of two circular beams, i.e.,  $\eta = \eta_x = \eta_y$ , we can find a closed solution if the image is outside the region of the caustic cusp  $\text{Im}(\eta^2) < 0$  which is caused by the spherical aberrations (see Section 2.3). We will elaborate this analytic solution to get a feeling for the relevant parameters of the spherical aberrations. By using polar coordinates  $r$  and  $\varphi$  and by substituting  $t = r^2 = x^2 + y^2$  we can show that the coupling loss due to spherical aberration is then given by

$$\gamma_{SA}^{(\text{circ})} = -10 \log \left| \frac{\sqrt{\pi}}{2\eta} \operatorname{erfc} \left( \frac{\sqrt{-i}}{2\eta} \right) e^{-i/(4\eta^2)} \right|^2 \quad (5.62)$$

where  $\operatorname{erfc}(\xi)$  designates the complementary error function. For small spherical aberrations ( $\eta \rightarrow 0$ ) we can replace the complementary error function by its asymptotic expansion and thus obtain

$$\gamma_{SA}^{(\text{circ})} \approx -10 \log |1 - 2i\eta^2|^2.$$

We will now consider optical system which would map the chip near-field perfectly onto the fiber endface if the optical system did not exhibit spherical aberrations  $\gamma_x^{(\tau)} = \gamma_y^{(\tau)} = 0$ . Using

$$\gamma_{SA}^{(\text{circ})} \approx 20 \log(e) |h_{SA}| k_0 n f \tan^4 \theta_c$$

we can see that the coupling loss due to spherical aberrations increases linearly with increasing focal length  $f$  and the aberration parameter  $\eta$  and that it grows with the fourth power of the diffraction angle (precisely;  $\tan \theta_c$ ) of the Gaussian beam emitted from the chip endface. We see furthermore that the coupling losses caused by spherical aberrations can be reduced by keeping the focal lengths of the critical lenses small and by reducing the aberration parameter  $h_{SA}$ . As a result, the endface emitting the more divergent beam, in our case the chip endface, must be located close to the surface of the first lens.

Figure 5.11 shows the coupling loss of an optical system exhibiting spherical aberrations as a function of the axial position of one of the endfaces. The optical system of our tutorial example is chosen such that it would allow for perfect coupling provided it were aberration-free. We see that the two curves showing the coupling losses of the true optical system and of its aberration-free variant are shifted against each other. The ray trajectories mounted on top of the coupling curves indicate that the plane of “minimal confusion” where best coupling is obtained and the plane of the intermediate image do not match. It becomes clear that – as a consequence of the distorted image – the tolerance toward defocusing will increase. Naturally, the same argument holds for transverse misalignment.

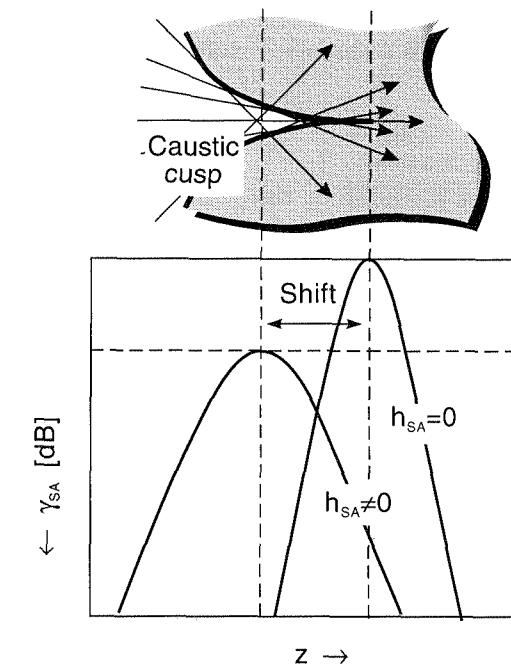


Figure 5.11. Influence of spherical aberrations on fiber-to-chip coupling.

### 5.3.4 Chromatic Aberrations

Optical systems operated over a large range of wavelengths  $\lambda_{\max} \leq \lambda \leq \lambda_{\min}$  exhibit chromatic aberrations which are caused by dispersion inside the elementary optical subsystems (usually by dispersion inside the lenses). The influence of dispersion is characterized by the generalized Abbe number

$$\nu = \frac{f(\bar{\lambda})}{f(\lambda_{\max}) - f(\lambda_{\min})} \quad (5.63)$$

where  $\bar{\lambda} = (\lambda_{\max} + \lambda_{\min})/2$  designates the mean wavelength of operation. This quantity coincides with its classical analog if we use the corresponding wavelengths  $\lambda_{\max} = 643.8 \text{ nm}$  and  $\lambda_{\min} = 483.0 \text{ nm}$  in the visible region. For rough estimations of the effect of chromatic aberrations it is useful to replace the first derivative of the focal length

$$\frac{\partial f}{\partial \lambda} \approx \frac{f(\bar{\lambda})}{\nu (\lambda_{\max} - \lambda_{\min})}$$

by the generalized Abbe number,

The dispersion of the material of a thin lens  $i$ , for example, results in a change of its focal lengths and thus in a chromatic shift, i.e., in a change of the position

$$\frac{\partial d_{\text{out}}^{(i)}}{\partial \lambda} = (2 - m^{(i)})^2 \frac{\partial d_{\text{in}}^{(i)}}{\partial \lambda} - (1 - m^{(i)})^2 \frac{\partial f^{(i)}}{\partial \lambda} \quad (5.64)$$

of the intermediate image produced by this lens. The first term on the right-hand side of Equations (5.64) accounts for the chromatic shift  $\partial d_{\text{in}}^{(i)} / \partial \lambda$  of the intermediate image which acts as object for the thin lens  $i$ . The second term describes the chromatic shift caused by the variation of the focal length of the lens. A thin lens will also exhibit a chromatic magnification shift

$$\frac{\partial m^{(i)}}{\partial \lambda} = \frac{2 - m^{(i)}}{f} \left( (1 - m^{(i)}) \frac{\partial f}{\partial \lambda} - (2 - m^{(i)}) \frac{\partial d_{\text{in}}^{(i)}}{\partial \lambda} \right), \quad (5.65)$$

i.e., the spot magnification changes with changing wavelength. In addition to the changes in the individual optical subsystems, the field radii of the near-field patterns of chip and fiber<sup>6</sup> will vary, and so will the required spot magnification  $m_{\text{req}}$  of the optical system.

All three effects contribute to the coupling loss caused by the chromatic aberrations of an optical system. In the spectral vicinity of the design wavelength  $\lambda_0$  which does not exhibit any misalignment at  $\lambda_0$ , the coupling loss (see Section 5.2) is given by

$$\begin{aligned} & \gamma_{\text{CA}}(\lambda_0 + \Delta\lambda) \\ & \approx 10 \log \left( 1 + \Delta\lambda^2 \left[ \left( \frac{1}{m_{\text{req}}} \frac{\partial m}{\partial \lambda} - \frac{1}{m_{\text{req}}^2} \frac{\partial m_{\text{req}}}{\partial \lambda} \right)^2 + \left( \frac{1}{2z_c^{(f)}} \frac{\partial d_{\text{out}}^{(i)}}{\partial \lambda} \right)^2 \right] \right) \\ & \approx \Delta\lambda^2 10 \log(e) \left[ \left( \frac{1}{m_{\text{req}}} \frac{\partial m}{\partial \lambda} - \frac{1}{m_{\text{req}}^2} \frac{\partial m_{\text{req}}}{\partial \lambda} \right)^2 + \left( \frac{1}{2z_c^{(f)}} \frac{\partial d_{\text{out}}^{(i)}}{\partial \lambda} \right)^2 \right]. \end{aligned} \quad (5.66)$$

Figure 5.12 shows the effects of chromatic aberration on a diagram of the coupling loss versus the position of the image. We see immediately that the amount of chromatic aberrations increases with increasing focal length.

Up to now, we have not included the effects of aberrations discussed in the previous section. In fact, there is an interaction between the spherical and chromatic aberrations since the change in wavelength changes the spot size and consequently the angle of diffraction and the illumination of the principal planes of the critical lenses. The axial shifts caused by chromatic and spherical aberrations can partly compensate.

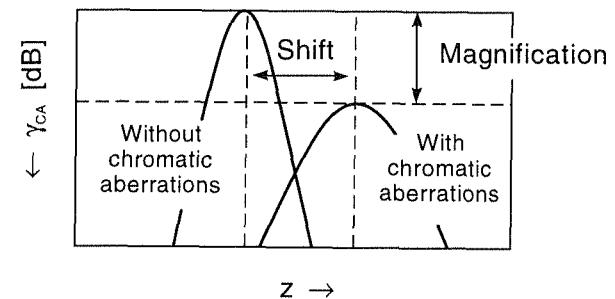


Figure 5.12. Effects of chromatic aberrations.

#### 5.4 FIBER-TO-CHIP COUPLING BY MICROLENSES

Optical systems based on microlenses are widely used to couple waveguides which are located on a semiconductor chip to single-mode fibers. The development of fiber-pigtailed modules is aimed at minimizing the size of the optical system in order to ensure simple and reliable packaging, i.e., to minimize potential misalignments due to variations in environmental conditions (especially in temperature). We saw in the previous section that theoretical considerations, i.e., the minimization of spherical and chromatic aberrations, aim at the same goal, namely, to minimize the focal lengths and thus to reduce the physical size of the corresponding hardware. We will now discuss the most prominent example of fiber-to-chip coupling: laser-to-fiber coupling for "standard" lasers, i.e., for index-guided InGaAsP laser diodes driven in the near infrared region ( $1.3/1.5 \mu\text{m}$ ). The application of the results to other examples of fiber-to-chip coupling is straightforward.

Due to the high contrast in the dielectric profile which is required for operating such lasers, the half-axes of the near-field at the chip endface are  $r_0^{(x)} = 0.5 \dots 1 \mu\text{m}$  (perpendicular to the epitaxial layers) and  $r_0^{(y)} = 1.5 \dots 3 \mu\text{m}$  (parallel to the epitaxial layers). We saw in Section 5.2 that the best coupling for an elliptical near-field emitted from a chip to a single-mode fiber is obtained if the optical system maps the geometrical mean of both half-axes of the chip near-field, i.e.,  $r_0 = \sqrt{r_0^{(x)} r_0^{(y)}} = 0.8 \dots 2 \mu\text{m}$ , onto the fiber near-field. A typical field radius of a single-mode fiber is  $5 \mu\text{m}$ . We can conclude that *an optical system for the coupling an InGaAsP laser diode to a single-mode fiber requires a spot magnification of  $m = 2.5 \dots 6$* . In the following, we will treat one- and two-lens assemblies for fiber-to-chip coupling based on the above specifications.

6. For the sake of simplicity, we assume circular near-fields of chip and fiber.

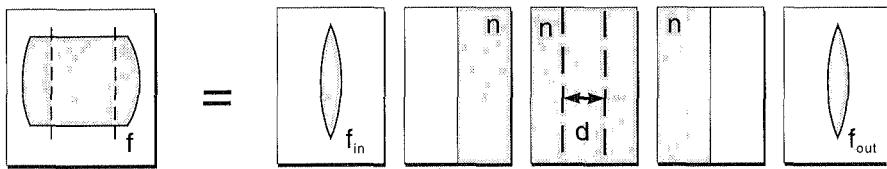


Figure 5.13. Elementary building blocks of a conventional thick lens.

#### 5.4.1 Microlenses

Let us consider the fundamental properties of the relevant microlenses, i.e., their focal length, the position of their principal planes and the parameter  $h_{SA}$  of their spherical aberration.

##### Ball and Plano-Convex Lenses

The principal planes of a conventional thick lens can be found by stacking the lens up from elementary building blocks (see Figure 5.13) and analyzing the  $ABCD$  matrix of the assembly. The focal length of the convex thin lenses on the left- and right-hand sides of the stack are given by

$$f = \frac{r}{n - 1} \quad (5.67)$$

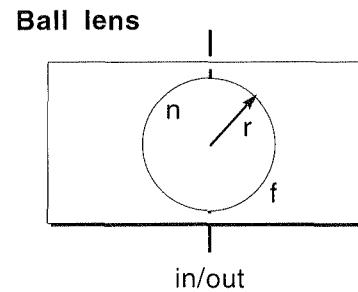
where  $r$  designates the radius of curvature of the surface of the lens and  $n$  the refractive index of its material.<sup>7</sup> The focal length of the corresponding concave lens is given by  $-f$ . The two dielectric interfaces forming the next shell of our lens are required to account for the refractive index  $n \neq 1$  inside the lens. The spacer forming the center of the lens stands for the physical thickness  $d$  of the lens. By inserting the  $ABCD$  matrices of the elementary building blocks we obtain

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{\text{lens}} = \begin{pmatrix} 1 - d/(nf_{\text{out}}) & d/n \\ -1/f_{\text{in}} - 1/f_{\text{out}} + d/(nf_{\text{in}}f_{\text{out}}) & 1 - d/(nf_{\text{in}}) \end{pmatrix} \quad (5.68)$$

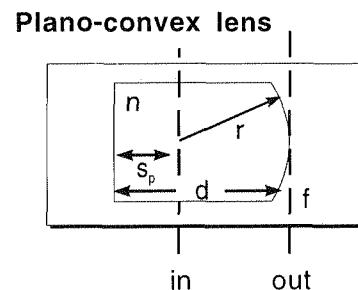
for the  $ABCD$  matrix of the thick lens, in which  $f_{\text{in}}$  and  $f_{\text{out}}$  designate the focal lengths of the input and output surfaces, respectively. We should note here that the  $ABCD$  matrix of the thick lens is unimodular, i.e.,  $AD - BC = 1$ , since assume the same material (free space) on both sides of the lens. The focal lengths on the input and output sides of the thick lens will therefore coincide, and we obtain the following according to Equations (5.43)–(5.46)

$$f_{\text{lens}} = \left( \frac{1}{f_{\text{in}}} + \frac{1}{f_{\text{out}}} - \frac{d}{nf_{\text{in}}f_{\text{out}}} \right)^{-1} \quad (5.69)$$

7. For the derivation of Equation (5.67), Snell's law (2.42) is replaced by its paraxial approximation  $\sin \theta \approx \theta$ .

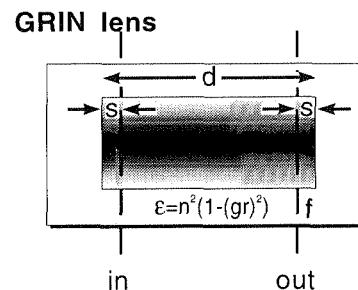


$$f = \frac{nr}{2(n-1)}$$



$$f = \frac{r}{n-1}$$

$$s_p = \frac{d}{n}$$



$$f = \frac{1}{ng \sin(gd)}$$

$$s = (1 - \cos(gd)) f$$

Figure 5.14. Focal lengths and position of the principal planes for the ball lens, plano-convex and GRIN lens.

for the focal length. The principal planes ( $s = \text{in}, \text{out}$ ) will then be found at a distance

$$s_s = \frac{d}{n} \frac{f_{\text{lens}}}{f_s} \quad (5.70)$$

apart from the corresponding surfaces of the lens. Using  $f_{\text{in}} = f_{\text{out}} = r/(n - 1)$  and  $d = 2r$  we obtain the parameters of the ball lens, by using the parameters  $f_{\text{in}} \rightarrow \infty$  and  $f_{\text{out}} = r/(n - 1)$  those of the plano-convex lens. The results of these calculations are presented in Figure 5.14.

The calculation of the parameters  $h_{SA}$  requires a more detailed analysis of the rays traveling through the lens. The corresponding calculations are found in the literature. The results are<sup>8</sup>

$$h_{SA} = \frac{1}{16} \left[ \frac{n}{(n-1)^2} - 1 \right] \quad (5.71)$$

for the ball lens [2] and

$$h_{SA} = \frac{1}{2} \left[ 1 - \frac{n-2}{n(n-1)^2} - \frac{d(n+1)(n-2)^2}{r n^3} \right] \quad (5.72)$$

for the plano-convex lens [3]. The parameters of the lenses are shown in Figure 5.14.

### GRIN Lenses

In the previous section we derived the  $ABCD$  matrix

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}_{\text{GRIN lens}} = \begin{pmatrix} \cos(gd) & \sin(gd)/(ng) \\ -ng \sin(gd) & \cos(gd) \end{pmatrix}$$

of a GRIN lens of length  $d$  as an example of a compound optical system (see Equation (5.36)). The dielectric profile of the GRIN rod was given by

$$\epsilon(r) = n^2 \left[ 1 - (gr)^2 + \sum_{l=2}^{\infty} h_{2l} (gr)^{2l} \right].$$

We see that the GRIN lens is described by a unimodular  $ABCD$  matrix with  $A = D$ . From Equations (5.43)–(5.46) we can conclude that it possesses only one focal length

$$f = \frac{1}{ng \sin(gd)} \quad (5.73)$$

and that the distances between the principal planes and the corresponding surfaces are both given by

$$s = (1 - \cos gd) f. \quad (5.74)$$

Figure 5.14 shows an outline of the GRIN lens and its relevant parameters.

For a “quarter-pitch” lens ( $gd = \pi/2$ ), the focal points are located at the endfaces of the GRIN rod. The parameter of spherical aberrations is then given by [4]

$$h_{SA} = \frac{3\pi}{16n^2} \left( h_4 - \frac{2}{3} \right). \quad (5.75)$$

We should note here that, in contrast to ball lenses, the parameter of spherical aberration  $h_{SA}$  depends on the choice of the material of the GRIN rod and on the fabrication process.

**Table 5.1**  
Properties of some commercial lenses (PC=plano-convex lens, B=ball lens).

Type	PC	B	B	B	GRIN(0.23)
	Si	Sk5	LaF22	LaSF9	
$n(1.55 \mu\text{m})$	3.4777	1.5711	1.7504	1.8131	1.590 (0.326)
$n(1.3 \mu\text{m})$	3.5053	1.5742	1.7545	1.8177	1.592 (0.327)
$\nu$	90.22	290.54	321.64	321.79	200.00
$(f-s)/f$	< 0.1	0.27	0.14	0.10	0.13
$f$	0.4r	1.37r	1.16r	1.11r	1.94 mm
$ h_{SA} $	< 0.05	0.24	0.13	0.11	< 0.1

We will now discuss some properties of typical microlenses used for fiber-to-chip coupling. In Table 5.1 we have listed the refractive indices at  $\lambda = 1.3 \mu\text{m}$  and  $1.55 \mu\text{m}$  and derived from it the generalized Abbe number for this wavelength region. For the GRIN lens with pitch = 0.23 we have added the quadratic parameter  $g[\text{mm}]$  in parentheses. We see that the Abbe number is 2–3 times higher than the corresponding values in the visible wavelength region; i.e., *the dispersion of the glass lenses decreases with increasing wavelength*. The Abbe number of the Si plano-convex lens is significantly smaller than that of most glass lenses due to the high refractive index of the lens material. The focal length of the GRIN lens is much greater than that of its conventional counterparts, whose lens radii are typically of  $r = 0.2 \dots 1 \text{ mm}$ . We can therefore expect that GRIN lenses will exhibit large chromatic shifts. We should, however, note here that achromatization by a pair of focusing and defocusing GRIN lenses has been demonstrated. The last line of our table shows the parameter of spherical aberration  $h_{SA}$ . To keep the spherical aberrations small, the focal points of all lenses are located near the surface.<sup>9</sup> We see that this parameter decreases for the ball lenses with increasing refractive index. For this reason, ball lenses made from high-index glasses are preferred for fiber-to-

8. To obtain a linear lens operator here we have used the rays starting from an effective source at the focus of the lens.

9. The thickness  $d \approx 1.4r$  of the plano-convex lens is chosen such that one focus is located on its plane surface.

chip coupling. The parameter of spherical aberration for the other two types of lenses (plano-convex and GRIN lenses) is fabrication-dependent but always lower than that of ball lenses.

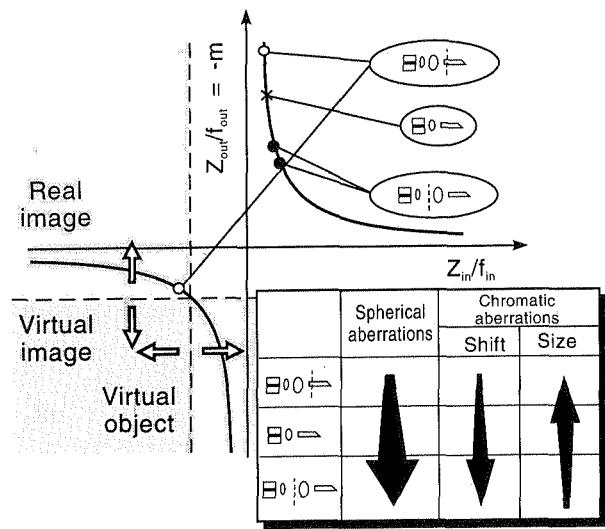


Figure 5.15. One- and two-lens assemblies.

#### 5.4.2 One- and Two-Lens Assemblies

For a one-lens assembly, the lens must magnify the chip near-field by a factor  $m = 2.5 \dots 6$ . Although two-lens assemblies must provide the same overall spot magnification  $m = m_1 m_2$ , the magnification factors of the two lenses  $m_1$  and  $m_2$  are free parameters of the optical system. Figure 5.15 shows Newton's diagram and three essentially different options for fiber-to-chip coupling by convex lenses: one-lens assembly and two-lens assemblies with and without a real intermediate image. The position of the intermediate images within the optical system is indicated by dashed lines. The two-lens assemblies are classified by their relative magnification

$$M = \frac{m_1}{m}$$

We can see from Figure 5.15 that for an optical system exhibiting a real intermediate image the relative spot magnification is given by  $-1 < M < -1/m$ ; i.e., the first lens accomplishes only a part of the required spot magnification. The intermediate

image of the other two-lens assembly is greater  $M > 1$  than the final one. The chip side of the optical system is always the critical part with respect to the spherical aberrations. Since the illumination of the first principle plane decreases with increasing relative magnification we can conclude that the two-lens assembly without an intermediate image will exhibit the smallest spherical aberrations. Thus, for fiber-to-chip couplings which are driven by only one wavelength such optical systems are a good solution. However, for cost reasons one-lens optical systems are increasingly preferred.

The situation becomes more complicated if a fiber-to-chip coupling driven at two different wavelengths is required. Using Equation (5.64) we see that

$$\frac{\partial d_{\text{out}}}{\partial \lambda} = \left( \frac{1}{M} - m_1 \right)^2 \frac{\partial f_1}{\partial \lambda} + \left( 1 - \frac{1}{M} \right)^2 \frac{\partial f_2}{\partial \lambda} \quad (5.76)$$

the chromatic shifts will sum up if both lenses are convex  $f_1/f_2 > 0$ . We see furthermore that the chromatic shift will increase with increasing focal lengths of the lenses; i.e., lenses with small focal lengths offer benefits. We observe that the chromatic shift will increase with decreasing relative magnification  $M$  of the first lens. The chromatic magnification error, in contrast, increases with increasing  $M$ .

## 5.5 OPTICAL TAPERS

Up to now we have studied the mode conversion by butt coupling and imaging, i.e., by modifying beams which propagate through a homogeneous region with appropriate optical systems. We will discuss in this section an alternative way to adapt the near-field to a new waveguide, tapering. Optical tapers are, roughly speaking, waveguide structures which vary slowly along the propagation direction. Provided this modification is adiabatic, i.e., “small enough”, the distribution of the optical power over the local eigenmodes of the taper will be conserved during propagation.

We will start this section with a derivation of the local normal mode theory in order to understand the physical mechanisms governing the tapering process. The second part of this section will deal with more practical questions of tapering.

### 5.5.1 Local Normal Mode Theory

The local normal mode theory treats the mode conversion in a tapered, i.e., a slowly varying waveguide structure by the eigenmode based BPM (see Section 4.4.1). For this type of treatment we divide the tapered structure into a sequence of thin plates (see Figure 5.16), and we assume as in Chapter 4 that the dielectric profile remains unchanged within each of these plates.

Let us now concentrate on a single slice which runs from  $z$  to  $z + \Delta z$ . As indicated in Figure 5.16 we must first expand the entire optical field  $|\phi\rangle_z$  into the

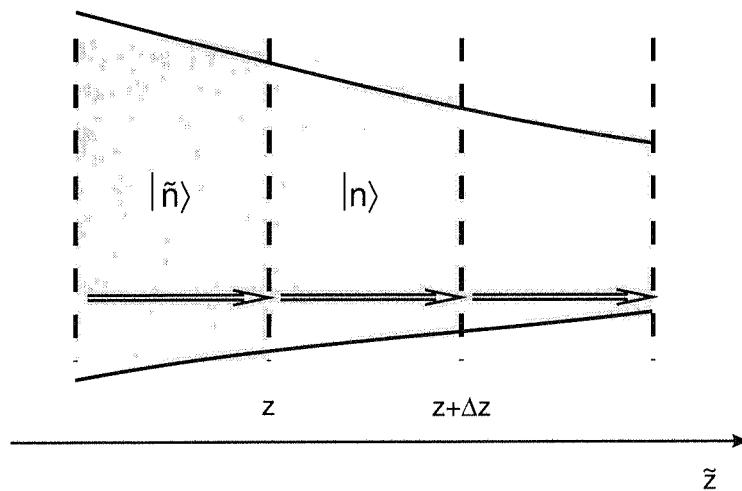


Figure 5.16. Tapered waveguide structure.

eigenmodes  $|n\rangle_z$  of the actual plate. We obtain

$$\sum_m a_n(z) |\tilde{n}\rangle_z = |\phi\rangle_z = \sum_m a_n(z_+) |n\rangle_z.$$

The left-hand equation shows the expansion of the optical field into the eigenmodes  $|\tilde{n}\rangle_z$  of the previous plate, the right-hand equation stands for the expansion into the actual eigenmodes. By scalar multiplication with an eigenmode  $\langle m|_z$  of the actual slice we obtain an equation

$$a_m(z_+) = a_m(z) + \sum_n C_{mn}(z) a_n(z) \Delta z \quad (5.77)$$

which expresses the expansion coefficients  $a_m(z_+)$  of the actual plate in terms of the previous ones  $a_m(z)$ . The coupling coefficient  $C_{mn}$  is given by the scalar product

$$C_{mn}(z) = \frac{\langle m | \tilde{n} - n \rangle_z}{\Delta z}. \quad (5.78)$$

We have formulated Equation (5.77) such that the adiabatic character of mode conversion becomes obvious, i.e., that the coupling to the other eigenmodes which is governed by the second term on the right-hand side of Equation (5.77) decreases with decreasing thickness of the actual plate. The evolution of the actual expansion coefficients inside the plate is simply given by

$$\begin{aligned} a_m(z + \Delta z) &= a_m(z_+) e^{iq_m \Delta z} \\ &= a_m(z_+) (1 + iq_m \Delta z) + O(\Delta z^2). \end{aligned} \quad (5.79)$$

Using Equations (5.77) and (5.80) and the representation

$$\frac{da_m}{dz} = \lim_{\Delta z \rightarrow 0} \frac{a_m(z + \Delta z) - a_m(z)}{\Delta z}$$

for the first derivative of the expansion coefficient  $a_m(z)$  we can formulate a set of ordinary linear differential equations

$$\frac{da_m}{dz} = iq_m(z) a_m(z) + \sum_n C_{mn}(z) a_n(z) \quad (5.80)$$

for the evolution of the expansion coefficients  $a_m(z)$ . We should point out here that the coupling coefficients are  $z$ -dependent; i.e., the dynamic system described by Equation (5.80) is non-autonomous.

Before we discuss Equation (5.80) in more detail we will derive explicit expressions for the coupling coefficients  $C_{mn}$  in the framework of the first-order Rayleigh-Schrödinger perturbation theory (see Section 3.5.2). For this purpose we regard the Hamiltonian describing the actual plate as the unperturbed Hamiltonian. The operator difference describing the difference between the actual and the previous plate is treated as a perturbation. In the framework of a first-order perturbation theory the difference of eigenmodes  $|\tilde{n} - n\rangle$  of the previous and the actual plate is then given by the first-order correction of the actual eigenmode. Inserting the corresponding expression (3.52) into Equation (5.78) we obtain

$$\begin{aligned} C_{mn}(z) &= (1 - \delta_{mn}) \frac{\langle m | \frac{\partial H}{\partial z} n \rangle}{\epsilon_n - \epsilon_m} \\ &\approx (1 - \delta_{mn}) \frac{\langle m | \frac{\partial \epsilon}{\partial z} n \rangle}{\epsilon_n - \epsilon_m} \end{aligned} \quad (5.81)$$

Obviously, the last expression holds exactly for weakly guiding waveguide structures (see Section 3.6). For the vectorial formulations it approximates the true coupling coefficient by its dominant term. Equation (5.81) clearly exhibits the resonant character of mode conversion. Using the resonance denominator  $1/(\epsilon_n - \epsilon_m)$  we can see that only the few next neighboured eigenmodes will significantly contribute to the mode conversion. We can further conclude that the conversion rate will increase with decreasing guidance of the waveguide. We will return to this point in the following section.

We will now start with the solution of the local normal mode equations (5.80) by eliminating the first term via the transformation

$$a_m(z) = \alpha_m(z) \exp \left( i \int_0^z d\tilde{z} q_m(\tilde{z}) \right) \quad (5.82)$$

The transformed local normal mode equation is then given by

$$\frac{d\alpha_m}{dz} = \sum_n C_{mn} \alpha_n \quad (5.83)$$

with the initial values  $\alpha_m(0) = a_m(0)$ . For small conversion rates we can tackle Equation (5.80) by successive iterations. For this purpose we expand the function

$$\alpha_m(z) = \sum_{l=0}^{\infty} \alpha_m^{(l)}(z) \quad (5.84)$$

describing the evolution of the  $m$ th eigenmode into a series over the coefficients

$$\alpha_m^{(l)}(z) = \int_0^z d\tilde{z} \sum_n C_{mn}(\tilde{z}) \alpha_n^{(l-1)}(z) \quad (5.85)$$

which are determined by successive iterations starting at  $\alpha_m^{(0)}(z) = a_m(0)$ . By differentiating the Equations (5.84) and using the coefficients (5.85) we can easily verify that the expansions (5.84) are indeed a solution of the system (5.83) of differential equations.

We can see by successive iteration of  $\alpha_m^{(l)}(z)$  according to Equation (5.85) that the  $l$ th iteration coefficient  $\alpha_m^{(l)}(z_0)$  describes the influence of mode conversion via  $l$  intermediate states during the propagation from  $z = 0$  to  $z_0$  on the evolution of the  $m$ th eigenmode. This physical interpretation makes clear why the quality of the results obtained by the iteration process will increase with decreasing conversion rates. The condition for adiabatic mode conversion rates can thus be formulated as

$$\left\| \left\langle m \left| \frac{\partial \epsilon}{\partial z} n \right. \right\rangle \right\| \ll |q_n(z) - q_m(z)|, \quad (5.86)$$

i.e., the dielectric profile of a adiabatic taper must not significantly vary over one beat to the next neighbouring eigenmode.

### 5.5.2 Design of Optical Tapers

Let us make a few remarks on the practical design of optical tapers. Figure 5.17 shows the field radius of a step-index waveguide as a function of its transverse extent. We see that the field radius can be enlarged by both “up-sizing” and “down-sizing” the waveguide cross section; i.e., spot magnification can be obtained by up-tapering and down-tapering of the waveguide. Using the local normal mode theory we see that there is a big difference between both methods of tapering. Up-tapering yields a successive improvement of guidance. We can conclude from Equation (5.81) that the mode-conversion rate of a linear taper of this type will decrease with increasing spot size. Tapers with a constant conversion rate – practical devices are usually designed in this way – will then exhibit a horn-like shape. Down-tapering, in contrast, results in a successive deterioration of the guidance and, in consequence, into conversion rates which increase with increasing spot size. We can conclude that up-tapering is less critical than down-tapering. Devices based on up-tapering

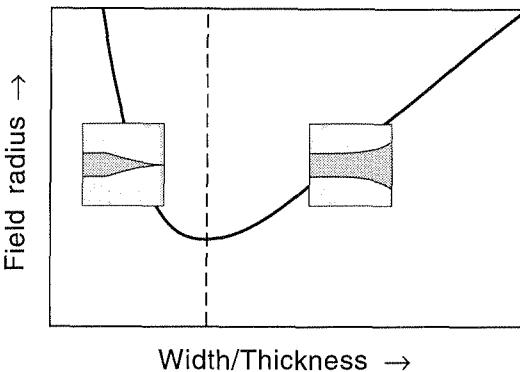


Figure 5.17. Field radius of a step-index waveguide vs. its transverse extent.

are usually shorter than those operating by down-tapering. However, up-tapering will become increasingly critical if the waveguide supports an increasing number of guided modes.

Let us now return to the most important application of optical tapers, to fiber-to-chip coupling. We have seen in the previous sections of this chapter that fiber-to-chip coupling causes problems which gain importance with increasing spot magnification. The essential reasons for this problems were decreasing tolerances towards transverse misalignment and increasing aberrations. Obviously, optical on-chip tapers are an attractive tool to tackle both problems at the same time. Let us consider an optical taper in the InGaAsP/InP material system as an illustrative example. The lateral mode extension (parallel to the epitaxial layers) will usually be adjusted by up-tapering. For the vertical axes, down-tapering is an attractive option since the

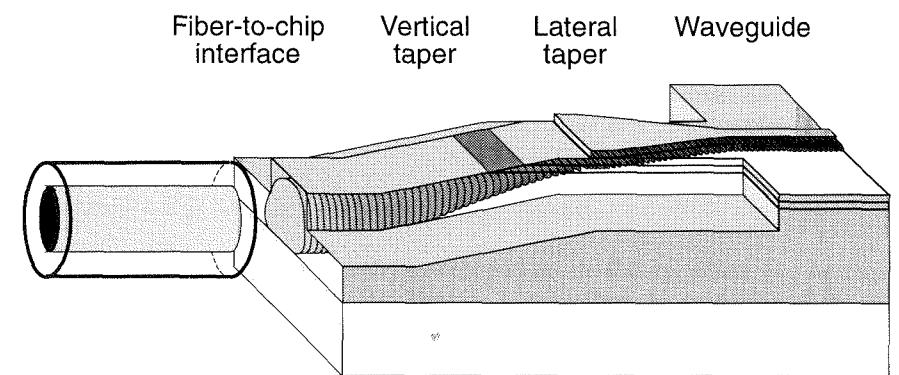


Figure 5.18. Waveguide taper in InGaAsP/InP (by courtesy of Dr. M. Schienle, Siemens AG).

fabrication of the corresponding epitaxial layers results in technological problems. Nevertheless, up-tapering is available at the expense of thick epitaxial layers (see Figure 5.18). Although the local normal mode theory can act as a design tool for optical tapers, most devices are constructed using the beam propagation method (BPM) today.

## REFERENCES

- [1] Iga, K., Kokubun Y., and Oikawa M., *Fundamentals of Microoptics*, Tokyo, Academic Press Japan, 1984.
- [2] Sumida, M., and Takemoto, K., "Lens Coupling of Laser Diodes to Single-Mode Fibers", *IEEE Journal of Lightwave Technology*, Vol. 2, 1984, pp. 305–311.
- [3] Karstensen, H., and Frankenberger, R., "High-Efficiency Two Lens Laser Diode to Single-Mode Coupler with a Silicon Plano Convex Lens", *IEEE Journal of Lightwave Technology*, Vol. 7, 1989, pp. 244–249.
- [4] Sumida, M., and Takemoto, K., "Aberrations of GRIN-Rod Lenses in Single-Mode Optical Devices", *Electronics Letters*, Vol. 22, 1986, pp. 1318–1319.
- [5] Haus, H. A., *Waves and Fields in Optoelectronics*, Englewood Cliffs, NJ, Prentice-Hall, 1984.
- [6] Kogelnik, H., and Li, T., "Laser Beams and Resonators", *Proceedings of the IEEE*, Vol. 54, 1966, pp. 1312–1329.
- [7] Snyder, A. W., and Love, J. D., *Optical Waveguide Theory*, London, Chapman and Hall, 1983.
- [8] Milton, A. F., and Burns, W. K., "Mode Coupling in Optical Waveguide Horns", *IEEE Journal of Quantum Electronics*, Vol. 13, 1977, pp. 828–835.

## Chapter 6

### Codirectional Coupling

Codirectional coupling, i.e., the interaction of guided modes traveling in the same direction, forms the basis for the operation of a large class of guided wave components in integrated optics. The coupled mode theory, which acts as the theoretical foundation of codirectional coupling, was originally developed to analyze the behavior of microwave directional couplers and dates back to middle of the twentieth century. Since the use of strip waveguides and the coupling of guided modes fits perfectly into the framework of integrated optics, directional couplers were among the first devices to be realized in the pioneer days of integrated optics. Their applications are widespread today. The examples treated within this chapter include optical filters for large and to some extent also for small channel spacings, optical switches or modulators as well as guided wave beam splitters.

This chapter covers the design tools for codirectional coupling, i.e., the coupled mode theory describing the operation of uniform devices and the transfer matrix technique which turns out to be crucial for the description of compound components and networks. We will further study the most important elementary devices – ideal (uniform) symmetrical and asymmetrical directional couplers and realistic (non-uniform) devices which are affected by  $\kappa$ -tapering, Y-branches, crossings – as well as compound components – Mach-Zehnder interferometers and couplers and directional couplers with a periodic overlay, i.e.,  $\Delta\beta$ - and  $\Delta\kappa$ -couplers.

#### 6.1 COUPLED MODE THEORY

In the first section we will establish the coupled mode theory which can be regarded as the theoretical foundation of this chapter. It describes the evolution of the optical field under the influence of codirectional coupling. We will derive this theory for a general coupled waveguide array which consists of  $N$  single-mode waveguides. In the later sections of this chapter we will restrict ourselves to the directional coupler, i.e., to the most elementary coupled waveguide array consisting of only two waveguides.

The coupled mode theory is based on the assumption that the supermodes –

the eigenmodes of the coupled waveguide array – can be represented by the guided modes of the individual waveguides. In consequence, our discussion starts by elucidating the physical background of this assumption, i.e., we will derive the conservation law for the optical power stored in the guided modes and draw some conclusions about the spectrum of the Hamiltonian of the coupled waveguide array. We will continue with an examination of the phenomenological coupled mode theory which regards the coupling coefficients as purely phenomenological parameters. In the framework of this discussion we will derive a general solution of the coupled mode equations for a longitudinally uniform coupled waveguide array, i.e., to a waveguide structure consisting of parallel waveguides. In addition, we will elucidate the interdependence of the symmetry relations for the coupling coefficients and the conservation law of the optical power. What follows is a derivation and discussion of the coupling coefficients from first principles. The end of this section is formed by an outline of the theory of tapered coupling.

### 6.1.1 Eigenmodes of Coupled Waveguide Arrays

We will start our examination of coupled mode theory with a discussion of the spectra of Hamiltonians describing coupled waveguide arrays. To do this, we will collect some information about the number and spectral position of eigenmodes. The main purpose of this discussion is to examine the basic assumption of coupled mode theory: *A coupled waveguide array, which is built up from  $N$  marginally<sup>1</sup> coupled waveguides, supports  $N$  guided modes*

$$|m\rangle = \sum_{n=1}^N c_{nm} |0^{(n)}\rangle \quad (6.1)$$

whose amplitudes  $|m\rangle$  can be represented by those of the guided modes  $|0^{(n)}\rangle$  of the individual waveguides. To avoid confusion between the two sets of modes occurring in Equation (6.1), we will designate the eigenmodes  $|m\rangle$  of the composite waveguide structure as supermodes of the coupled waveguide array.

In mathematical terms, Equation (6.1) means that the eigenspace of the guided supermodes of a coupled waveguide array is spanned by the guided modes of its individual waveguides. This means in physical terms that *the optical power launched into a coupled waveguide array via one of its waveguides is interchanged between the guided modes of the waveguides forming the array, and it is not transferred into the radiation modes*. We can further conclude that *an arbitrary (rapidly varying) optical field*

$$|\Phi(z)\rangle = \sum_{n=1}^N a_n(z) |0^{(n)}\rangle$$

which can be represented by the guided modes of a bundle of lossless waveguides, i.e., of a coupled waveguide array formed by lossless waveguides, will exhibit lossless propagation:

$$\frac{\partial}{\partial z} \langle \phi | \phi \rangle = \frac{\partial}{\partial z} \sum_{n=1}^N |a_n|^2 = 0. \quad (6.2)$$

Since the eigenmodes of the individual waveguides form a set of non-orthogonal functions, we must use the “covariant” set of eigenmodes indicated by the bra-vector  $\langle 0^{(n)} |$  in order to guarantee the orthogonality relations  $\langle 0^{(m)} | 0^{(n)} \rangle = \delta_{mn}$ . The proof of the conservation law for the optical power is then straightforward.

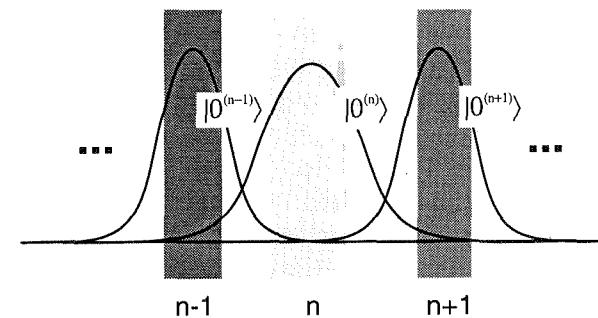


Figure 6.1. Coupled waveguide array.

To clarify this scenario, we will use the Rayleigh-Schrödinger perturbation theory (see Section 3.5). In our discussion we will concentrate on one arbitrary single-mode waveguide  $n$  of a coupled waveguide array. In the framework of a first-order perturbation analysis, the effective dielectric constant  $\epsilon_0^{(n)}$  of its guided mode is then modified by

$$\Delta\epsilon_0^{(n)} = \langle 0^{(n)} | \mathcal{H}_1 | 0^{(n)} \rangle \quad (6.3)$$

in which the Hamiltonian  $\mathcal{H}_1$  stands for the perturbation caused by all other waveguides of the coupled waveguide array. Figure 6.1 shows for a one-dimensional coupled waveguide array the waveguide  $n$ , its guided mode and the immediately adjacent waveguides. We see that, if the waveguides are far away from each other, no optical power of the guided mode  $|0^{(n)}\rangle$  will be confined in the other waveguides of the coupled waveguide array; i.e., the coupling between the waveguides will be marginal. For the same reasons, the first-order corrections of the guided mode

$$\langle 0^{(l)} | \Delta 0^{(n)} \rangle = \frac{\langle 0^{(l)} | \mathcal{H}_1 | 0^{(n)} \rangle}{\epsilon_0^{(n)} - \epsilon_0^{(l)}} \quad (6.4)$$

1. Two waveguides are called *marginally coupled* if their coupling coefficient vanishes ( $\kappa \rightarrow 0$ ).

by the radiation modes  $l$  will be negligible. Thus, we can conclude that for the “no-coupling” limit the effective dielectric constant  $\epsilon_0^{(n)}$  of the guided mode of waveguide  $n$  will also be an eigenvalue of the coupled waveguide array.

If we carry out the same procedure for all waveguides of our waveguide array we see that the marginally coupled waveguide array supports  $N$  guided modes. The spectrum of its guided eigenvalues is given by the effective dielectric constants of the individual waveguides. This means in particular that *if  $n$  waveguides of a marginally coupled waveguide array support a guided mode with the same effective dielectric constant  $\epsilon_0^{(n)}$ , then the corresponding eigenvalue of the coupled waveguide array will be  $n$ -fold degenerated*. We can further conclude that the guided modes of the individual waveguides will span the eigenspace of the guided modes of the coupled waveguide array; i.e., the supermodes can be expanded according to Equation (6.1).

With increasing coupling between the waveguides of a coupled waveguide array the expansion (6.1) usually remains valid to a good approximation. However, Equation (6.4) shows that the influence of the radiation on the guided modes of the individual waveguides will increase with increasing coupling. The resonance denominator indicates that this effect will occur more quickly if the guidance of the individual waveguides is poor. The following simple example will demonstrate that this effect can result in a complete breakdown of the expansion (6.1). Let us consider a one-dimensional coupled waveguide array with a background dielectric constant  $\epsilon_B$  which is built up from  $N$  single-layer slab waveguides each having a width  $d$  and a dielectric constant  $\epsilon_L$ . We have seen in Section 3.10 that the individual waveguides will support only one guided mode if the V parameter satisfies the condition

$$V = k_0 d \sqrt{\epsilon_L - \epsilon_B} < \pi.$$

The most efficient coupling between the waveguides of the coupled waveguide array can be obtained if the waveguides are located close together, i.e., if the coupled waveguide array represents a single-layer slab of width  $d_{\text{CWA}} = Nd$ . The V parameter is then given by  $V_{\text{CWA}} = NV < N\pi$ . Obviously, the coupled waveguide array has  $N$  guided modes if the V parameter of the individual waveguides satisfies the condition

$$V \geq \left(1 - \frac{1}{N}\right)\pi.$$

If this condition is violated, the coupled waveguide array has less than  $N$  guided supermodes, and expansion (6.1) becomes obsolete. We see that the required V parameter increases with an increasing number of waveguides in the coupled waveguide array. For the most elementary coupled waveguide array, the symmetrical directional coupler ( $N = 2$ ),  $V \geq \pi/2$  is required.

### 6.1.2 Phenomenological Coupled Mode Theory

We will now derive a coupled mode theory which is based on the results of the previous section, i.e., on the expansion of the optical field

$$|\Phi(z)\rangle = \sum_{n=1}^N a_n(z) |0^{(n)}\rangle \quad (6.5)$$

into the eigenmodes of the individual waveguides  $|0^{(n)}\rangle$  and on the conservation law of optical power

$$\frac{\partial}{\partial z} \sum_{n=1}^N |a_n|^2 = 0 \quad (6.6)$$

which holds for lossless coupled waveguide arrays.

On the basis of the law of expansion (6.5), we can conclude that *the evolution of the expansion coefficients  $a_n(z)$  is governed by the coupled mode equations, i.e., by the dynamic system*

$$-i \frac{\partial a_m}{\partial z} = \sum_{n=1}^N \kappa_{mn} a_n. \quad (6.7)$$

In the “no-coupling” limit the diagonal matrix elements

$$\kappa_{mm} = q_0^{(m)} \quad (6.8)$$

are given by the propagation constants of the individual waveguides. The non-diagonal matrix elements describe the coupling between two waveguides  $m$  and  $n$ .

We will understand our theory for the moment as a purely phenomenological coupled mode theory; i.e., we will regard the coupling coefficients as purely phenomenological parameters. As long as the waveguides forming the coupled waveguide array run parallel, the coupling coefficients represent constants: the coupled mode equations (6.7) form a set of autonomous ordinary differential equations. If a coupled waveguide array comprises longitudinal structures such as tapers or periodic overlays, the coupling coefficients will become  $z$ -dependent; i.e., the dynamic system (6.7) describing the evolution of the expansion coefficients will become a non-autonomous set of differential equations. We will concentrate for the moment on longitudinally invariant structures and leave the other cases for our later discussions.

Using the conservation law for the optical power (6.6) of a lossless coupled waveguide array, we see that

$$\frac{\partial}{\partial z} \sum_{m=1}^N |a_m|^2 = i \sum_{m,n=1}^N (\kappa_{mn} - \kappa_{nm}^*) a_m a_n^* \stackrel{!}{=} 0,$$

This means that *the optical power is conserved if and only if the propagation constants of the individual waveguides  $\kappa_{mm} = q_m$  are real and the coupling coefficients satisfy the symmetry relations*

$$\kappa_{mn} = \kappa_{nm}^*. \quad (6.9)$$

We will now start solving the coupled mode equations. Using the matrix formulation

$$\frac{\partial \mathbf{a}}{\partial z} = i\mathcal{K}\mathbf{a} \quad (6.10)$$

of Equation (6.7) with the vector elements  $(\mathbf{a})_n = a_n$  and the matrix elements  $(\mathcal{K})_{mn} = \kappa_{mn}$  we see that the coupled mode equations define a beam propagation problem which is restricted to the subspace of guided modes. A comparison with Section 4.1 shows that the coupling matrix  $\mathcal{K}$  is related to the Hamiltonian  $\mathcal{H}$  of the coupled waveguide array via  $\mathcal{K} \Leftrightarrow k_0\sqrt{\mathcal{H}}$ . That is, the eigenvalues of the matrix  $\mathcal{K}$  represent the propagation constants of the supermodes expanded into the guided modes of the individual waveguides. In analogy to our derivation in Section 4.1 we can show that the evolution of the optical field

$$\mathbf{a}(z_0 + \Delta z) = \mathcal{U}_{CM}(\Delta z)\mathbf{a}(z_0) \quad (6.11)$$

is governed by the coupled mode propagator

$$\mathcal{U}_{CM}(\Delta z) = \exp(i\mathcal{K}\Delta z). \quad (6.12)$$

We should, however, note the mathematical differences between the coupled mode equations and the forward Helmholtz equation treated in Chapter 4, i.e., the difference between a set of ordinary differential equations and a partial differential equation. As a consequence,  $\mathcal{K}$  represents a matrix acting in a vector space of dimension  $N^2$  and not a linear operator in Hilbert space. Therefore, a direct solution of the coupled mode equations (6.10) via the eigenmodes of the matrix  $\mathcal{K}$  is usually preferred.<sup>2</sup> For this purpose, we consider the representation of the optical field

$$\mathbf{a}(z) = \sum_{n=1}^N c_n \exp(iq_n z) \mathbf{a}_n \quad (6.13)$$

in terms of the eigenvectors

$$\mathcal{K}\mathbf{a}_n = q_n \mathbf{a}_n \quad (6.14)$$

of the matrix  $\mathcal{K}$  as a starting point. As mentioned previously, the eigenvectors  $\mathbf{a}_n$  in Equation (6.14) represent the supermodes of the coupled waveguide array decomposed into the eigenmodes of the individual waveguides. The coefficients

$$c_n = \bar{\mathbf{a}}_n \cdot \mathbf{a}(0) \quad (6.15)$$

in Equation (6.13) in turn describe the decomposition of the optical field into supermodes. The vectors  $\bar{\mathbf{a}}_n$  form the covariant set of eigenvectors, i.e.,  $\bar{\mathbf{a}}_m \cdot \mathbf{a}_n = \delta_{mn}$ .

We have shown at the beginning of this discussion that the coupling matrix of a lossless waveguide array is Hermitian;  $\mathcal{K} = \mathcal{K}^\dagger$ . We can directly conclude that the eigenvalues of the matrix  $\mathcal{K}$ , i.e., the propagation constants of the supermodes, are

real and that the propagator  $\mathcal{U}_{CM}$  represents a unitary matrix; i.e., the eigenvectors  $\mathbf{a}_n$  of the matrix  $\mathcal{K}$  are orthogonal to each other. In more mathematical terms, *the propagator for a lossless waveguide array with N waveguides is a member of the unitary group U(N)*. By relating all phase shifts to the mean phase shift  $\bar{\varphi}$ , i.e.,

$$\mathcal{U}_{CM} = e^{i\bar{\varphi}} \tilde{\mathcal{U}}_{CM}$$

*we can formulate the properties of the coupled waveguide array in terms of the (unit-modular unitary) transfer matrix  $\tilde{\mathcal{U}}_{CM}$  which is a member of the subgroup SU(N).*

### 6.1.3 Coupling Coefficients

Up to now, we have dealt with the problem of mode coupling on a purely phenomenological level. We will now derive expressions for the coupling coefficients  $\kappa_{mn}$ . For this purpose we will calculate the evolution of the optical field  $|\Phi\rangle$  by the methods of beam propagation derived in Section 4.1. Using the formal solution (4.9) of the beam propagation problem, we see that the evolution of the optical field is governed by

$$|\Phi(z)\rangle = \exp(ik_0\sqrt{\mathcal{H}}z) |\Phi(0)\rangle. \quad (6.16)$$

where  $\mathcal{H}$  stands for the Hamiltonian describing the coupled waveguide array. The evolution of the expansion coefficients  $a_m$  is then given by

$$a_m(z) = \langle 0^{(m)} | \exp(ik_0\sqrt{\mathcal{H}}z) \Phi(0) \rangle. \quad (6.17)$$

Let us now differentiate Equation (6.17) with respect to  $z$ . Incidentally we formulate Parseval's theorem according to the basic assumptions of the coupled mode theory; i.e., we represent the unity operator  $\mathcal{E}$  by

$$\mathcal{E} \approx \sum_{m=1}^N |0^{(m)}\rangle \langle 0^{(m)}| \approx \sum_{m=1}^N |m\rangle \langle m|. \quad (6.18)$$

Inserting both expressions we obtain

$$\begin{aligned} -i \frac{\partial a_m}{\partial z} &= \langle 0^{(m)} | k_0\sqrt{\mathcal{H}} \exp(ik_0\sqrt{\mathcal{H}}z) \Phi(0) \rangle \\ &\approx \sum_{n=1}^N \langle 0^{(m)} | k_0\sqrt{\mathcal{H}} 0^{(n)} \rangle \langle 0^{(n)} | \exp(ik_0\sqrt{\mathcal{H}}z) \Phi(0) \rangle \\ &= \sum_{n=1}^N \kappa_{mn} a_n \end{aligned}$$

for the first derivative of the expansion coefficient  $a_m$ . Comparing the last two equations it becomes clear that the coupling coefficients are given by

$$\kappa_{mn} = \langle 0^{(m)} | k_0\sqrt{\mathcal{H}} 0^{(n)} \rangle. \quad (6.19)$$

2. The corresponding mathematical technique for the original problem of beam propagation was introduced in Section 4.4.1.

Using the expansion of the unity operator into the supermodes  $|l\rangle$  we can derive more explicit expressions

$$\kappa_{mn} = \sum_{l=1}^N q_l \langle 0^{(m)} | l \rangle \langle l | 0^{(n)} \rangle. \quad (6.20)$$

for the coupling coefficients. Using Equation (6.20) we see that the matrix elements  $\kappa_{mn}$  of a lossless coupled waveguide array satisfy the symmetry relations ( $\kappa_{mn} = \kappa_{nm}^*$ ), i.e., the law of power conservation (see Section 6.1.2 for a discussion of this point). By an appropriate choice of eigenmodes it is always possible to keep the coupling coefficients  $\kappa_{mn}$  real.

We will continue our discussion with a rough estimation of the coupling coefficients from first principles,<sup>3</sup> i.e., based on the eigenmodes of the individual waveguides; and we will restrict ourselves to lossless coupled waveguide arrays which are, in addition, only weakly coupled. To guarantee the symmetry relations of the matrix elements within the framework of our further approximations we will use Equation (6.19) in the more symmetrical form

$$\kappa_{mn} \approx \frac{1}{2} \{ \langle 0^{(m)} | k_0 \sqrt{\mathcal{H}} 0^{(n)} \rangle + \langle 0^{(n)} | k_0 \sqrt{\mathcal{H}} 0^{(m)} \rangle^* \}. \quad (6.21)$$

Figure 6.1 shows a coupled waveguide array described by a Hamiltonian  $\mathcal{H}$ . If we intend to apply the Hamiltonian to a guided mode  $|0^{(m)}\rangle$  of a waveguide  $m$  we can divide it into a Hamiltonian  $\mathcal{H}_0^{(m)}$  describing the unperturbed waveguide  $m$  and a Hamiltonian  $\mathcal{H}_1^{(m)}$  which stands for the (small) perturbation of the eigenmode  $|0^{(m)}\rangle$  caused by all other waveguides. Using this decomposition of the Hamiltonian we can expand its square root into a Taylor series<sup>4</sup> and obtain

$$\begin{aligned} k_0 \sqrt{\mathcal{H}} &= \sqrt{\mathcal{H}_0^{(m)} + \mathcal{H}_1^{(m)}} \\ &= \sqrt{\mathcal{H}_0^{(m)}} + \frac{1}{2} \mathcal{H}_0^{(m)}^{-1/2} \mathcal{H}_1^{(m)} \pm \dots \\ &\approx q_0^{(m)} + \frac{k_0}{2\bar{n}} \mathcal{H}_1^{(m)}. \end{aligned} \quad (6.22)$$

In Equation (6.22)  $q_0^{(m)}$  stands for the propagation constant of the guided mode of waveguide  $m$  and  $\bar{n}$  for a reference refractive index. Inserting Equation (6.22) into one of the overlap integrals in Equation (6.21) yields

$$\langle 0^{(n)} | k_0 \sqrt{\mathcal{H}} 0^{(m)} \rangle = q_0^{(m)} \delta_{mn} + \frac{k_0}{2\bar{n}} \langle 0^{(n)} | \mathcal{H}_1^{(m)} 0^{(m)} \rangle.$$

Looking again at Figure 6.1, we see that the essential contributions to the overlap integrals on the right-hand side are caused by that part of the perturbation which describes the waveguide  $n$ . Thus, we can substitute the perturbation by all waveguides by the perturbation caused by the waveguide  $n$ , i.e.,

$$\mathcal{H}_1^{(m)} \approx 2\bar{n} \Delta n^{(n)}(\mathbf{r}_t) (1 - \delta_{mn})$$

where  $\Delta n^{(n)}(\mathbf{r}_t)$  stands for the refractive index profile forming the waveguide  $n$ . By carrying out the same procedure for the other overlap integral in Equation (6.21) and by inserting all results we obtain<sup>5</sup>

$$\kappa_{mn} \approx \begin{cases} q_0^{(m)} & \text{for } m = n \\ \frac{k_0}{2} \langle 0^{(m)} | \{ \Delta n^{(m)}(\mathbf{r}_t) + \Delta n^{(n)}(\mathbf{r}_t) \} 0^{(n)} \rangle & \text{for } m \neq n. \end{cases} \quad (6.23)$$

At a brief glance it looks as if the coupling coefficient would increase with increasing refractive index contrasts  $\Delta n$  of the waveguides. The simultaneous modification of the near-fields of the guided modes, however, prevents the formulation of such a rule. However, we can derive a rule of thumb for the behavior of the coupling coefficients with increasing spacing of the waveguides. In doing so, we restrict ourselves to the coupling of TE modes in a one-dimensional array of single-mode strip waveguides (see Figure 6.2). The second overlap integral is then restricted to the region  $x_n - d_n/2 < x < x_n + d_n/2$  of the waveguide  $n$ , i.e.,

$$\langle 0^{(m)} | \Delta n^{(n)}(\mathbf{r}_t) 0^{(n)} \rangle = \Delta n^{(n)} \int_{x_n - d_n/2}^{x_n + d_n/2} dx e_y^{(m)*} e_y^{(n)}$$

We have seen in Section 3.8 that the guided mode of the strip waveguide  $m$  will then exhibit exponential behavior

$$e_y^{(m)} \propto \exp(-\alpha_m |x_m - x|)$$

where the decay constant describing the exponential tail of the guided mode of waveguide  $m$  is given by

$$\alpha_m = k_0 \sqrt{\epsilon_0^{(m)} - \epsilon_B}$$

with the effective dielectric constant  $\epsilon_0^{(m)}$  of the guided mode of waveguide  $m$  and the background dielectric constant  $\epsilon_B$ . As a consequence of this exponential behavior, we can conclude that both overlap integrals and thus the coupling coefficient will exhibit the same shape,

$$\kappa_{mn} \propto \exp\left(-k_0 \sqrt{\epsilon_0^{(m)} - \epsilon_B} |x_m - x_n|\right). \quad (6.24)$$

3. Obviously, an alternative starting point for the calculation of the coupling coefficients is given by Equation (6.20).

4. Within the framework of our discussion of beam propagation we have frequently used an analogous procedure, for example, in Section 4.1, to derive the paraxial Helmholtz equation.

5. By taking higher order corrections into account we obtain effective propagation constants  $\tilde{q}_0^{(m)}$  which are “dressed” by self-coupling.

Equation (6.24) can also be applied to other waveguide structures as a *rule of thumb*; the coupling coefficient  $\kappa_{mn}$  decreases usually exponentially with increasing spacing between waveguides  $m$  and  $n$ . We can further conclude that only the coupling coefficients between immediately adjacent waveguides will contribute significantly to the coupling. For this reason the coupling matrices of large coupled waveguide arrays will be sparse.

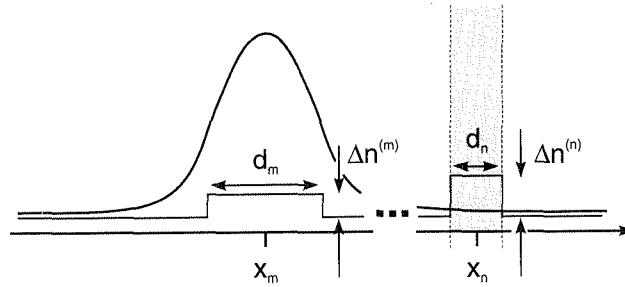


Figure 6.2. Coupling between waveguides  $m$  and  $n$  of a coupled waveguide array.

A great number of publications (see, e.g., [1, 2]) deals with the derivation of more sophisticated expressions for the coupling coefficients, often by considering the radiation field. From a practical point of view, however, it has become increasingly attractive to express the coupling coefficients  $\kappa_{mn}$  in terms of the supermodes (see, for example, Equation (6.20)) in order to take advantage of the flexible eigenmode solvers which are standard equipment in laboratories today. For this reason we will present in the next section an extremely simple formula for the coupling coefficients of directional couplers. This representation has the additional advantage of delivering the correct beat lengths irrespective of the strength of the coupling.

A very last remark at the end of this section: Most of our derivations concerning the coupling coefficients are independent of the underlying approximations; i.e., they can be carried out on the basis of a full vectorial, a scalar or another theory (e.g., the effective index method) which incorporates even more approximations. However, the calculation of coupling coefficients can be erroneous even though the propagation constants predicted by the theory are fairly correct. The reason is that the coupled mode theory deals with the spacing of eigenvalues and not with the eigenvalues themselves.

#### 6.1.4 Tapered Coupling

Up to now, we have discussed longitudinally uniform coupled waveguide arrays, i.e., bundles of parallel waveguides. Real devices require fan-in and fan-out regions where

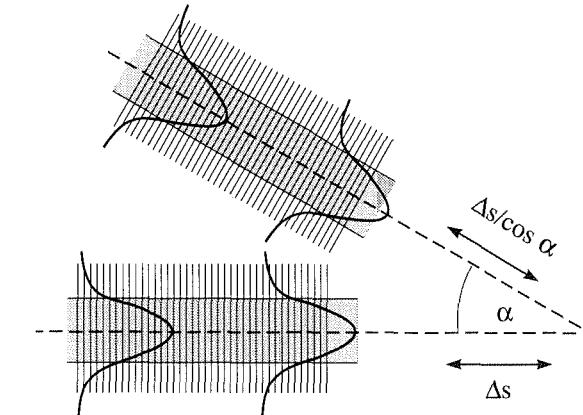


Figure 6.3. Coupling phenomena in a non-uniform coupled waveguide array.

the waveguides forming the coupled waveguide array are successively separated from each other. Thus, real coupled waveguide arrays are usually built up from non-uniform structures. A short glance at Figure 6.3 shows the physical phenomena which must be taken into account by a theory of tapered coupling. The most obvious effect is the tapering of the coupling coefficient  $\kappa_{mn}$ , i.e., the decrease in coupling with increasing spacing of the waveguides. In addition, the non-parallel paths of the waveguides result in a tilt of the phase fronts of the guided modes of the individual waveguides and in a phase shift due to the different optical light paths.

Roughly speaking, coupled mode theories for tapered coupling are hybrids between the coupled mode theory for longitudinally uniform structures and the local normal mode theory derived in Section 5.5. This means that all theories of this type divide the coupled waveguide array into a sequence of thin plates each describing a piece of the coupled waveguide array.

The most obvious way to derive a theory of tapered waveguide coupling is to formulate a local normal mode theory in terms of the local supermodes, i.e., in terms of the true local normal modes of the coupled waveguide array. This means that the optical field must be decomposed in the “no-coupling” region on the input side by the operator  $\mathcal{C}(z_{in})$  into the supermodes of the coupled waveguide array. The supermodes are then propagated through the waveguide structure by the operator  $\mathcal{U}_{LNM}(z_{in}, z_{out})$  describing the propagation of the local normal modes. On the output side the supermodes are decomposed into the eigenmodes of the individual waveguides by the operator  $\mathcal{C}^{-1}(z_{out})$ . In summary, the evolution of the optical field according to this theory is described by the propagator

$$\mathcal{U}_{CM}(z_{in}, z_{out}) = \mathcal{C}^{-1}(z_{out}) \mathcal{U}_{LNM}(z_{in}, z_{out}) \mathcal{C}(z_{in}). \quad (6.25)$$

Figure 6.4a illustrates the underlying mathematical procedure. Using Equations

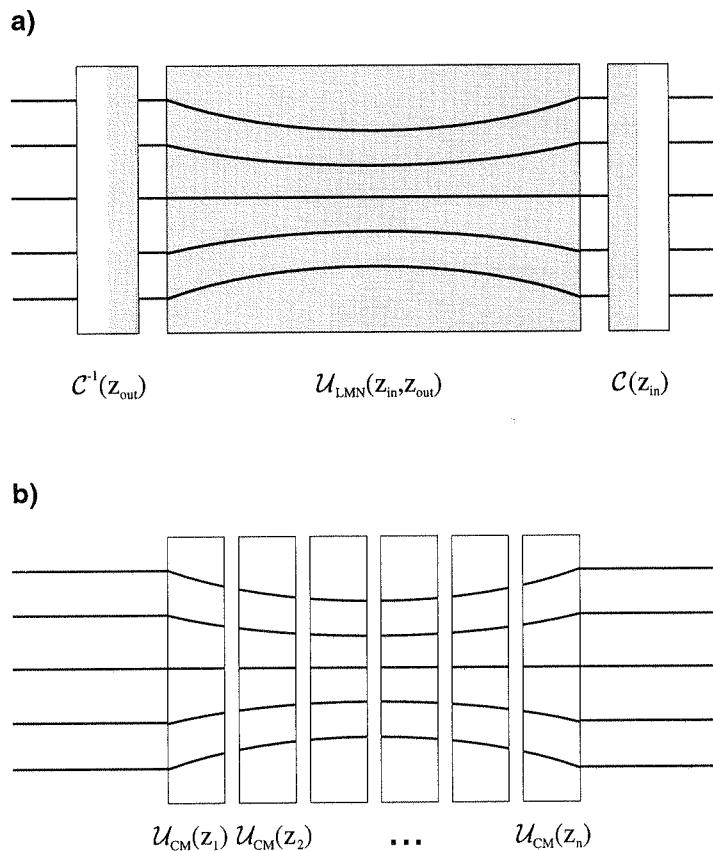


Figure 6.4. Tapered coupling described by the local supermodes (a) and the local guided modes of the individual waveguides (b).

(5.83) and (5.84) we see that the first iteration of the local normal mode propagator for the tapered coupling is given by a diagonal matrix

$$\mathcal{U}_{\text{LNM}}(z_{\text{in}}, z_{\text{out}}) = \begin{pmatrix} \exp\left(i \int_{z_{\text{in}}}^{z_{\text{out}}} dz q_0(z)\right) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp\left(i \int_{z_{\text{in}}}^{z_{\text{out}}} dz q_{N-1}(z)\right) \end{pmatrix}. \quad (6.26)$$

describing the overall phase shifts of the supermodes. We have seen in Section 5.5 that two local supermodes are converted into each other by the tapered structure

if the dielectric profile varies significantly over their beat length, i.e.,

$$\left\| \left\langle m \frac{\partial \epsilon}{\partial z} n \right\rangle \right\| \ll |q_n(z) - q_m(z)|.$$

Figure 6.4b shows an alternative way to describe tapered coupling by stacking

$$\mathcal{U}_{\text{CM}}(z_{\text{in}}, z_{\text{out}}) = \prod_{n=1}^{N_p} \mathcal{U}_{\text{CM}}(z_n) \quad (6.27)$$

the transfer matrices  $\mathcal{U}_{\text{CM}}(z_n)$  for the  $N_p$  individual plates. We should note that these transfer matrices operate on the guided modes of the local individual waveguides. This approach makes it easier to incorporate the effects of tilted waveguides. Proposals have been made [3] to tackle tilted waveguides by applying phase corrections

$$\Delta\varphi_n = q_0^{(n)} \Delta s^{(n)}(z)$$

which account after each propagation step for the relative path difference  $\Delta s^{(1)}(z)$  between the individual waveguides, i.e., the propagator for a single plate is replaced by

$$\mathcal{U}_{\text{CM}}^{(c)}(z) = \mathcal{U}_{\text{CM}}(z) \begin{pmatrix} \exp(i\Delta\varphi_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \exp(i\Delta\varphi_N) \end{pmatrix}. \quad (6.28)$$

We will see later on (see Section 6.5) that this type of propagator offers benefits for the description of networks consisting of coupled waveguide arrays and other guided wave components such as phase shifters, power splitters and combiners and X-switches.

Up to now, we have concentrated on the construction of propagators for tapered coupling. The main physical effect caused by the successive reduction of the coupling coefficients in the fan-in and fan-out regions of an asymmetrical coupled waveguide array is the reduction of the sidelobes in their filter curves. We will discuss this point in Section 6.4 for the example of an asymmetrical coupler in more detail.

## 6.2 UNIFORM DIRECTIONAL COUPLERS

We will start our discussion of codirectional couplers with an investigation of the most elementary example, the longitudinally uniform directional coupler. This device comprises only two single-mode waveguides, so its coupled mode analysis must consider only two guided modes.<sup>6</sup> We have seen in the previous section that the

6. Strictly speaking, a single-mode waveguide has two guided vectorial modes, and for weakly guiding waveguide structures these are TE and TM modes. Since most directional couplers do not couple TE and TM modes or their vectorial equivalents to each other it is sufficient to study the coupling of each set of guided modes separately.

coupled mode equations of this device are given by

$$-i \frac{d}{dz} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \tilde{q}_0^{(1)} & \kappa_{12} \\ \kappa_{21} & \tilde{q}_0^{(2)} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}. \quad (6.29)$$

We recall that the expansion coefficients  $a_m(z)$  describe the local decomposition of the optical field into the eigenmodes of the individual waveguides forming the coupler. The parameters  $\tilde{q}_0^{(m)}$  stand for the effective propagation constants of the individual waveguides, and the tilde above this symbol indicates that the propagation constants of the individual waveguides are “dressed”; i.e., corrected by the influence of the other waveguide (see Section 6.1.3). The parameters  $\kappa_{12}$  and  $\kappa_{21}$  describe the coupling of the waveguides to each other, and we saw in Section 6.1.2 that the directional coupler conserves the optical power if the coupling coefficients satisfy the symmetry relation  $\kappa_{21} = \kappa_{12}^*$ .

### 6.2.1 Solution of the Coupled Mode Equations

In a first step we will derive a closed expression for the coupled mode propagator  $\mathcal{U}_{CM}$ . For this purpose we replace the rapidly varying coefficients  $a_m(z)$  by their slowly varying “amplitudes”:

$$\alpha_m(z) = a_m(z) e^{-i\bar{q}z} \quad (6.30)$$

where

$$\bar{q} = \frac{\tilde{q}_0^{(1)} + \tilde{q}_0^{(2)}}{2} \quad (6.31)$$

designates the mean propagation constant. Straightforward calculations yield the transformed coupled mode equations

$$-i \frac{d}{dz} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \delta & \kappa_{12} \\ \kappa_{21} & -\delta \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}. \quad (6.32)$$

in which

$$\delta = \frac{\tilde{q}_0^{(1)} - \tilde{q}_0^{(2)}}{2} \quad (6.33)$$

describes the detuning between the effective propagation constants of the individual waveguides. The deviations of the propagation constants of the supermodes, i.e., of the eigenmodes of the compound waveguide structure forming the directional coupler, from the reference value  $\bar{q}$  can be calculated by solving the characteristic equation

$$\begin{vmatrix} \bar{q} + \delta - q & \kappa_{12} \\ \kappa_{21} & \bar{q} - \delta - q \end{vmatrix} = 0.$$

The deviations of the propagation constants  $q_0$  and  $q_1$  of the supermodes

$$\begin{aligned} q_0 &= \bar{q} + \delta_{eff} \\ q_1 &= \bar{q} - \delta_{eff} \end{aligned} \quad (6.34)$$

can be understood as an effective detuning

$$\delta_{eff} = \sqrt{\delta^2 + \kappa_{21}\kappa_{12}}$$

of the supermodes.

On the basis of the results of the eigenmode analysis, we can build up the fundamental solutions of the coupled mode equations, and from these the transfer matrix of the coupler, i.e.,

$$\tilde{\mathcal{U}}_{CM}(z) = \begin{pmatrix} \cos(\delta_{eff}z) + i\delta \sin(\delta_{eff}z) / \delta_{eff} & i\kappa_{12} \sin(\delta_{eff}z) / \delta_{eff} \\ i\kappa_{21} \sin(\delta_{eff}z) / \delta_{eff} & \cos(\delta_{eff}z) - i\delta \sin(\delta_{eff}z) / \delta_{eff} \end{pmatrix} \quad (6.35)$$

governing the evolution of the amplitudes  $\alpha_m(z)$ . It is related to the full propagator describing the evolution of the rapidly varying coefficients  $a_m(z)$  by

$$\mathcal{U}_{CM}(z) = e^{i\bar{q}z} \tilde{\mathcal{U}}_{CM}(z).$$

For the rest of this section we will restrict ourselves to lossless devices; i.e., we assume  $\kappa = \kappa_{21} = \kappa_{12}^*$  and in addition  $\text{Im}(\kappa) = 0$ . The transfer matrix  $\tilde{\mathcal{U}}_{CM}(z)$  for a lossless directional coupler is given by the matrix

$$\tilde{\mathcal{U}}_{CM}(z) = \begin{pmatrix} A^\ominus & A^\otimes \\ -A^{\otimes*} & A^{\ominus*} \end{pmatrix} \quad (6.36)$$

with the two coefficients

$$\begin{aligned} A^\ominus &= \cos(\delta_{eff}z) + i\delta \sin(\delta_{eff}z) / \delta_{eff} \\ A^\otimes &= i\kappa \sin(\delta_{eff}z) / \delta_{eff}. \end{aligned} \quad (6.37)$$

For a symmetrical device consisting of two waveguides of the same type ( $\delta = 0$ ) the expressions for the coefficients simplify to

$$\begin{aligned} A^\ominus &= \cos(|\kappa| z) \\ A^\otimes &= i \sin(|\kappa| z). \end{aligned} \quad (6.38)$$

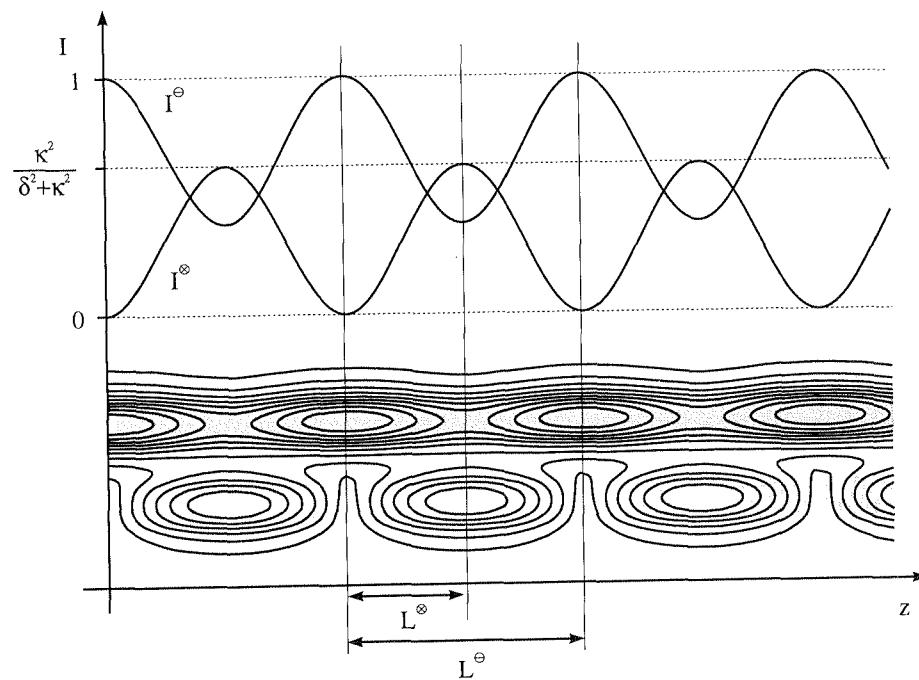


Figure 6.5. Beam propagating through a directional coupler.

### 6.2.2 Physical Properties of Directional Couplers

For most applications, the optical power is launched into the directional coupler via one of the ports at the input side of the device (see Figure 6.5). The directional coupler will then distribute the input power between the two ports on the output side of the device. The filter curve for the “bar” port, i.e., the relative optical power at the end of the input waveguide, is given by

$$I^\Theta = |A^\Theta|^2 = 1 - \left| \frac{\kappa}{\delta_{\text{eff}}} \right|^2 \sin^2(\delta_{\text{eff}}L) \quad (6.39)$$

in which  $L$  designates the length of the directional coupler. The response at the “cross” port is given by

$$I^\otimes = |A^\otimes|^2 = \left| \frac{\kappa}{\delta_{\text{eff}}} \right|^2 \sin^2(\delta_{\text{eff}}L). \quad (6.40)$$

First of all, we recover the conservation law of the optical power in the form

$$I^\Theta + I^\otimes = 1.$$

We see furthermore that the directional coupler is in a perfect bar state ( $I^\otimes = 0$ ) if its length represents any multiple of the beat length

$$L^\Theta = \frac{\pi}{\delta_{\text{eff}}}, \quad (6.41)$$

and is in its best cross state for the coupling length

$$L^\otimes = \frac{L^\Theta}{2}. \quad (6.42)$$

Only symmetrical directional couplers ( $\delta = 0$ ) exhibit a perfect cross state ( $I^\otimes = 0$ ).

The phase difference between two signals leaving the device via the two output ports – another quantity which is of some interest – is given by

$$\varphi^\otimes - \varphi^\Theta = -\tan^{-1} \left( \frac{\delta_{\text{eff}}}{\delta} \cot(\delta_{\text{eff}}L) \right). \quad (6.43)$$

We should note here that the filter curves  $I^\Theta$  and  $I^\otimes$  for the bar and cross ports do not depend on the choice of input port; i.e., the filter curve for input and output through waveguide 1 will look like that for input and output through waveguide 2. The lower part of Figure 6.5 shows for illustration contours of the optical field inside a directional coupler which was excited via one port. The evolution of the response functions  $I^\Theta$  and  $I^\otimes$  along the coupler is indicated by the diagrams on top of the BPM plot. The coupling length  $L^\otimes$  describing the length for one crossover and the beat length  $L^\Theta$  describing the coupling there and back are also shown.

We see that the directional coupler oscillates between a perfect bar state  $I^\otimes = 0$  and the best cross state  $I^\otimes = |\kappa/\delta_{\text{eff}}|^2$  there and back. As pointed out before, perfect crossover can be obtained only if the effective detuning  $\delta_{\text{eff}}$  equals to the coupling coefficient  $|\kappa| = |\delta_{\text{eff}}|$ , i.e., only for vanishing detuning  $\delta$ . Symmetrical couplers will always satisfy this condition. *Asymmetrical directional couplers, however, will exhibit no detuning only at the crossover of the dispersion curves of the individual waveguides.* For completely dispersionless devices we cannot use the crossover since it is located at an infinite wavelength  $\lambda \rightarrow \infty$ . However, we will show in Section 6.3 that devices with waveguide dispersion can exhibit a crossover at finite wavelengths. In Section 6.4 we will demonstrate that periodic structures offer an additional way to shift the crossover wavelength to the spectral region of operation.

### 6.2.3 Coupling Coefficients

We have shown in Section 6.1.3 how to derive expressions for the coupling coefficients  $\kappa_{mn}$  from first principles, i.e., on the basis of the eigenmodes of the individual waveguides, and we have seen that it is difficult to elaborate the corresponding overlap integrals. Therefore, we will now show two alternative ways to calculate the coupling coefficients.

The first calculation scheme makes use of the supermodes of the directional coupler; i.e., it requires a flexible eigenmode solver for the analysis of the compound waveguide structure. Using Equation (6.34) we see easily that the effective detuning is given by

$$\delta_{\text{eff}} = \frac{q_0 - q_1}{2} \quad (6.44)$$

half the spacing between the two supermodes. For symmetrical directional couplers the effective detuning and the coupling coefficient coincide ( $\delta_{\text{eff}} = |\kappa|$ ), and the phase factor of the coupling coefficient can be freely chosen. For the asymmetrical coupler we must in addition determine the detuning  $\delta$  of the individual waveguides, which is given by

$$\delta = \frac{\tilde{q}_0^{(1)} - \tilde{q}_0^{(2)}}{2} \quad (6.45)$$

the spacing of the corrected propagation constants  $\tilde{q}_0^n$  of the individual waveguides. For weakly coupled waveguides, we can replace the corrected propagation constants by the true propagation constants of the individual waveguides. For strongly coupled waveguides, we have to account for the influence of the other waveguide by

$$\begin{aligned} \tilde{q}_0^{(1)} &\approx q_0^{(1)} + k_0 \langle 0^{(1)} | \Delta n^{(2)} | 0^{(1)} \rangle \\ \tilde{q}_0^{(2)} &\approx q_0^{(2)} + k_0 \langle 0^{(2)} | \Delta n^{(1)} | 0^{(2)} \rangle. \end{aligned} \quad (6.46)$$

For step-index waveguide structures, the overlap integrals are given by the corresponding confinement factors multiplied by the refractive index contrast of the other waveguide. For a derivation of Equation (6.46) see Section 6.1.3.

A second more phenomenological approach is based on a BPM analysis of the coupler. Using Equation (6.41) we can derive the effective detuning from the beat length

$$\delta_{\text{eff}} = \frac{\pi}{L^\ominus}, \quad (6.47)$$

the coupling coefficient can then be determined from the maximum crossover  $I_{\max}^\otimes$  via Equation (6.40), i.e.,

$$|\kappa|^2 = |\delta_{\text{eff}}|^2 I_{\max}^\otimes. \quad (6.48)$$

From a purely theoretical point of view both approaches are unsatisfactory since they require knowledge about the compound structure for the calculation of the coupling coefficients. Nevertheless, they are useful from a more practical point of view since the required design tools are widespread today.

#### 6.2.4 Transfer Matrix of the Directional Coupler

The transfer matrices will turn out to be powerful tools for the treatment of compound directional couplers. We will now briefly list some properties of these matrices. We should note here once more that the following formulas apply to *lossless* directional couplers.

#### General Properties

The transfer matrix describing a compound directional coupler is a unimodular unitary  $2 \times 2$  matrix, i.e.,  $\det \tilde{\mathcal{U}}_\prec(z) = 1$ .

The unimodular unitary  $2 \times 2$  matrices form the group  $SU(2)$ , and it is sufficient to prove this property for the transfer matrix describing a uniform directional coupler, which is straightforward. As a consequence, we can use many of the results for unimodular matrices derived in Section 2.2.4.

Since the transfer matrix of a directional coupler is unitary, i.e.,

$$\tilde{\mathcal{U}}_\prec^{-1}(z) = \tilde{\mathcal{U}}_\prec^\dagger(z), \quad (6.49)$$

its inverse is given by

$$\tilde{\mathcal{U}}_\prec^{-1}(z) = \tilde{\mathcal{U}}_\prec(-z). \quad (6.50)$$

The absolute value of the trace of the transfer matrix

$$|\text{Tr } \tilde{\mathcal{U}}_\prec^{-1}(z)| \leq 2 \quad (6.51)$$

is less than or equal to 2; i.e., a directional coupler can never exhibit stopbands. See Section 2.2.4 for a more detailed discussion of this point.

The proof of the first two formulas is straightforward. For the derivation of the third, we use the representation of the trace by the sum over the eigenvalues of a matrix and obtain

$$|\text{Tr } \tilde{\mathcal{U}}_\prec^{-1}| = 2 |\text{Re}(A^\ominus)| \leq 2 |A^\ominus| \leq 2.$$

#### Compound Directional Couplers

The transfer matrix of a compound directional coupler is given by

$$\tilde{\mathcal{U}}_\prec(z) = \begin{pmatrix} A^\ominus & A^\otimes \\ -A^{\otimes*} & A^{\ominus*} \end{pmatrix}, \quad (6.52)$$

its eigenvalues are pure phase shifts, i.e.,

$$\delta_{1/2} = e^{\pm i\varphi} \quad (6.53)$$

with

$$\varphi = \tan^{-1} \left\{ \frac{\sqrt{1 - \text{Re}(A^\ominus)^2}}{\text{Re}(A^\ominus)} \right\}.$$

The coefficients of the corresponding transfer matrix satisfy the relation

$$|A^\ominus|^2 + |A^\otimes|^2 = 1. \quad (6.54)$$

Obviously, we must prove these assertions only for a stack of two different directional couplers. The derivation of Equation (6.52) is straightforward. For the derivation of the other two equations, we must account for the fact that  $\tilde{\mathcal{U}}_\prec(z)$  belongs to the group  $SU(2)$ .

### Uniform Directional Couplers

The transfer matrix of a uniform directional coupler represents a periodic function

$$\tilde{U}_{\times}(z + 2L^{\ominus}) = \tilde{U}_{\times}(z) \quad (6.55)$$

whose period is given by twice the beat length. The transfer matrix describing a coupler of beat length  $L^{\ominus}$  is given by

$$\tilde{U}_{\times}(L^{\ominus}) = - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad (6.56)$$

the matrix

$$\tilde{U}_{\times}(L^{\otimes}) = \frac{i}{\delta_{\text{eff}}} \begin{pmatrix} \delta & \kappa \\ \kappa^* & -\delta \end{pmatrix} \quad (6.57)$$

stands for a coupler of coupling length  $L^{\otimes}$ .

### Compound Symmetrical Couplers

A compound symmetrical coupler consisting of  $N$  uniform couplers of lengths  $L^{(n)}$  and coupling coefficients  $\kappa^{(n)}$  is described by the transfer matrix

$$\tilde{U}_{\times} = \begin{pmatrix} \cos \varphi & i \sin \varphi \\ i \sin \varphi & \cos \varphi \end{pmatrix} \quad (6.58)$$

with

$$\varphi = \sum_{n=1}^N |\kappa^{(n)}| L^{(n)}.$$

The quantity  $\varphi$  can be regarded as the overall phase thickness of the symmetrical directional coupler.

Obviously, it is sufficient to derive this formula for a stack consisting of two symmetrical directional couplers. A symmetrical directional coupler of a quarter-beat length acts as a 1 : 1 power splitter. Its transfer matrix is given by

$$\tilde{U}_{\times}\left(\frac{L^{\ominus}}{4}\right) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}. \quad (6.59)$$

### 6.3 SYMMETRICAL COUPERS

In this section we will study the spectral properties of a symmetrical directional coupler. We recall from our discussion of uniform couplers (see the previous section) that a symmetrical coupler always allows for a perfect cross state, i.e., for transfer of all optical power to the “cross” port. This is a consequence of the device symmetry,

which maintains the detuning  $\delta = 0$  over the full spectral range. We have further seen that the coupling coefficient

$$|\kappa| = \frac{q_0 - q_1}{2} = k_0 \frac{n_0(k_0) - n_1(k_0)}{2} \quad (6.60)$$

is given by half the spacing of the propagation constants  $q_m$  of the supermodes of the directional coupler. The second equation shows an alternative representation of the coupling coefficient in terms of the wave number  $k_0$  and the effective refractive indices  $n_m$  of the supermodes. We see that the coupling coefficient of a dispersionless coupler, i.e., of a device without material and waveguide dispersion ( $n_0(k_0) - n_1(k_0)$ ), will increase linearly with increasing wave number, material and waveguide dispersion and will consequently result in non-linearities of the coupling coefficients.

For our further discussion we consider a “real” device, a symmetrical coupler with fan-in and fan-out regions for the separation of its waveguides. We saw in the previous section (see Equation (6.58)) that the transfer matrix describing a compound symmetrical coupler is given by

$$\tilde{U}_{\times} = \begin{pmatrix} \cos \varphi & i \sin \varphi \\ i \sin \varphi & \cos \varphi \end{pmatrix}. \quad (6.61)$$

The overall phase thickness of the device comprising a continuously varied waveguide structure

$$\varphi = \int_{-L/2}^{L/2} dz |\kappa(z)| \quad (6.62)$$

is an integral over the coupling coefficient which varies along the coupler. Using the transfer matrix we obtain

$$I^{\otimes} = \sin^2 \varphi \quad (6.63)$$

for the filter characteristic of the device. We can conclude that *the symmetrical directional coupler has infinitely many perfect cross states which are found as solutions of the equation*

$$\varphi(k_0^{\otimes}) = (2m - 1)\frac{\pi}{2}. \quad (6.64)$$

*It has also infinitely many bar states at*

$$\varphi(k_0^{\ominus}) = m\pi. \quad (6.65)$$

In practice, only a finite number of these bar and cross states can be used, since the fundamental conditions for a correct operation of the coupler – exactly two guided supermodes for the coupler and one guided mode for each of its individual waveguides – is satisfied only within a finite spectral range.

In the following, we will consider a symmetrical coupler which shall be operated as an optical filter at two different wave numbers. At one of these wave numbers  $k_0^{\otimes}$

the device is assumed in a bar state with  $n$  beats inside it. For the analysis of device operation in the spectral region around  $k_0$  we can replace the overall phase thickness  $\varphi$  by the leading terms of its Taylor expansion

$$\varphi(k_0) \approx n\pi + \frac{d\varphi}{dk_0} \Big|_{k_0^\ominus} (k_0 - k_0^\ominus). \quad (6.66)$$

We further assume that the device operates at the wave number  $k_0^\otimes$  in a cross state with  $2m - 1$  beats inside it. To guarantee the desired operation at both wave numbers, the derivative of the phase thickness at  $k_0^\ominus$  must satisfy the condition

$$\frac{d\varphi}{dk_0} \Big|_{k_0^\ominus} = \left( m - n - \frac{1}{2} \right) \frac{\pi}{k_0^\otimes - k_0^\ominus}. \quad (6.67)$$

This condition in fact represents a severe restriction of design space since it requires special material compositions and/or waveguide geometries for a symmetrical directional coupler which often turn out to be incompatible with the requirements of other devices located on the same chip.

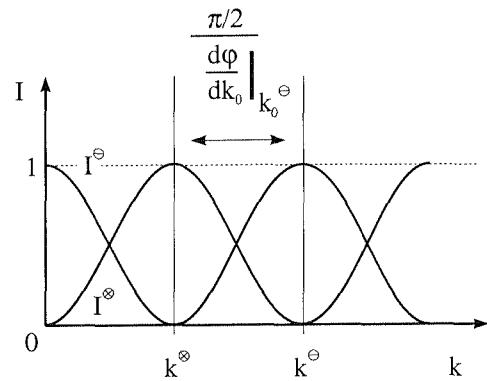


Figure 6.6. Filter curves of a symmetrical directional coupler.

Figure 6.6 shows for illustration the filter curves of a symmetrical directional coupler which exhibits neither material nor waveguide dispersion; i.e., the spacing of the effective refractive indices  $n_0(k_0) - n_1(k_0)$  remains constant. As indicated in this figure, usual designs aim at putting the wave numbers  $k_0^\otimes$  and  $k_0^\ominus$  on adjacent extrema of the filter curve, i.e.,  $m = n$  or  $m = n + 1$ , in order to increase both spectral and fabrication tolerances.

A few remarks on the dispersion properties of the coupling coefficients are ap-

propriate. The derivative of the overall phase thickness of the coupler is given by

$$\frac{d\varphi}{dk_0} = \int_{-L/2}^{L/2} dz \left[ \frac{n_0 - n_1}{2} + \frac{k_0}{2} \left( \frac{\partial n_0}{\partial k_0} - \frac{\partial n_1}{\partial k_0} \right) \right] \quad (6.68)$$

where  $n_0$  and  $n_1$  designate the local effective refractive indices of the supermodes. The first term on the right-hand side of Equation (6.68) accounts for the linear dispersion of the supermodes of the coupler. Since directional couplers are mostly used as filters for large channel spacings it is usually necessary to include the effects of material and waveguide dispersion described by the second term on the right-hand side of Equation (6.68). Waveguide dispersion can lead to significant variations of the coupling coefficient, in particular at the transition between the weak coupling and strong coupling regimes of the directional coupler.

## 6.4 ASYMMETRICAL COUPLERS

We have seen in the previous section that a symmetrical directional coupler has many perfect bar and cross states along the wavelength axis. The asymmetrical device, in contrast, exhibits resonant behavior; i.e., perfect crossover can occur only for one wave number  $k_0^{(c)}$ . We will start our discussion with a phenomenological theory of this device. We will then continue with an analysis of the required waveguide structures. The end of this section we will examine the influence of  $\kappa$ -tapering on the filter curves of the device.

### 6.4.1 Filter Characteristic

We will start our discussion of asymmetrical couplers with a purely phenomenological analysis of their filter characteristic. For this purpose we consider a wave number  $k_0^{(c)}$  where the detuning  $\delta$  of the asymmetrical coupler vanishes, and we assume in addition that the coupling coefficient of the coupler does not vanish at this wave number. In the spectral region close to the crossover of the dispersion curves of the individual waveguides, we can expand the detuning

$$\delta \approx \nabla_k \delta \Delta k_0 \quad (6.69)$$

into a Taylor series. In Equation (6.69)  $\Delta k_0 = k_0 - k_0^{(c)}$  designates the wave number relative to the crossover of the dispersion curves. The detuning parameter

$$\nabla_k \delta = \left[ \frac{dn^{(1)}}{dk_0} - \frac{dn^{(2)}}{dk_0} \right]_{k_0^{(c)}} \quad (6.70)$$

is determined by the angle of intersection of the two dispersion curves which describe the behavior of the individual waveguides. In the same way we can replace the coupling coefficient  $\kappa$  by its Taylor expansion. We shall restrict ourselves here to a zeroth-order expansion; i.e., we will neglect the influence of the dispersion of the coupling coefficient on the filter characteristic of the coupler.<sup>7</sup>

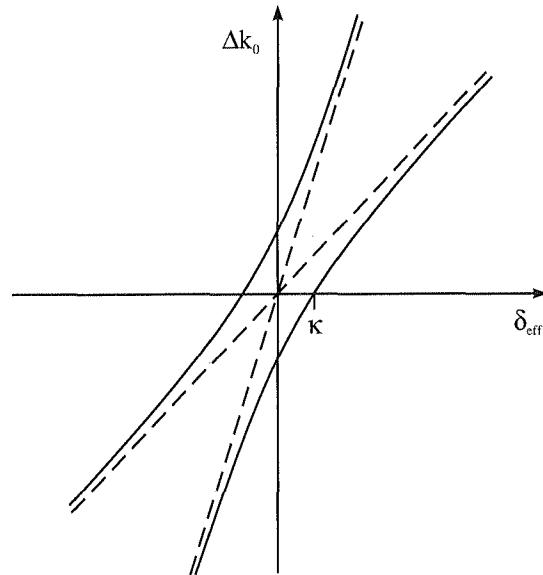


Figure 6.7. Dispersion curves of the asymmetrical directional coupler.

The effective detuning of the asymmetrical coupler in the region of the crossover of the dispersion curves is then given by

$$\delta_{\text{eff}}(\Delta k_0) \approx \sqrt{(\nabla_k \delta)^2 (\Delta k_0)^2 + |\kappa|^2}. \quad (6.71)$$

Figure 6.7 shows the dispersion diagram of a directional coupler. The dispersion curves of the individual waveguides, i.e., the filter characteristics of the directional coupler in the “no-coupling” limit, are indicated by dashed lines. We see that reasonable coupling of the modes occurs only if the detuning  $\delta$  is of the same order of magnitude as the coupling coefficient  $|\kappa|$ , i.e., in the spectral vicinity of the crossover of the dispersion curves of the individual waveguides.

By using Equation (6.71) we see that the effective detuning has a minimum at the wave number  $k_0^{(c)}$  where the detuning  $\delta$  vanishes. We can conclude from

7. This restriction has been made only for the sake of simplicity. If we take the dispersion of the coupling coefficient into account we end up with a slight asymmetry of the filter characteristic.

Equation (6.40) that the filter characteristic of an asymmetrical directional coupler

$$I^{\otimes}(k_0) = F(\Delta k_0) \sin^2(\delta_{\text{eff}} L). \quad (6.72)$$

exhibits resonant behavior. Its maxima

$$F(\Delta k_0) = \frac{(\chi_c/2)^2}{(\chi_c/2)^2 + (\Delta k_0)^2}. \quad (6.73)$$

are located on an envelope of Lorentzian shape. The full-width half-maximum (FWHM) of this envelope

$$\chi_c = 2 \left| \frac{\kappa}{\nabla_k \delta} \right|$$

is governed by the ratio of the coupling coefficient  $\kappa$  to the detuning parameter  $\nabla_k \delta$  defined previously.<sup>8</sup> In the next paragraph of this section we will see how to influence the parameters of the couplers by choosing appropriate waveguide structures.

Using the filter characteristic (6.72) we see that the directional coupler exhibits a perfect cross state at  $k_0^{(c)}$  if its length is given by  $(2n - 1)L^{\otimes}$ , and

$$L^{\otimes} = \frac{\pi}{2|\kappa|} \quad (6.74)$$

designates the coupling length of the asymmetrical coupler. The immediately adjacent bar state is found at

$$\Delta k_0^{\ominus} = \frac{\sqrt{4n - 1}}{2n - 1} \frac{\chi_c}{2}. \quad (6.75)$$

As expected, the spectral spacing to the next bar state decreases with an increasing number of beats. We should note, however, that the envelope  $F(\Delta k_0)$  is not affected by the choice of device length; i.e., we can reduce the channel spacing by lengthening the device only at the expense of higher sidelobes.

Figure 6.8 shows the filter characteristic of an asymmetrical directional coupler of length  $L^{\otimes}$  in the spectral vicinity of the crossover of the dispersion curves of the individual waveguides.

#### 6.4.2 Coupling Coefficients

In the first part of this section we have examined asymmetrical couplers from a purely phenomenological point of view. We now will show how to control the spectral position of the crossover of the dispersion curves by an appropriate layout of the coupler.

8. The full width half maximum of the envelope of a symmetrical coupler is infinite since  $\nabla_k \delta = 0$ . In physical terms, the symmetrical device exhibits no suppression of the sidelobes (see the previous section for a more detailed discussion of this device).

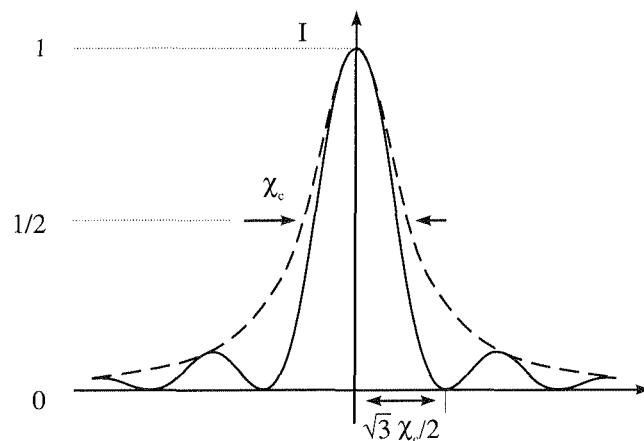


Figure 6.8. Filter characteristic of an asymmetrical directional coupler.

We will start with a discussion of the spectral properties of waveguides and will do so on the basis of a wavelength scale since this parameter accords better with our physical intuition. We consider an arbitrary step-index waveguide of a dielectric constant  $\epsilon_S$  which is embedded in background material of dielectric constant  $\epsilon_B$ . We have seen in Section 3.10 that such a waveguide will support at least one guided mode. For vanishing wavelengths  $\lambda \rightarrow 0$ , however, it will support infinitely many guided modes, and the effective dielectric constant  $\epsilon_0$  of its fundamental mode will approach the dielectric constant of the waveguide. With increasing wavelength, the guidance of the waveguide will decrease, and the effective dielectric constant  $\epsilon_0$  will successively approach the dielectric constant of the background material. To obtain more information about the “speed” of this process, we go back to the representation

$$\epsilon_0 = \langle 0 | \mathcal{H} | 0 \rangle \quad (6.76)$$

of the effective dielectric constant of the fundamental mode in terms of the corresponding mean value of the Hamiltonian. We see that the decrease of the effective dielectric constant is accompanied by a successive spread of its near-field over the region filled with background material. Therefore, we expect that *a waveguide of small cross section will approach the background dielectric constant faster than a device of large cross section*.

We can draw several further conclusions from the preceding discussion. First, *the dispersion curves of two step-index waveguides, which consist of the same material but different geometry, do not intersect at all*. Such structures are obviously useless for uniform asymmetrical couplers. *Devices which exhibit the desired operation, in contrast, must be assembled from one high-contrast waveguide of small cross section and one low-contrast waveguide of large cross section*. Figure 6.9 shows the dispersion curves of two waveguides which could be used to assemble an asymmet-

rical directional coupler. We see that the wavelength of crossover decreases with increasing “mismatch” of the waveguide properties. We should note that the shift of the crossover is usually accompanied by a change in the angle of intersection and thus by a change in the width of the optical filter (see Equations (6.70) and (6.73)). The required single-mode operation of both waveguides at the crossover of the dispersion curves results in an additional boundary condition for the choice of waveguide material and geometry. The material dispersion, which was completely neglected so far, will naturally affect the spectral position of the crossover. However, as long as both waveguides are based on the same material system (e.g., InGaAsP/InP) the general scenario shown in Figure 6.9 will be conserved.

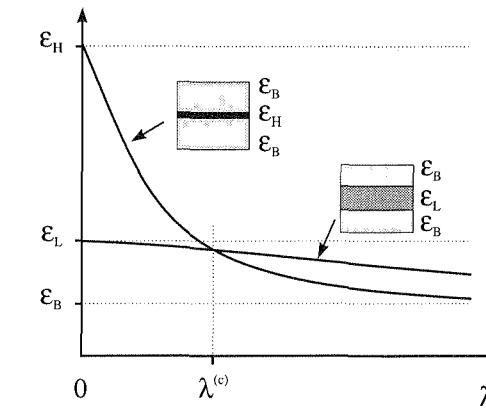


Figure 6.9. Crossover of two dispersion curves.

#### 6.4.3 Tapered Coupling

For the analysis of tapered coupling it is useful to replace the coupled mode equations by a first-order non-linear differential equation. For the derivation of this equation we introduce the ratio

$$\tau_\alpha(z) = \frac{\alpha_2(z)}{\alpha_1(z)} \exp \left( 2i \int_0^z d\tilde{z} \delta(\tilde{z}) \right) \quad (6.77)$$

of the amplitudes of the eigenmodes of the individual waveguides. For our further discussion<sup>9</sup> we imagine a directional coupler which is excited via waveguide 1 (see

9. If the waveguide is excited via waveguide 2 we can analyze the evolution of the optical field by means of the inverse ratio of amplitudes  $1/\tau_\alpha(z)$ .

Figure 6.10). That is, we assume that the ratio of amplitudes vanishes at the beginning of the directional coupler at  $z = -L/2$ , i.e.,  $\tau_\alpha(-L/2) = 0$ .

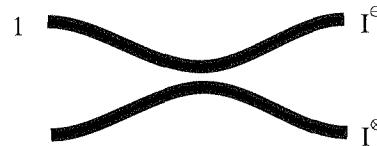


Figure 6.10. Tapered coupling.

We can then easily derive the differential equation for the evolution of the ratio of amplitudes

$$\frac{1}{\tau_\alpha} \frac{d\tau_\alpha}{dz} = \frac{1}{\alpha_2} \frac{d\alpha_2}{dz} - \frac{1}{\alpha_1} \frac{d\alpha_1}{dz} + 2i\delta$$

by using the coupled mode equations (see Equations (6.29)), and we obtain

$$\frac{d\tau_\alpha}{dz} = i\kappa(z) \exp\left(2i \int_0^z d\tilde{z} \delta(\tilde{z})\right) \left[1 - \tau_\alpha^2 \exp\left(-4i \int_0^z d\tilde{z} \delta(\tilde{z})\right)\right]. \quad (6.78)$$

Obviously, Equation (6.78) represents a non-linear first-order differential equation for the ratio of amplitudes, which means that the reformulation of the coupled mode equation was successful. It is a Riccati equation, and one can show that this equation has no general closed-form solution. However, in the spectral regions of marginal crossover ( $|\tau_\alpha| \ll 1$ ) we can derive an approximate solution of this differential equation since the quadratic term  $O(\tau_\alpha^2)$  on the right-hand side of Equation (6.78) is in these regions much smaller than the constant contributions. We can conclude that an approximate solution is then given by

$$\frac{d\tau_\alpha}{dz} \approx i\kappa(z) \exp\left(2i \int_0^z d\tilde{z} \delta(\tilde{z})\right). \quad (6.79)$$

However, this differential equation can be solved by integration. Using the initial condition  $\tau_\alpha(-L/2) = 0$  we obtain the following for the ratio of amplitudes at the end of the coupler

$$\tau_\alpha \approx i \int_{-L/2}^{L/2} dz \kappa(z) \exp\left(2i \int_0^z d\tilde{z} \delta(\tilde{z})\right). \quad (6.80)$$

If the detuning remains constant along the coupler, the result further simplifies to

$$\tau_\alpha \approx i \int_{-L/2}^{L/2} dz \kappa(z) \exp(2i\delta z). \quad (6.81)$$

In the regions outside the coupler  $|z| > L/2$  the coupling coefficient will also vanish; i.e., far away from the resonance, the response of the asymmetrical directional coupler with constant detuning

$$\tau_\alpha \approx i\mathcal{F}[\kappa]_{-2\delta} \quad (6.82)$$

is given by the Fourier transform of its coupling coefficient. In more technical terms; it is possible to suppress the sidelobes of an asymmetrical coupler by appropriate tapering. The classical window functions used for digital filtering are appropriate candidates for  $\kappa$ -tapering. Prominent examples are the Hamming, raised cosine, Blackman and Kaiser windows. As a rule of thumb [4] we should note here that all sidelobes except the first one are governed by the Fourier transform according to Equation (6.82).

## 6.5 TRANSFER MATRIX THEORY

In the last two sections of this chapter we have dealt with directional couplers and have formulated their properties in terms of the transfer matrices  $\tilde{\mathcal{U}}$ . This approach has turned out to be predestined for the description of composite waveguide structures such as tapered couplers. We will now derive compatible transfer matrices for other classes of elementary guided wave components and thus establish a general transfer matrix theory which can be applied to many problems of codirectional coupling. At the end of this section we will show how to assemble more complex networks by stacking the transfer matrices of their elementary constituents in an appropriate way.

### 6.5.1 Elementary Components

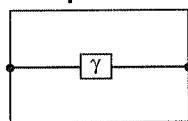
Figure 6.11 shows the transfer matrices for some elementary components. Before we start their derivation we recall that the transfer matrix  $\tilde{\mathcal{U}}$  is related to the full propagator  $\mathcal{U}$  via

$$\mathcal{U} = e^{i\bar{\varphi}} \tilde{\mathcal{U}} \quad (6.83)$$

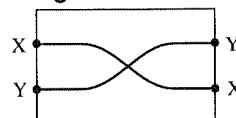
where  $\bar{\varphi}$  designates the “essential” phase shift<sup>10</sup> which is impressed onto a signal traveling through a device. For the coupled waveguide arrays, and in particular for the directional couplers,  $\bar{\varphi}$  was given by the mean phase shift, and, in consequence, the transfer matrix became a member of the unimodular unitary group  $SU(2)$ <sup>11</sup>

10. We will see in the following section that the description of networks will sometimes require a knowledge of the mean phase shift for the elementary components.

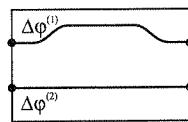
11. It is interesting to note here that the group  $SU(2)$  can be mapped onto the group  $SO(3)$  which describes the rotations in three-dimensional space. This means that the action of lossless 4-ports can always be represented by motions on the sphere  $S^2$  [5].

**Active and passive waveguide**

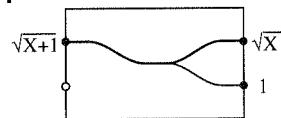
$$\tilde{U}_- = \sqrt{\gamma}$$

**Crossing**

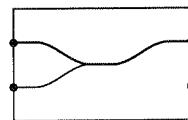
$$\tilde{U}_x = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

**Phase shifter**

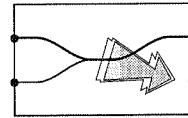
$$\tilde{U}_{\Delta\varphi} = \begin{pmatrix} e^{i\Delta\varphi/2} & 0 \\ 0 & e^{-i\Delta\varphi/2} \end{pmatrix}$$

**Splitter**

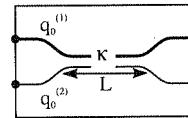
$$\tilde{U}_s = \frac{1}{\sqrt{X+1}} \begin{pmatrix} \sqrt{X} & 0 \\ 1 & 0 \end{pmatrix}$$

**Combiner**

$$\tilde{U}_r = \frac{1}{\sqrt{X+1}} \begin{pmatrix} \sqrt{X} & 1 \\ 0 & 0 \end{pmatrix}$$

**Combiner with radiation port**

$$\tilde{U}_z = \frac{i}{\sqrt{X+1}} \begin{pmatrix} \sqrt{X} & 1 \\ 1 & -\sqrt{X} \end{pmatrix}$$

**Directional coupler**

$$\delta = (q_0^{(1)} q_0^{(2)})/2$$

$$\delta_{\text{eff}} = \sqrt{\delta^2 + K^2}$$

$$\tilde{U}_x = \begin{pmatrix} A^\oplus & A^\otimes \\ -A^\otimes & A^\oplus \end{pmatrix}$$

$$A^\oplus = \cos(\delta_{\text{eff}} L) + i \frac{\delta}{\delta_{\text{eff}}} \sin(\delta_{\text{eff}} L)$$

$$A^\otimes = i \frac{K}{\delta_{\text{eff}}} \sin(\delta_{\text{eff}} L)$$

We will always try to choose the “essential” phase shift such that the corresponding transfer matrices are also members of the unimodular unitary group.

**Active and Passive Waveguides**

The single-mode waveguide is the only 2-port in our list of elementary components. The transfer factor describing the action of the device is simply given by

$$\tilde{U}_- = \sqrt{\gamma} \quad (6.84)$$

where  $\gamma$  designates the (power) amplification factor. For a lossless waveguide ( $\gamma = 1$ ) the transfer factor belongs formally to  $SU(1)$ . The essential phase shift of a longitudinally invariant waveguide of length  $L$  is given by

$$\Delta\varphi = q_0 L$$

where  $q_0$  stands for the effective propagation constant of the guided mode of the waveguide.

**Phase Shifter**

A phase shifter device consists of two completely decoupled waveguides. By implementing a physical path difference between both waveguides and/or by using different materials or waveguide geometries we can introduce a phase imbalance

$$\Delta\varphi = \Delta\varphi^{(1)} - \Delta\varphi^{(2)}$$

at the output ports. Obviously, the transfer matrix

$$\tilde{U}_{\Delta\varphi} = \begin{pmatrix} \exp(i\Delta\varphi/2) & 0 \\ 0 & \exp(-i\Delta\varphi/2) \end{pmatrix}, \quad (6.85)$$

describes the phase imbalance between the two signals at the end of the device. We should note that the transfer matrix of the phase shift is a member of the unimodular unitary group  $SU(2)$ , i.e.,  $\det \tilde{U}_{\Delta\varphi} = 1$ .

**Crossing**

A perfect waveguide crossing does nothing other than exchange the input ports. The transfer matrix describing its operation is given by

$$\tilde{U}_x = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}. \quad (6.86)$$

The waveguides forming a “real” crossing couple slightly in the regions of their crossover. Such devices are usually treated as directional couplers. The transfer matrix of the crossing is a member of the unimodular unitary group  $SU(2)$ , i.e.,  $\det \tilde{U}_x = 1$ .

Figure 6.11. Transfer matrices of some elementary guided wave components.

### Y-Branch Operated as a Power Splitter

The treatment of the Y-branch is slightly more complicated than the examples just treated, since this device acts as a 3-port and not as a 4-port. However, we will regard it as a 4-port by adding a fourth “grounded” port. The transfer matrix for a Y-branch which acts as a  $1 : X$  power splitter is then given by

$$\tilde{\mathcal{U}}_{\prec} = (\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} & 0 \\ 1 & 0 \end{pmatrix}. \quad (6.87)$$

This formulation assumes that the open input port is described by the first element of the amplitude vector (see Figure 6.11).

To prove the correctness of Equation (6.87) we will apply two linearly independent “test signals” to the Y-branch. If we launch unity power into the “grounded” input port

$$(\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = (\sqrt{X+1})^{-1/2} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

we see immediately that the transfer matrix does not deliver a response. If, in contrast, we launch unity power into the open input port, we obtain

$$(\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} \\ 1 \end{pmatrix}$$

two output signals whose powers have the extinction ratio  $1 : X$ . The coefficients of the transfer matrix are chosen such that the optical power is conserved. We can conclude that the handling of optical power is well described by the transfer matrix  $\tilde{\mathcal{U}}_{\prec}$ . To show that the balancing of the phase fronts is well described by the transfer matrix, we will follow a beam propagating through the Y-branch. On its input side the Y-branch supports only one guided supermode, and the total optical power is stored in that eigenmode. In the branching region the Y-branch begins to support a second guided supermode. However, the whole optical power is tapered into the fundamental supermode of the compound waveguide structure<sup>12</sup> which can be represented by the sum (and not by the difference)

$$|\phi_{\text{out}}\rangle = (\sqrt{X+1})^{-1/2} (\sqrt{X} |0^{(1)}\rangle + |0^{(2)}\rangle) \quad (6.88)$$

of the guided modes of the individual waveguides, i.e., there is no phase imbalance between the two output signals.

The transfer matrix of the Y-branch operating as a power splitter acts as a projector, i.e.,  $\det \tilde{\mathcal{U}}_{\prec} = 0$ .

### Y-Branch Operated as a Power Combiner

The transfer matrix for a Y-branch which acts as a  $1 : X$  power combiner is given by

$$\tilde{\mathcal{U}}_{\succ} = (\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} & 1 \\ 0 & 0 \end{pmatrix}. \quad (6.89)$$

This formulation assumes that the open output port is described by the first element of the amplitude vector (see Figure 6.11).

The most elegant way to verify the transfer matrix  $\tilde{\mathcal{U}}_{\succ}$  is to launch the two guided supermodes (see Equation (6.88)) into the Y-branch and control the correct operation of the device for these two examples. Applying the fundamental supermode as the input signal, we obtain

$$(\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} & 1 \\ 0 & 0 \end{pmatrix} (\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

i.e., the tapering of the fundamental supermode into the guided supermode of the single-mode region is recovered. The propagation of the excited supermode yields

$$(\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} & 1 \\ 0 & 0 \end{pmatrix} (\sqrt{X+1})^{-1/2} \begin{pmatrix} 1 \\ -\sqrt{X} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

This result describes the complete radiation of the excited supermode in the single-mode output region. We see furthermore that the “grounded” output port does not provide any signal irrespective of the input. The verification of the transfer matrix for these two input signals is sufficient, since corresponding vectors form a basis of the two dimensional vector space.

As for the power splitter, the transfer matrix of the Y-branch operating as a power combiner acts as a projector, i.e.,  $\det \tilde{\mathcal{U}}_{\succ} = 0$ .

### Y-Branch Equipped with a Virtual “Radiation” Port

We have seen in the preceding discussion that the transfer matrices describing the actions of a Y-branch as a power splitter and combiner are projectors, i.e.,  $\det \tilde{\mathcal{U}}_{\prec} = \det \tilde{\mathcal{U}}_{\succ} = 0$ . This mathematical property of the transfer matrix is a consequence of the radiation loss exhibited by Y-branches which are used as power combiners. We can therefore make the transfer matrix unitary by replacing the “grounded” port with a virtual “radiation” port which traps the radiated power. The transfer matrix of such a Y-branch power combiner is given by

$$\tilde{\mathcal{U}}_{\Sigma} = i (\sqrt{X+1})^{-1/2} \begin{pmatrix} \sqrt{X} & 1 \\ 1 & -\sqrt{X} \end{pmatrix}. \quad (6.90)$$

Obviously, this transfer matrix is a member of the unimodular unitary group ( $SU(2)$ ); i.e., we expect that the transfer matrix for the corresponding power splitter is given by

$$\tilde{\mathcal{U}}_{\Sigma} = \tilde{\mathcal{U}}_{\prec}. \quad (6.91)$$

12. See Section 5.5 for a more detailed analysis of the underlying tapering process.

A comparison with Equation (6.90) verifies this assumption.

The transfer matrix for the Y-branch which is equipped with a “radiation” port must be handled with great care. In particular, no optical power must be launched into a network via a “radiation” port and no radiated power must be launched into another port. Nevertheless, it is possible to keep a network unitary and benefit from this property at the expense of some virtual extra-ports.

Most of the devices listed in Figure 6.11 are non-symmetrical with respect to an exchange between the upper and lower ports, i.e., with respect to the transformation  $\alpha_1^{(in)} \Leftrightarrow \alpha_2^{(in)}$ ,  $\alpha_1^{(out)} \Leftrightarrow \alpha_2^{(out)}$ . The transfer matrix describing a device with exchanged upper and lower ports can be obtained directly from the results already listed by using the transformation

$$\tilde{\mathcal{U}}_{\ddagger} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tilde{\mathcal{U}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} \tilde{U}_{22} & \tilde{U}_{21} \\ \tilde{U}_{12} & \tilde{U}_{11} \end{pmatrix}. \quad (6.92)$$

### 6.5.2 Compound Components and Networks

The transfer matrices for the elementary components listed in the previous section form an appropriate toolbox for the construction of many compound components and even for integrated optical circuits, i.e., for more complex optical networks.

We will start our discussion of these topics with some remarks about the most complex “networks” treated so far, these being compound 4-ports. We recall (see Section 6.3) that we have already tackled the tapered symmetrical coupler as an infinite stack of uniform couplers each exhibiting a different spacing between its individual waveguides. We have assembled this device by simply stacking the transfer matrices of its constituents. The total phase shift caused by the building blocks will often be of minor interest, only the relative phase shifts between the two output signals are usually required for the further analysis of a device. This means that we should not worry too much about the mean phase shifts of the elementary components. Obviously, the same remarks apply to compound  $N$ -ports which consists of a stack of elementary  $N$ -ports. In the following sections of this chapter we will derive the stacked transfer matrices describing several types of Mach-Zehnder devices and directional couplers comprising periodic structures.

We will now turn our interest to networks of more complex topology. For the further analysis we will decompose such networks into stages which consist only of unconnected  $N$ -ports, for example 4-ports and 2-ports. A 4-port of such a stage can represent an elementary 4-port, i.e., one of the items in the list presented in the previous part of this section, or even a compound 4-port. In the same way the 2-ports can represent elementary components, i.e., active or passive waveguides, or compound devices. In contrast to the construction compound of  $2N$ -ports, it is now necessary to keep strict track of the different signals and in particular of their phase shifts relative to each other. Figure 6.12 shows for illustration one stage of a network

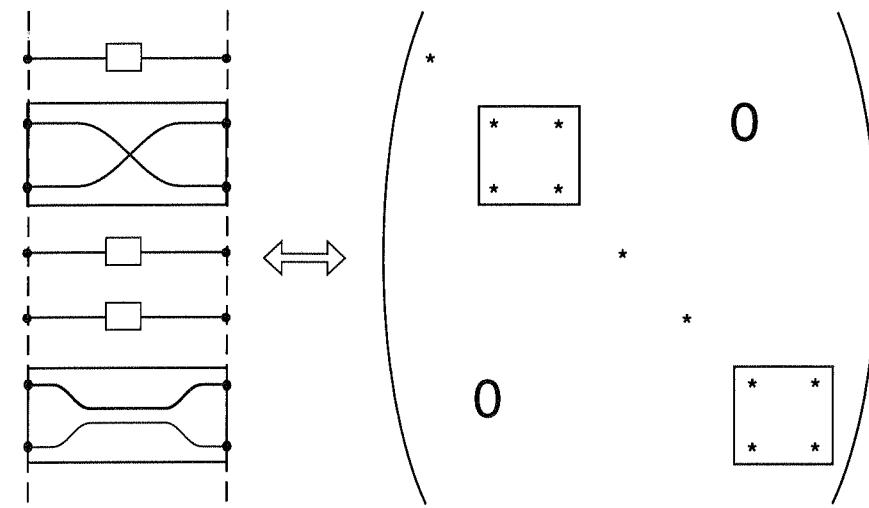


Figure 6.12. Transfer matrix for one stage of a  $2N$ -port optical network.

consisting of elementary 2-ports and 4-ports. We see that the corresponding transfer matrix exhibits a box diagonal form where the  $2 \times 2$ -boxes stand for the 4-ports, and the single diagonal elements describe the 2-ports. The determinant of one stage of the network consisting of  $P$  2- and 4-ports is given by

$$\det \mathcal{U} = \prod_{m=1}^P \exp(ip_m \Delta \varphi_m) \det \mathcal{U}_m \quad (6.93)$$

where  $p_m$  with  $\sum_{m=1}^P p_m = N$  designates half the number of ports in the subnetwork  $m$ . For our example

$$p_m = \begin{cases} 1 & \text{for 2-ports} \\ 2 & \text{for 4-ports} \end{cases}$$

The transfer matrix describing one stage of a network can be made unimodular unitary, i.e., a member of the group of unimodular unitary  $N \times N$  matrices  $SU(N)$ , if the transfer matrices describing its constituents are unitary and if all phase shifts are related to the mean phase shift of the actual stage, i.e.,

$$\sum_{m=1}^P p_m \Delta \varphi_m = 0. \quad (6.94)$$

The transfer matrix of the full network is then obtained by stacking the transfer matrices of its individual stages.

As a very simple tutorial example, let us now construct the general transfer matrix describing a  $1 : 4$  beam splitter built up from two stages of symmetrical

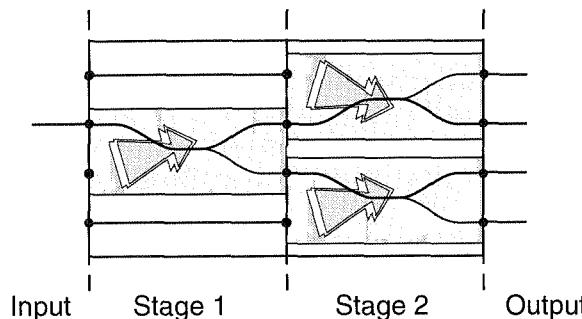


Figure 6.13. Transfer matrix for a 1 : 4 beam splitter.

Y-branches ( $X = 1$ ). Figure 6.13 shows a schematic diagram of the power splitter. Its transfer matrix for a structure equipped with radiation ports is given by

$$\begin{aligned}\tilde{\mathcal{U}} &= \frac{1}{2} \begin{pmatrix} -i & i & 0 & 0 \\ i & i & 0 & 0 \\ 0 & 0 & i & i \\ 0 & 0 & i & -i \end{pmatrix} \begin{pmatrix} \sqrt{2} & 0 & 0 & 0 \\ 1 & i & i & 0 \\ 0 & i & -i & 0 \\ 0 & 0 & 0 & \sqrt{2} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} -i\sqrt{2} & -1 & -1 & 0 \\ i\sqrt{2} & -1 & -1 & 0 \\ 0 & -1 & 1 & i\sqrt{2} \\ 0 & -1 & 1 & -i\sqrt{2} \end{pmatrix}. \quad (6.95)\end{aligned}$$

The transfer matrix describing the true 4-port is found by projecting the general transfer matrix  $\tilde{\mathcal{U}}$  to the input port 2 of the 8-port, i.e.,

$$\tilde{\mathcal{U}}_{1:4} = \tilde{\mathcal{U}} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = -\frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}. \quad (6.96)$$

As expected, the optical power launched into the device is equally distributed among the output ports, and all partial waves leaving the device are in phase.

Since the general 8-port is equipped with radiation ports we can use the inverse transfer matrix to derive equations for the complementary device, i.e., for the 4 : 1 power combiner. By choosing port 2 of the general 8-port as output we obtain

$$\tilde{\mathcal{U}}_{4:1} = (0 \ 1 \ 0 \ 0) \tilde{\mathcal{U}}^\dagger = -\frac{1}{2} (1 \ 1 \ 1 \ 1). \quad (6.97)$$

As expected, Equation (6.97) describes the projection of the input field onto the fundamental supermode of the waveguide array on the input side. If we, for example,

excite the power combiner via port 3 we obtain

$$-\frac{1}{2} (1 \ 1 \ 1 \ 1) \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} = -\frac{1}{2}. \quad (6.98)$$

We see that three quarters of the optical power are lost by radiation. By using Equation (6.98) we see easily that we obtain the same result if the optical power is launched via one of the other input ports.

We shall now make a few remarks on the conservation laws for optical networks. If the transfer matrices  $\tilde{\mathcal{U}}_m$  describing the stages of a compound  $2N$ -port are unitary, i.e., members of the unitary group  $U(N)$  or of its unimodular subgroup  $SU(N)$ , then the complete network described by the transfer matrix  $\tilde{\mathcal{U}}$  conserves the optical power. That is,

$$\sum_{m=1}^N |\alpha_m^{(in)}|^2 = \sum_{m=1}^N |\alpha_m^{(out)}|^2. \quad (6.99)$$

The response of the network to backward running signals is described by the adjoint matrix  $\tilde{\mathcal{U}}^\dagger$ , and the optical power is also conserved.

## 6.6 MACH-ZEHNDER DEVICES

Mach-Zehnder devices represent compound components consisting of a phase shifter which is embedded in two Y-branches or directional couplers. We will discuss three variants of this device here. The first one, the Mach-Zehnder interferometer, is a 2-port which allows phase modulations applied within the phase shifter to be translated into intensity modulations. The second and third devices, a 3-port and a 4-port Mach-Zehnder coupler, can be used as an optical filter or switch. The operation of all Mach-Zehnder devices is driven by the phase shifter, the Y-branches and/or directional couplers at the input and output serve only as power splitters and combiners.

### 6.6.1 Mach-Zehnder Interferometer

Figure 6.14a shows a schematic diagram of a Mach-Zehnder interferometer. We see that this device acts a 2-port, the information being impressed by phase modulation onto one of the arms of the phase shifter; i.e., the phase shift  $\Delta\varphi$  acts as a parameter describing the effect of modulation. The phase shifter of the Mach-Zehnder interferometer is embedded between two Y-branches. The transfer factor of the device is given by

$$\tilde{\mathcal{U}} = (0 \ 1) 1/\sqrt{2} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \exp(i\Delta\varphi/2) & 0 \\ 0 & \exp(-i\Delta\varphi/2) \end{pmatrix} 1/\sqrt{2} \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} (1 \ 0)$$

$$= \cos(\Delta\varphi/2). \quad (6.100)$$

The vectors at the beginning and end of the matrix product on the right-hand side of Equation (6.100) are used to extract the two active ports of the Mach-Zehnder interferometer. The center matrix stands for the phase shifter, while the other two matrices describe the Y-branches in front of and behind the phase shifter. The power response of the device is thus given by

$$I^\otimes = \cos^2(\Delta\varphi/2). \quad (6.101)$$

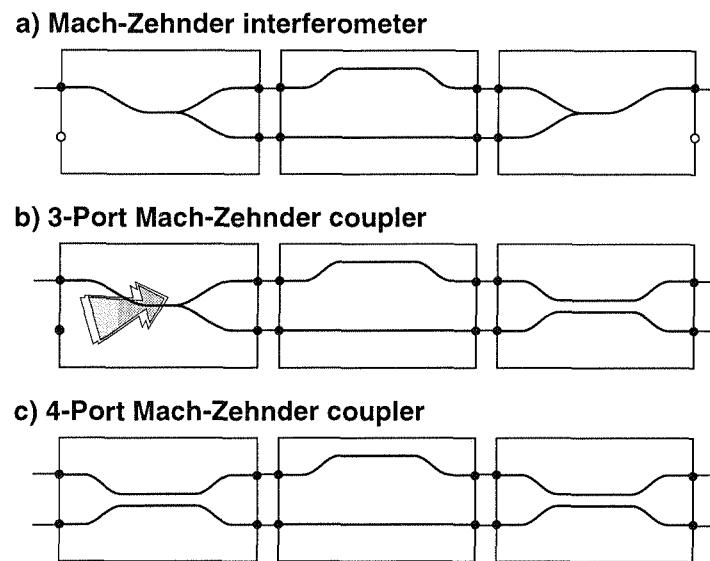


Figure 6.14. Mach-Zehnder interferometer (a), 3-port (b) and 4-port (c) Mach-Zehnder couplers.

Let us discuss the operation of the Mach-Zehnder interferometer in physical terms. We see that the incoming guided mode is tapered by the Y-branch into the fundamental supermode of the Y-branch on its output side. If a phase shift of  $\Delta\varphi = 2m\pi$  is applied, the fundamental supermode is launched into the Y-branch, which is used as a power combiner, and transformed into the guided mode of the output waveguide. In technical terms, the 2-port is open. If the phase shift equals an odd multiple of  $\pi$ , i.e.,  $\Delta\varphi = 2m\pi - 1$ , the first excited mode is launched into the power combiner and radiated later on: the 2-port is closed. Obviously, this is exactly the scenario described by Equation (6.101).

### 6.6.2 Mach-Zehnder Couplers

We will now discuss the two types of Mach-Zehnder couplers shown in Figure 6.14b and 6.14c for comparison. In contrast to the Mach-Zehnder interferometer treated in the first part of this section, these two devices are candidates for optical filters and switches.

Let us start the discussion with the 3-port. This device (see Figure 6.14b) consists of a phase shifter which is embedded between a Y-branch on the input side and a symmetrical coupler on the output side. The transfer matrix of the phase-quadrature coupler is given by

$$\begin{aligned} \tilde{\mathcal{U}} &= \begin{pmatrix} \cos\phi & i\sin\phi \\ i\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} \exp(i\Delta\varphi/2) & 0 \\ 0 & \exp(-i\Delta\varphi/2) \end{pmatrix} 1/\sqrt{2} \begin{pmatrix} i & i \\ i & -i \end{pmatrix} \\ &= \begin{pmatrix} A^\oplus & A^\otimes \\ -A^{\otimes*} & A^{\oplus*} \end{pmatrix} \end{aligned} \quad (6.102)$$

with the coefficients

$$\begin{aligned} A^\oplus &= \frac{1}{\sqrt{2}} [-\sin(\phi + \Delta\varphi/2) + i\cos(\phi - \Delta\varphi/2)] \\ A^\otimes &= \frac{1}{\sqrt{2}} [\sin(\phi - \Delta\varphi/2) - i\cos(\phi + \Delta\varphi/2)] \end{aligned}$$

In Equation (6.102) the parameter  $\phi$  stands for the phase thickness of the symmetrical coupler on the output side. The phase shift applied within the phase shifter is given by

$$\Delta\varphi = q_0\Delta L \quad (6.103)$$

provided the phase shifter consists of two waveguides of the same type. The propagation constant of the guided mode is then given by  $q_0$ , the physical path difference between the two waveguides of the phase shifter by  $\Delta L$ . We should note that the Y-branch is equipped with a “radiation” port in order to keep the transfer matrix of the compound device unitary. The relative optical power which is launched into the port belonging to the long waveguide of the phase shifter is given by

$$I^\oplus = \frac{1 + \sin(2\phi)\sin(\Delta\varphi)}{2}. \quad (6.104)$$

For the sake of simplicity we will designate this port as the bar port of the device. We assume for the further discussion that the symmetrical coupler acts approximately as a power splitter ( $\phi = \pi/4 + \Delta\phi$  with  $\Delta\phi \ll 1$ ). Then the best bar states are obtained for

$$\Delta\varphi^\oplus = (4m - 3)\frac{\pi}{2} \quad (6.105)$$

and the condition for the best cross states is given by

$$\Delta\varphi^\otimes = (4m - 1)\frac{\pi}{2}. \quad (6.106)$$

For the phase shift  $\Delta\phi = m\pi$  the device acts as a power splitter. We see that perfect bar and cross states can be achieved only if the symmetrical coupler on the output side is a perfect power splitter ( $\phi = \pi/4$ ). If this condition is violated, the minimum relative power level at both cross and bar ports is given by

$$\min(I^\otimes) = \min(I^\ominus) = \sin^2 \Delta\phi \approx \Delta\phi^2 \quad (6.107)$$

where  $\Delta\phi$  stands for the deviation of the phase thickness  $\phi$  of the directional coupler from the ideal value ( $\phi = \pi/4$ ). We see that crosstalk caused by the imperfections of the directional coupler is balanced for both output ports.

The 4-port Mach-Zehnder coupler (see Figure 6.14c) exhibits, as we will see, a different behavior. The transfer matrix of this device, which consists of a phase shifter embedded between two identical symmetrical couplers each of phase thickness  $\phi$ , is given by

$$\begin{aligned} \tilde{\mathcal{U}} &= \begin{pmatrix} \cos \phi & i \sin \phi \\ i \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} \exp(i\Delta\varphi/2) & 0 \\ 0 & \exp(-i\Delta\varphi/2) \end{pmatrix} \begin{pmatrix} \cos \phi & i \sin \phi \\ i \sin \phi & \cos \phi \end{pmatrix} \\ &= \begin{pmatrix} A^\ominus & A^\otimes \\ -A^{\otimes*} & A^{\ominus*} \end{pmatrix} \end{aligned} \quad (6.108)$$

with the coefficients

$$\begin{aligned} A^\ominus &= \cos(2\phi) \cos(\Delta\varphi/2) + i \sin(\Delta\varphi/2) \\ A^\otimes &= i \sin(2\phi) \cos(\Delta\varphi/2). \end{aligned}$$

The parameter  $\Delta\varphi$  again stands for the relative phase shift inside the phase shifter. The relative optical power at the cross port is then given by

$$I^\otimes = \sin^2(2\phi) \cos^2(\Delta\varphi/2), \quad (6.109)$$

i.e., the device exhibits cross states at

$$\Delta\varphi^\otimes = 2m\pi, \quad (6.110)$$

its bar states are found at

$$\Delta\varphi^\ominus = (2m-1)\pi. \quad (6.111)$$

We see that this device exhibits perfect bar and cross states if the directional couplers act as perfect power splitters ( $\phi = \pi/4$ ). For deviations  $\Delta\phi$  from this operating point, however, the bar state remains perfect

$$\min I^\otimes = 0, \quad (6.112)$$

and the cross state becomes imperfect, i.e.,

$$\min I^\ominus = \sin^2(2\Delta\phi) \approx 4\Delta\phi^2. \quad (6.113)$$

If we compare the results of the 3-port and 4-port Mach-Zehnder couplers we see that *as long as the built-in directional couplers represent perfect power splitters, both devices yield the same filter characteristic, i.e., several perfect bar and cross states along the wavelength axis*.<sup>13</sup> If the directional couplers become imperfect, however, the 3-port Mach-Zehnder coupler exhibits a decreasing extinction ratio which is balanced for both output ports. The 4-port Mach-Zehnder coupler, in contrast, maintains a perfect bar state at the expense of a degraded cross state.

If we compare the Mach-Zehnder couplers with their direct competitor, the symmetrical directional coupler, it looks at first glance as if the Mach-Zehnder devices had only drawbacks. After all, the symmetrical coupler exhibits perfect bar and cross states irrespective of the material parameters, whereas Mach-Zehnder couplers offer advantages for integration with other devices. Examples of this kind are cascaded filters or integration with completely different components. It also looks as if the fabrication of the specified phase shifters, i.e., the control of wavelengths, is easier than the control of fabrication tolerances for the symmetrical directional couplers.

## 6.7 DIRECTIONAL COUPLERS WITH PERIODIC OVERLAY

In the following we will analyze directional couplers with periodic variations of both detuning  $\delta$  and coupling coefficient  $\kappa$ . However, before we start our discussion we should clearly point out that this section deals with the influence of periodic structures on codirectional coupling, i.e., on the coupling of modes traveling in the same direction. A discussion of contradirectional coupling, the coupling of modes traveling in opposite directions via periodic structures, is the subject of the following chapter. We start with a derivation of the transfer matrix for a directional coupler with  $N$  periods and discuss the two most important examples of periodic couplers on the basis of the general analysis, the  $\Delta\beta$ -couplers which represent an appropriate structure for optical switches and the  $\Delta\kappa$ -couplers which are used as fixed or tunable filters.

### 6.7.1 Some Properties of Periodic Couplers

If a unimodular transfer matrix

$$\tilde{\mathcal{U}} = \begin{pmatrix} A^\ominus & A^\otimes \\ -A^{\otimes*} & A^{\ominus*} \end{pmatrix}$$

designates a single period of a periodic directional coupler then the transfer matrix

$$\tilde{\mathcal{U}}^N = \begin{pmatrix} A_N^\ominus & A_N^\otimes \\ -A_N^{\otimes*} & A_N^{\ominus*} \end{pmatrix} \quad (6.114)$$

13. The phase shift of both filter curves by  $\Delta\varphi = \pi/2$  is of minor importance.

with the coefficients

$$\begin{aligned} A_N^\ominus &= \cos(N\theta) + i \operatorname{Im}(A^\ominus) \frac{\sin(N\theta)}{\sin \theta} \\ A_N^\otimes &= |A^\otimes|^2 \frac{\sin(N\theta)}{\sin \theta} \end{aligned}$$

describes the behavior of a device with  $N$  periods. The angle  $\theta$  is defined via

$$\cos \theta = \frac{1}{2} \operatorname{Tr} \tilde{\mathcal{U}} = \operatorname{Re}(A^\ominus). \quad (6.115)$$

Using the transfer matrix (6.114), we can easily show that the relative optical power at the bar port of a directional coupler with  $N$  periods is given by

$$I_N^\ominus = \cos^2(N\theta) + \operatorname{Im}(A^\ominus)^2 \frac{\sin^2(N\theta)}{\sin^2 \theta}, \quad (6.116)$$

and the relative power level at the cross port of the device is

$$I_N^\otimes = |A^\otimes|^2 \frac{\sin^2(N\theta)}{\sin^2 \theta}. \quad (6.117)$$

For a proof of the above assertions we proceed as follows. We have seen in Section 6.2 that the transfer matrix of a compound coupler has the same mathematical form as that describing a uniform coupler (see Equation (6.52)); i.e., the unimodular transfer matrix  $\tilde{\mathcal{U}}$  is the most general transfer matrix describing a lossless directional coupler. It even includes potential phase shifters in the fundamental period. We saw in Section 2.2.4 that the transfer matrix of an  $N$ -period device is given by

$$\tilde{\mathcal{U}}_N^\times = U_{N-1} \left( \frac{\operatorname{Tr} \tilde{\mathcal{U}}_\times}{2} \right) \tilde{\mathcal{U}}_\times - U_{N-2} \left( \frac{\operatorname{Tr} \tilde{\mathcal{U}}_\times}{2} \right) \mathcal{E}, \quad (6.118)$$

In our case,  $\cos \theta = \operatorname{Tr} \tilde{\mathcal{U}} / 2 = \operatorname{Re}(A^\ominus) \leq 1$ , the trace of the transfer matrix represents a real quantity; i.e., we can represent the Chebyshev polynomials of second kind in the form

$$U_N(\cos \theta) = \frac{\sin((N+1)\theta)}{\sin \theta}.$$

Using  $\sin((N-1)\theta) = \cos \theta \sin(N\theta) - \sin \theta \cos(N\theta)$  we can verify the expression for the coefficient  $A_N^\ominus$ , while the derivation of the expression for the coefficient  $A_N^\otimes$  is straightforward. The expressions for the relative optical power at the two output ports were derived in Section 6.2.

We will now discuss some general properties of periodic couplers. From Equation (6.117) we see that the periodic device exhibits several perfect bar states which can be divided into two groups, one describing the bar states of the single period itself, i.e.,

$$A^\otimes = 0, \quad (6.119)$$

and a second one describing the new bar states of the periodic structure. These bar states are obviously found at  $\sin^2(N\theta) = 0$ , i.e., at  $N\theta = m\pi$ . By using Equation (6.115) we replace this condition by

$$\operatorname{Re}(A^\ominus) = \cos\left(\frac{m\pi}{N}\right). \quad (6.120)$$

Clearly, the periodic coupler also has two types of cross states, one standing for the cross states of the single period, which will usually not be perfect, and another one describing the new cross states of the compound device. These cross states are perfect if and only if the conditions

$$\begin{aligned} \operatorname{Im}(A^\ominus) &= 0 \\ \operatorname{Re}(A^\ominus) &= \cos\left(\frac{(2m-1)\pi}{2N}\right) \end{aligned} \quad (6.121)$$

are satisfied. If the imaginary part of the coefficient  $A^\ominus$  vanishes over the whole  $\kappa$ - $\delta$  plane, i.e.,  $\operatorname{Im}(A^\ominus) \equiv 0$ , then the periodic coupler will exhibit many perfect cross states. If it varies over the  $\kappa$ - $\delta$  plane, then only a few or even none of the solutions of Equation (6.121) will represent perfect cross states.

### 6.7.2 $\Delta\beta$ -Coupler

The  $\Delta\beta$ -coupler is a symmetrical coupler which can be made asymmetrical by an electrical control, e.g., by electro-optical switching. One period of this device consists of two sections with alternating detuning  $\delta$ . The length of each section is given by half of the period  $\Lambda$ . Using the representation of the transfer matrix of a directional coupler (6.36), we see that the section with positive detuning is described by the coefficients

$$\begin{aligned} A_\delta^\ominus &= \cos\left(\delta_{\text{eff}} \frac{\Lambda}{2}\right) + i \frac{\delta}{\delta_{\text{eff}}} \sin\left(\delta_{\text{eff}} \frac{\Lambda}{2}\right) \\ A_\delta^\otimes &= i \frac{\kappa}{\delta_{\text{eff}}} \sin\left(\delta_{\text{eff}} \frac{\Lambda}{2}\right). \end{aligned} \quad (6.122)$$

The coefficients for the section with negative detuning are easily obtained by using the transformations  $A_{-\delta}^\ominus = A_\delta^\ominus$ \* and  $A_{-\delta}^\otimes = A_\delta^\otimes$ . We can now derive the transfer matrix for one period of the  $\Delta\beta$ -coupler. It is given by

$$\tilde{\mathcal{U}} = \begin{pmatrix} A_\delta^{\ominus*} & A_\delta^\otimes \\ -A_\delta^{\otimes*} & A_\delta^\ominus \end{pmatrix} \begin{pmatrix} A_\delta^\ominus & A_\delta^\otimes \\ -A_\delta^\otimes & A_\delta^{\ominus*} \end{pmatrix} = \begin{pmatrix} A^\ominus & A^\otimes \\ -A^\otimes & A^\ominus \end{pmatrix} \quad (6.123)$$

with

$$\begin{aligned} A^\ominus &= |A_\delta^\ominus|^2 - |A_\delta^\otimes|^2 = 1 - 2|A_\delta^\otimes|^2 \\ A^\otimes &= 2A_\delta^{\ominus*}A_\delta^\otimes \end{aligned}$$

We see that the coefficient  $A^\Theta$  is always real:  $\text{Im}(A^\Theta) \equiv 0$ . The expressions for the relative power levels at the output ports are then given by

$$\begin{aligned} I^\Theta &= \cos^2(N\theta) \\ I^\otimes &= \sin^2(N\theta), \end{aligned} \quad (6.124)$$

i.e., the  $\Delta\beta$ -coupler exhibits many perfect cross states. Using Equation (6.123) we can express the angle  $\theta$  by the trigonometrical equation

$$\sin^2 \frac{\theta}{2} = \sin^2 \theta_c \sin^2 \left( \frac{\rho}{2} \right)$$

with  $\tan \theta_c = \kappa/\delta$  and  $\rho = \delta_{\text{eff}}\Lambda$ . We should note that  $\theta_c$  and  $\rho$  are the polar coordinates spanning the  $\delta\Lambda$ - $\kappa\Lambda$  plane. Using these results in turn, we can replace the conditions (6.120) and (6.121) for the occurrence of bar and cross states by a single condition

$$\sin^2 \theta_c \sin^2 \left( \frac{\rho}{2} \right) = \sin^2 \left( \frac{m\pi}{4N} \right). \quad (6.125)$$

The even integers  $m$  in Equation (6.125) stand for bar states of the  $\Delta\beta$ -coupler, the odd integers designate the cross states. For the case of detuning  $\delta = 0$  the coupler represents a symmetrical coupler of length  $N\Lambda$ . Thus, alternating cross and bar states are equidistantly located along the  $\kappa$ -axis. We see furthermore that the solutions belonging to the integer  $m$  are restricted to the region defined by

$$\sin^2 \theta_c \geq \sin^2 \left( \frac{m\pi}{4N} \right), \quad (6.126)$$

i.e., by  $|\theta_c| \geq m\pi/(4N)$ . Two of the bar states for any period of the right-hand side of Equation (6.126) exhibit a special behavior. One class corresponding to the integers  $m = 2N(2l - 1)$  represents a set of isolated points in the  $\delta\Lambda$ - $\kappa\Lambda$  plane since the condition  $\sin^2 \theta_c = 1$ , i.e.,  $\kappa = 0$ , is satisfied only along the  $\kappa$ -axis. The other class characterized by the integers  $m = 4lN$  represents a set of circles,  $\rho = 2\pi l$ . The switching diagram of the  $\Delta\beta$ -coupler shown in Figure 6.15 confirms this scenario.

### 6.7.3 $\Delta\kappa$ -Cougplers

We have seen in Section 6.4 that the design of uniform asymmetrical couplers results in stringent requirements on the material and geometry of the waveguides forming the coupler. We will see in the following that  $\Delta\kappa$ -couplers, i.e., directional couplers with a periodically varying coupling coefficient  $\kappa$ , exhibit similar behavior. However, they allow the spectral position of the resonance to be shifted by choosing an appropriate period  $\Lambda$  and thus the requirements on the underlying waveguide structure to be reduced.

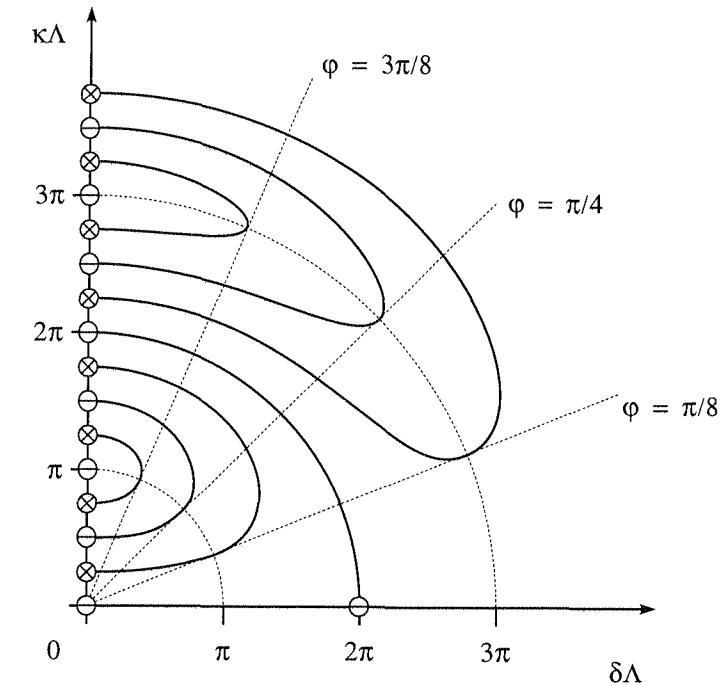


Figure 6.15. Switching diagram of the  $\Delta\beta$ -coupler.

In the “no-coupling” limit ( $\Delta\kappa \rightarrow 0$ ) the  $\Delta\kappa$ -coupler represents an asymmetrical coupler. The coefficients of the transfer matrix describing one period of the device are then given by

$$\begin{aligned} A_\delta^\Theta &= \cos(\delta_{\text{eff}}\Lambda) + i \frac{\delta}{\delta_{\text{eff}}} \sin(\delta_{\text{eff}}\Lambda) \\ A_\delta^\otimes &= i \frac{\kappa}{\delta_{\text{eff}}} \sin(\delta_{\text{eff}}\Lambda). \end{aligned}$$

Using the condition (6.121) we see that at the effective detuning  $\delta_{\text{eff}}$  a weakly coupled  $\Delta\kappa$ -coupler exhibits a resonance, a perfect cross state, if the period represents a multiple of the beat length of the uncorrugated device, i.e., if its period  $\Lambda$  satisfies the condition

$$\Lambda = \frac{m\pi}{\delta_{\text{eff}}}. \quad (6.127)$$

In the following discussion we will restrict ourselves to the analysis of the filter curves of the  $\Delta\kappa$ -coupler in the spectral vicinity of a resonance  $m$ . The most obvious way to tackle this problem is to assemble the transfer matrix for one period of the device by stacking the transfer matrices of its constituents and to expand the transfer matrix into a Taylor series around the resonance for further analysis.

Since this procedure necessitates extensive algebra which makes it difficult to keep track of the underlying physical effect, we will treat the  $\Delta\kappa$ -coupler by a different technique.

For this purpose we return to the (non-autonomous) coupled equations describing the  $\Delta\kappa$ -coupler and solve them within the framework of the rotating wave approximation. Since the coupling coefficient represents a real periodic function  $\kappa_{12}(z + \Lambda) = \kappa(z)$ , we can expand it into a Fourier series

$$\kappa(z) = \sum_{m=-\infty}^{\infty} \chi^{(m)} \exp\left(i \frac{2m\pi}{\Lambda} z\right). \quad (6.128)$$

We now introduce the transformation

$$\begin{aligned} \alpha_1(z) &= \tilde{\alpha}_1^{(m)}(z) \exp\left(i \frac{m\pi}{\Lambda} z\right) \\ \alpha_2(z) &= \tilde{\alpha}_2^{(m)}(z) \exp\left(-i \frac{m\pi}{\Lambda} z\right) \end{aligned} \quad (6.129)$$

and will formulate coupled mode equations in terms of the new amplitudes  $\tilde{\alpha}_m$ . With the shifted detuning

$$\delta^{(m)} = \delta - \frac{m\pi}{\Lambda}$$

the modified coupled mode equations can be written as

$$-i \frac{d}{dz} \begin{pmatrix} \tilde{\alpha}_1^{(m)} \\ \tilde{\alpha}_2^{(m)} \end{pmatrix} = \begin{pmatrix} \delta^{(m)} & \kappa(z) \exp\left(-i \frac{2m\pi}{\Lambda} z\right) \\ \kappa(z) \exp\left(i \frac{2m\pi}{\Lambda} z\right) & \delta^{(m)} \end{pmatrix} \begin{pmatrix} \tilde{\alpha}_1^{(m)} \\ \tilde{\alpha}_2^{(m)} \end{pmatrix} \quad (6.130)$$

Using our basic assumption (6.127) we see that for weakly coupled devices the transformation (6.129) filters out the most significant oscillations of the original amplitude  $\alpha_m$ . We can then neglect the oscillating terms on the right-hand side of the coupled mode equations (6.130) and obtain the autonomous differential equations

$$-i \frac{d}{dz} \begin{pmatrix} \tilde{\alpha}_1^{(m)} \\ \tilde{\alpha}_2^{(m)} \end{pmatrix} \approx \begin{pmatrix} \delta^{(m)} & \chi^{(m)} \\ \chi^{(m)*} & \delta^{(m)} \end{pmatrix} \begin{pmatrix} \tilde{\alpha}_1^{(m)} \\ \tilde{\alpha}_2^{(m)} \end{pmatrix} \quad (6.131)$$

for the  $\Delta\kappa$ -coupler. The procedure is called *rotating-wave approximation* (RWA). Equation (6.131) represents the system of coupled mode equations of a uniform asymmetrical coupler. However, the detuning  $\delta$  is replaced by the shifted detuning  $\delta^{(m)}$  and the coupling coefficient  $\kappa$  by its  $m$ th Fourier coefficient  $\chi^{(m)}$ . Using the fundamental solution of the coupled mode equations for the uniform directional coupler derived in Section 6.2, we obtain the following for the transfer matrix describing one period of the  $\Delta\kappa$ -coupler

$$\tilde{U} = \begin{pmatrix} A^\Theta & A^\otimes \\ -A^{\otimes*} & A^{\Theta*} \end{pmatrix} \quad (6.132)$$

with the coefficients

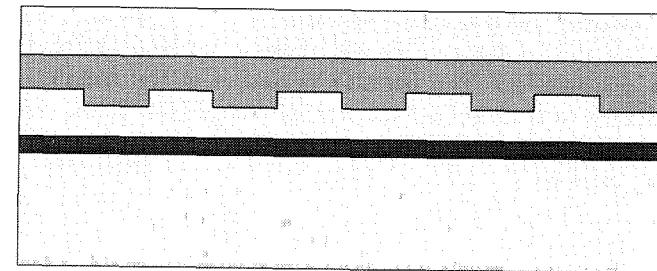
$$\begin{aligned} A_\delta^\Theta &= \cos(\delta_{\text{eff}}^{(m)} \Lambda) + i \frac{\delta^{(m)}}{\delta_{\text{eff}}^{(m)}} \sin(\delta_{\text{eff}}^{(m)} \Lambda) \\ A_\delta^\otimes &= i \frac{\chi^{(m)}}{\delta_{\text{eff}}^{(m)}} \sin(\delta_{\text{eff}}^{(m)} \Lambda). \end{aligned}$$

The parameter

$$\delta_{\text{eff}}^{(m)} = \sqrt{\left(\delta^{(m)} - \frac{m\pi}{\Lambda}\right)^2 + |\chi_m|^2} \quad (6.133)$$

designates the effective detuning of the  $\Delta\kappa$ -coupler.

### a) Vertical $\Delta\kappa$ -coupler



### b) Meander-coupler

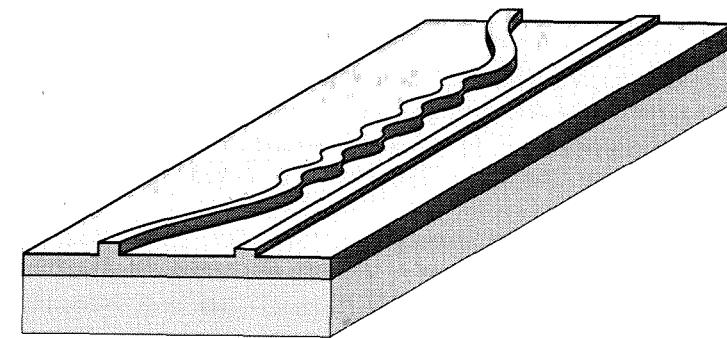


Figure 6.16. Two examples for the  $\Delta\kappa$ -coupler: (a) vertical structure, (b) meander coupler (by courtesy of Dr. H. P. Nolting, Heinrich-Hertz Institut).

We can conclude that the  $\Delta\kappa$ -coupler behaves like an asymmetrical coupler. It exhibits a detuning  $\delta^{(m)}$  which is shifted by a "grating constant"  $m\pi/\delta_{\text{eff}}^{(m)}$ . Its coupling coefficient is given by the corresponding Fourier coefficient  $\chi^{(m)}$  of the coupling coefficient  $\kappa$ . Thus, most of the results obtained within the framework of Section 6.4 can be applied directly to the  $\Delta\kappa$ -coupler. In particular, the maxima of the filter curve are located on an envelope of Lorentzian shape whose full-width half-maximum (FWHM) is given by

$$\chi_c = 2 \left| \frac{\chi^{(m)}}{\nabla_k \delta} \right|. \quad (6.134)$$

The parameter  $\nabla_k \delta$  designates, as for the asymmetrical coupler, the angle of intersection of the shifted dispersion curves at the resonance (see Equation (6.70)). A full crossover can be achieved only if the coupling length

$$L^\otimes = N\Lambda = \frac{\pi}{2\chi^{(m)}} \quad (6.135)$$

represents an integer multiple of the period  $\Lambda$ .

At the end of this paragraph we will make a few remarks on the practical realization of a  $\Delta\kappa$ -coupler. The periodic variation of the coupling coefficient can be realized as a vertical or horizontal grating (see Figure 6.16a) or as a horizontal meander structure (see Figure 6.16b). It is common to operate these devices in the first order, i.e.,  $m = 1$ .

## REFERENCES

- [1] Hardy, A., and Streifer, W., "Coupled Mode Theory of Parallel Waveguides", *IEEE Journal of Lightwave Technology*, Vol. 3, 1985, pp. 1135–1146.
- [2] Haus, H. A., Huang, W. P., Kawakami, S., and Whitaker, N. A., "Coupled-Mode Theory of Optical Waveguides", *IEEE Journal of Lightwave Technology*, Vol. 5, 1987, pp. 16–23.
- [3] Weinert, C. M., "Three Dimensional Coupled Mode Method for Simulation of Coupler and Filter Structures", *IEEE Journal of Lightwave Technology*, Vol. 10, 1992, pp. 1218–1225.
- [4] Alferness, R. C., and Cross, P. C., "Filter Characteristics of Codirectionally Coupled Waveguides with Weighted Coupling", *IEEE Journal of Quantum Electronics*, Vol. 14, 1978, pp. 843–847.
- [5] Korotky, S. K., "Three-Space Representation of Phase-Mismatch Switching in Coupled Two-State Optical Systems", *IEEE Journal of Quantum Electronics*, Vol. 22, 1986, pp. 952–958.
- [6] Louisell, W., *Coupled Mode and Parametric Electronics*, New York, Wiley, 1960.
- [7] Kogelnik, H., and Schmidt, R. V., "Switched Directional Couplers with Alternating  $\Delta\beta$ ", *IEEE Journal of Quantum Electronics*, Vol. 12, 1976, pp. 396–401.
- [8] Marcatili, E., "Improved Coupled Mode Equations for Dielectric Guides", *IEEE Journal of Quantum Electronics*, Vol. 6, 1986, pp. 988–993.
- [9] Broberg, B., Lindgren, B. S., Öberg, M. G., and Jiang, H., "A Novel Integrated Optics Wavelength Filter in InGaAsP-InP", *IEEE Journal of Lightwave Technology*, Vol. 4, 1986, pp. 196–203.
- [10] Walker, R. G., Urquhart, J., Bennion, I., and Carter, A. C., "1.3/1.5  $\mu\text{m}$  Mach-Zehnder Wavelength Duplexers for Integrated Optoelectronic Transceiver Modules", *IEE Proceedings*, Vol. 137, Pt. J, 1990, pp. 33–38.
- [11] Bornholdt, C., Kappe, F., Müller, R., Nolting, H.-P., Reier, F., Stenzel, R., Venghaus, H., and Weinert, C. M., "Meander Coupler, a Novel Wavelength Division Multiplexer/Demultiplexer", *Applied Physics Letters*, Vol. 57, 1990, pp. 2517–2519.
- [12] Lessard, S., and Huang, W. P., "Assessment of Coupled-Mode Theory for Tapered Optical Coupler", *IEEE Journal of Lightwave Technology*, Vol. 11, 1993, pp. 405–407.

## Chapter 7

### Contradirectional Coupling

Contradirectional coupling represents the integrated optical analog to the Bragg reflections occurring in any type of crystal. Whereas the distributed feedback in crystals is caused by the crystallographic grating itself, it must be initiated in integrated optics by impressing an appropriate periodic (or almost periodic) overlay onto the underlying waveguide structure. The spectral response exhibits the same scenario for both examples: strictly resonant behavior and stopbands, i.e., extended spectral regions of perfect reflectance, around the resonances. Today, contradirectional coupling is used for integrated optical filters with small channel spacings as well as to integrated optical resonators for distributed feedback (DFB) and distributed Bragg reflector (DBR) lasers. The smallness of the overlay period – typically  $0.1\text{--}0.5\ \mu\text{m}$  depending on the underlying wavelength region and material system – necessitates the use of enhanced fabrication technologies such as holographic exposure or electron beam writing for the definition of the periodic overlay. This serious technological drawback has hitherto prevented more widespread use of contradirectionally coupled devices.

In this chapter we will first derive the phenomenological theory of contradirectional coupling starting from the corresponding conservation laws. Within the framework of this discussion we will show that only the first-order Bragg reflections are unaffected by the radiation of optical power into sub- and superstrate. Next, we will examine Bragg gratings and derive their transfer matrix, which will turn out to be a useful tool for the analysis of compound devices such as quarter-wave shifted Bragg gratings. The analysis of Bragg gratings operated at oblique incidence will yield a rich variety of physical effects such as the TE-TM polarization conversion, the occurrence of a Brewster angle for TE polarization and the Goos-Hänchen shift of finite beams. The chapter ends with a discussion of contradirectional couplers, i.e., directional couplers with an appropriate periodic overlay, within the framework of a 4-wave coupled mode theory.

## 7.1 PHENOMENOLOGICAL COUPLED MODE THEORY

We will now examine the coupled mode theory for a contradirectionally coupled waveguide array which can in principle support many guided modes (see Figure 7.1). Starting from the conservation law governing contradirectional coupling we will derive symmetry relations for its coupling coefficients. We will then show that only first-order devices are unaffected by radiation into the sub- and superstrate. What follows is a brief elucidation of the effects caused by tapering and chirping of the periodic structure on the filter curves of the device. The end of this section comprises a mathematical assessment of contradirectional coupling and a comparison with related problems.

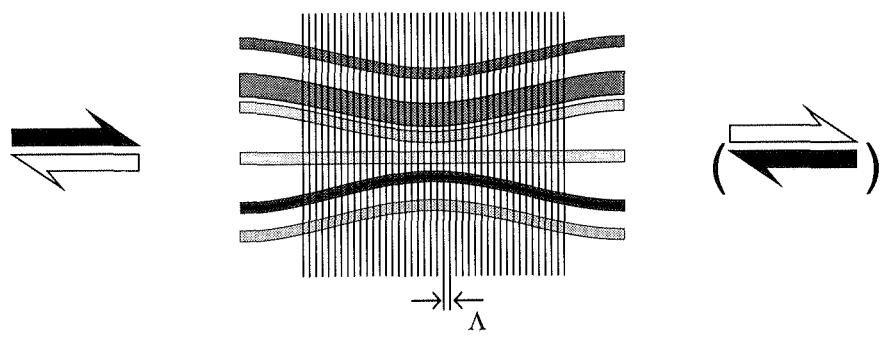


Figure 7.1. Contradirectionally coupled waveguide array.

### 7.1.1 Coupled Mode Equations

At the beginning of this chapter we will discuss the effects of contradirectional coupling from a purely phenomenological point of view. For this purpose we consider a contradirectionally coupled waveguide array with a period  $\Lambda$  which couples two sets of eigenmodes, one describing the forward propagating waves and a second one for the backward propagating waves. Thus, we expand the optical field

$$|\Phi(z)\rangle \propto \sum_{n=1}^N \{a_n^{(+)}(z) + \sigma a_n^{(-)}(z)\} |n\rangle \quad (7.1)$$

into a set of  $N$  modes  $|n\rangle$ . The signum

$$\sigma = \begin{cases} +1 & \text{for E-field formulations} \\ -1 & \text{for H-field formulations} \end{cases}$$

accounts for the symmetry of the electric and magnetic fields with respect to reverse propagation. These modes can be regarded either as a number of supermodes of the unperturbed coupled waveguide array or as a collection of modes of its individual waveguides. Equation (7.1) assumes that the coupling to any other eigenmodes can be neglected.<sup>1</sup>

The phenomenological coupled mode equations for the rapidly varying expansion coefficients of a waveguide structure exhibiting both co- and contradirectional coupling can be written as

$$\begin{aligned} -i \frac{da_m^{(+)}}{dz} &= \sum_{n=1}^N [\kappa_{mn}(z) a_n^{(+)} + \kappa_{mn}^{(+)}(z) a_n^{(-)}] \\ -i \frac{da_m^{(-)}}{dz} &= -\sum_{n=1}^N [\kappa_{mn}(z) a_n^{(-)} + \kappa_{mn}^{(-)}(z) a_n^{(+)}]. \end{aligned} \quad (7.2)$$

In Equations (7.2) the coupling coefficients  $\kappa_{mn}$  account for codirectional coupling, i.e., for the coupling between two modes traveling in the same direction. The other coupling coefficients  $\kappa_{mn}^{(\pm)}$  account for the contradirectional coupling, i.e., for the coupling between two modes traveling in opposite directions.

From a mathematical point of view the coupled mode equations (7.2) represent a non-autonomous system of ordinary differential equations with rapidly varying coupling coefficients

$$\begin{aligned} \kappa_{mn}(z) &= \kappa_{mn} + \sum_{l=-\infty}^{\infty} \chi_{mn,l}^{(\rightarrow)} e^{iQ_l z} \\ \kappa_{mn}^{(\pm)}(z) &= \sum_{l=-\infty}^{\infty} \chi_{mn,l}^{(\pm)} e^{iQ_l z} \end{aligned} \quad (7.3)$$

where

$$Q_l = \frac{2\pi l}{\Lambda}$$

designates the grating vector of a periodic structure of a period  $\Lambda$  which is operated in the  $l$ th order. For lossless waveguide arrays the coupling coefficients  $\kappa_{mn}(z)$  can be made real. This results in the following symmetry relations for the expansion coefficients:

$$\begin{aligned} \chi_{mn,l}^{(\rightarrow)} &= \chi_{mn,-l}^{(\rightarrow)*} \\ \chi_{mn,l}^{(\pm)} &= \chi_{mn,-l}^{(\pm)*}. \end{aligned}$$

Since waveguide structures without periodic overlay do not exhibit contradirectional coupling, we can further conclude that

$$\chi_{mn,0}^{(\rightarrow)} = \chi_{mn,0}^{(\pm)} = 0.$$

1. See Section 6.1 for a more detailed discussion of the basic assumptions of coupled mode theory.

To obtain an overview of the relevant contributions of the coupled mode equations (7.2) for a specific waveguide array, it is useful to reformulate the equations setting

$$\begin{aligned} a_m^{(+)} &= \tilde{a}_m^{(+)} e^{iQ_l z/2} \\ a_m^{(-)} &= \tilde{a}_m^{(-)} e^{-iQ_l z/2} \end{aligned} \quad (7.4)$$

We then obtain the modified coupled mode equations

$$\begin{aligned} -i \frac{d\tilde{a}_m^{(+)}}{dz} &= \sum_{n=1}^N \left[ (\kappa_{mn} - \delta_{mn} Q_l/2) \tilde{a}_n^{(+)} + \chi_{mn,l}^{(+)} \tilde{a}_n^{(-)} \right] + \text{osc.} \\ -i \frac{d\tilde{a}_m^{(-)}}{dz} &= - \sum_{n=1}^N \left[ (\kappa_{mn} - \delta_{mn} Q_l/2) \tilde{a}_n^{(-)} + \chi_{mn,l}^{(-)*} \tilde{a}_n^{(+)} \right] + \text{osc.} \end{aligned} \quad (7.5)$$

We saw within the framework of our discussion of the rotating wave approximation (RWA) that the oscillating terms can be neglected if the variables are (slowly varying) amplitudes, i.e., if the grating vector satisfies the condition

$$Q_l = q_m + q_n \quad (7.6)$$

where

$$q_m = \kappa_{mm}$$

stands for the propagation constant of the eigenmode  $m$  under the influence of self-coupling.<sup>2</sup> The coupled mode equations (7.5) then become a set of autonomous differential equations. We see that the diagonal elements

$$\delta_m = \kappa_{mm} - Q_l/2 = q_m - Q_l/2$$

represent the detuning of the mode, i.e., the deviation of its propagation constant from half the grating constant.

If we calculate the Poynting vector by using the representation (7.1) of the optical field and keep in mind that the expansion coefficients  $a_n^{(\pm)}$  of the optical field and thus any type of amplitudes (e.g.,  $\tilde{a}_n^{(\pm)}$ ) have the same absolute value, we see that the optical power is conserved along the contradirectionally coupled waveguide array if and only if the condition

$$\frac{d}{dz} \sum_{m=1}^N \left[ |\tilde{a}_m^{(+)}|^2 - |\tilde{a}_m^{(-)}|^2 \right] = 0 \quad (7.7)$$

is satisfied. By using the coupled mode equations (7.5) we can easily derive a set of symmetry relations for the coupling coefficients of lossless waveguide structures.

2. For a direct derivation of this formula, the coupled mode equations (7.2) are formulated in terms of the true amplitudes  $a_m^{(\pm)} = \tilde{a}_m^{(\pm)} e^{\pm iq_m z}$ .

Inserting the coupled mode equations into the representation (7.7) we obtain

$$\begin{aligned} \frac{\partial}{\partial z} \sum_{m=1}^N &\left[ |\tilde{a}_m^{(+)}|^2 - |\tilde{a}_m^{(-)}|^2 \right] \\ &= i \sum_{m=1}^N (\kappa_{mn} - \kappa_{nm}^*) \left( a_m^{(+)*} \tilde{a}_n^{(+)} + a_m^{(-)*} \tilde{a}_n^{(-)} \right) \\ &\quad + \left( \chi_{mn,l}^{(+)} - \chi_{nm,l}^{(-)*} \right) a_m^{(+)*} \tilde{a}_m^{(-)} - \left( \chi_{mn,l}^{(-)} - \chi_{nm,l}^{(+)*} \right) a_m^{(-)*} \tilde{a}_m^{(+)} \\ &\stackrel{!}{=} 0, \end{aligned}$$

and we can conclude that the coupling coefficients of a contradirectionally coupled lossless waveguide array must satisfy the following conditions

$$\kappa_{mn} = \kappa_{nm}^* \quad (7.8)$$

and

$$\chi_{mn,l} = \chi_{mn,l}^{(+)} = \chi_{nm,l}^{(-)*} \quad (7.9)$$

For symmetry reasons we can further assume that the conditions

$$\chi_{mn,l} = \chi_{nm,l}^* \quad (7.10)$$

are satisfied.

The matrix representation of the coupled mode equations describing a lossless waveguide structure is then given by

$$-i \frac{\partial}{\partial z} \begin{pmatrix} \tilde{\mathbf{a}}^{(+)} \\ \tilde{\mathbf{a}}^{(-)} \end{pmatrix} = \begin{pmatrix} \mathcal{K} & \mathcal{X}_l \\ -\mathcal{X}_l & -\mathcal{K} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{a}}^{(+)} \\ \tilde{\mathbf{a}}^{(-)} \end{pmatrix} \quad (7.11)$$

in which the vectors  $\tilde{\mathbf{a}}^{(+)}$  and  $\tilde{\mathbf{a}}^{(-)}$  represent the parameter vectors of the expansion coefficients  $\tilde{a}_n^{(+)}$  and  $\tilde{a}_n^{(-)}$ , respectively. The matrices  $\mathcal{K}$  and  $\mathcal{X}_l$  describe the codirectional and contradirectional coupling. Equation (7.11) shows that the contradirectionally coupled waveguide array exhibits contradirectional coupling via the Fourier coefficient matrix  $\mathcal{X}_l$  which is accompanied by codirectional coupling via the matrix  $\mathcal{K}$ . The corresponding matrix elements are given by

$$\begin{aligned} (\mathcal{K})_{mn} &= \kappa_{mn} \\ (\mathcal{X}_l)_{mn} &= \chi_{nm,l}. \end{aligned}$$

Considering the “no-contradirectional-coupling” limit ( $\mathcal{X}_l \rightarrow \mathcal{O}$ ) we see easily that the eigenmodes of the coupled mode equations represent the propagation constants of the coupled waveguide array which are related to the grating vector. We see furthermore that any formulation based on the supermodes of the full waveguide structure results in a diagonal matrix  $\mathcal{X}_l$  for codirectional coupling.

### 7.1.2 Radiation

In the following we will show that contradirectionally coupled waveguide arrays radiate under certain conditions. For the sake of simplicity we will start this discussion with the most elementary device, the planar Bragg grating operated at normal incidence. We can then restrict our discussion to the  $q_x$ - $q_z$  plane. By using the rotating wave approximation, we saw in Section 7.1 that a Bragg grating of a period  $\Lambda$  exhibits resonances if the condition

$$q_{\text{in}} \pm q_{\text{out}} = Q_l \quad (7.12)$$

with the grating vector  $Q_l = 2\pi l/\Lambda$  and the effective propagation constants  $q_{\text{in}}$  and  $q_{\text{out}}$  is satisfied. Equation (7.12) can be regarded as a “conservation law” for the  $z$ -component of the wave numbers. We further know that both incident and reflected waves must be proper eigenmodes of the underlying waveguide structure. In particular, the radiation modes must satisfy the additional condition

$$q_x^2 + q_z^2 = k_0^2 \epsilon_B \quad (7.13)$$

where  $\epsilon_B$  designates the background dielectric function of the substrate or superstrate.

Equations (7.12) and (7.13) are sufficient to carry out a kinetic analysis of radiation, i.e., an analysis of potential radiation by calculating the corresponding angles. The analysis of the coupling efficiencies, however, requires the full solution of the coupled mode equations. Let us start the discussion with a Bragg grating which is operated in the first order. Figure 7.2a shows the schematic diagram of the  $q_x$ - $q_z$  plane which applies in that case. The circles in this diagram indicate the proper radiation modes of the substrate ( $q_S$ ) and superstrate ( $q_A$ ). We see directly that a Bragg grating operated in first order allows only for back reflection. We see furthermore that radiation will start to occur on the short-wave side of the stopband if the condition

$$q_{\text{out}} = Q_1 - q_{\text{in}} \leq q_S \quad (7.14)$$

is satisfied. Consequently, a Bragg grating operated in second order (see Figure 7.2b) will exhibit both Bragg reflection back into the waveguide and radiation which is emitted normal to the chip surface into both the substrate and superstrate. A thirdorder Bragg grating (see Figure 7.2c) will already show a typical multi-beam environment. Beside the Bragg reflection we will observe two beams each radiated into the substrate and superstrate. Obviously, with an increasing order of the grating the number of radiated beams will successively increase.<sup>3</sup>

The extension of these results to planar Bragg gratings operated at oblique incidence is straightforward. The resonance condition (7.12) is replaced by

$$q_{\text{in}} - q_{\text{out}} = Q_l \quad (7.15)$$

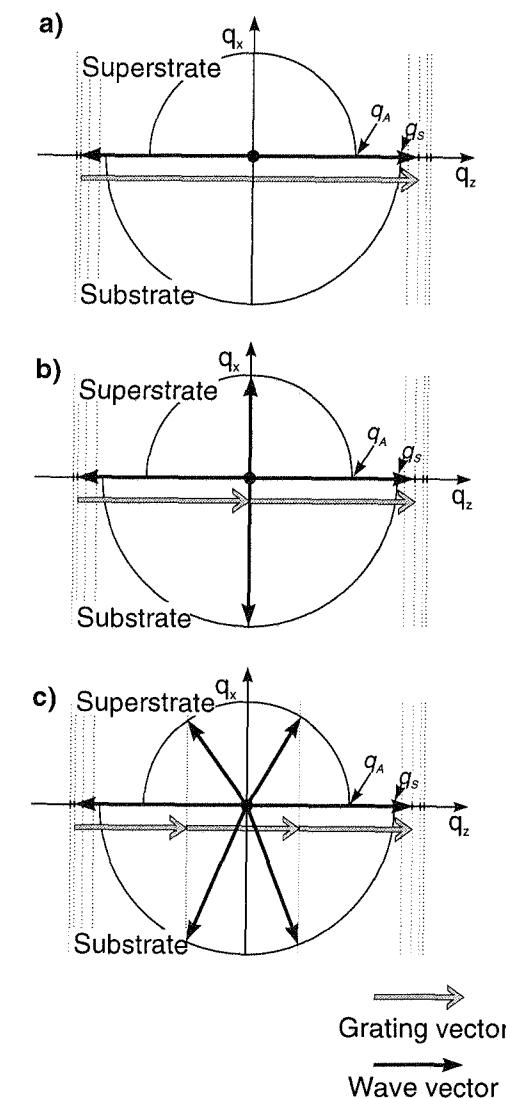


Figure 7.2. Vector diagrams for Bragg gratings operated in the first (a), second (b) and third (c) order

3. Within the framework of this discussion, the  $\Delta\kappa$ -coupler treated in Section 6.7 can be regarded as a contradirectional coupler which is operated in extremely high order. However, the small number of periods results in completely inefficient operation of the device.

where  $\mathbf{q}_{\text{in}}$ ,  $\mathbf{q}_{\text{out}}$  and  $\mathbf{Q}_l$  designate vectors in the  $q_y$ - $q_z$  plane. The radiation modes must now satisfy the additional condition

$$q_x^2 + q_y^2 + q_z^2 = k_0^2 \epsilon_B, \quad (7.16)$$

i.e., the radiation modes are found on a hemisphere in momentum space. The vector diagrams in the  $q_y$ - $q_z$  plane (see Figure 7.3) are commonly used to analyze the radiation conditions in that case. Non-planar waveguide structures will exhibit radiation of finite extent, i.e., beams which undergo diffraction during propagation.

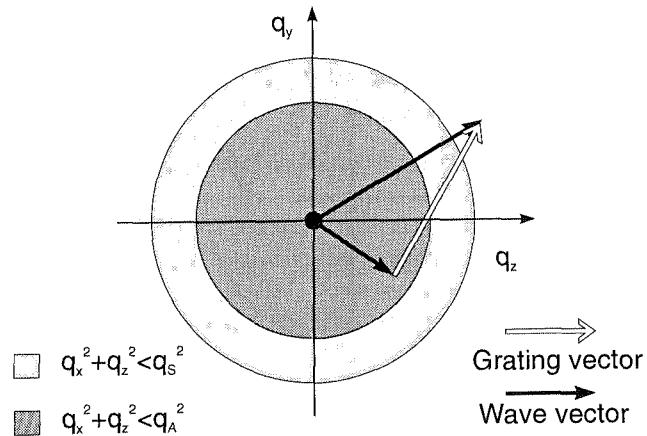


Figure 7.3. Vector diagrams for the waveguide plane.

From the analysis just carried out we can draw two conclusions which are of some importance for the operation of contradirectionally coupled waveguide arrays.

- Contradirectionally coupled waveguide arrays which are designed as Bragg reflectors should be operated in the first order. Otherwise, radiation losses will occur.
- Coupled waveguide arrays (usually Bragg gratings) can be operated as grating couplers, i.e., as devices which launch light into or out of waveguides. By appropriate chirping of the gratings, such devices can be made focusing, and their efficiency can be influenced by the use of blazed gratings, i.e., choosing an appropriate shape of grating relief.

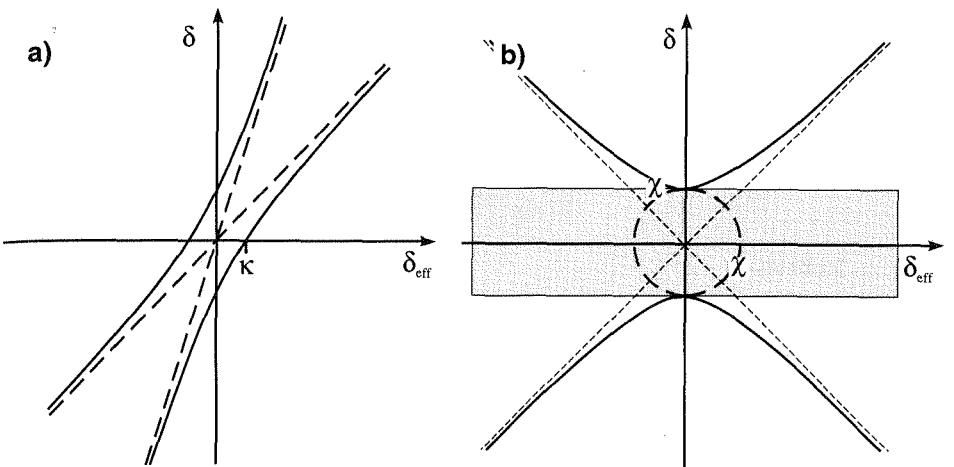


Figure 7.4. Dispersion curves of two codirectionally (a) and contradirectionally (b) coupled modes.

### 7.1.3 Dispersion and Filter Curves

We will now perform a qualitative analysis of the nature of contradirectional coupling in the spectral vicinity of a crossover of two dispersion curves of a contradirectionally coupled waveguide array. We will derive the corresponding equations in the following two sections of this chapter. Figure 7.4 shows the dispersion curves, i.e., the detuning  $\delta$  as a function of the effective detuning  $\delta_{\text{eff}}$ , of two codirectionally and contradirectionally coupled modes in the spectral region of its crossover. We see that both dispersion curves split up at the crossover. However, the codirectional coupler exhibits real propagation constants in the whole spectral region (see the following paragraph for the mathematical reasons) whereas the contradirectionally coupled modes exhibit a stopband, i.e., a finite spectral region where the propagation constant becomes purely imaginary even if the underlying waveguide structure is lossless. Obviously, the propagator of such a mode is given by

$$\mathcal{U} \propto e^{-\text{Im}(q)z}.$$

We see directly that the optical power launched into the device is successively coupled to the backward traveling modes and that a device of infinite length will not transmit any optical power. As a rule of thumb, the width of the stopband is given by twice the absolute value of the corresponding coupling coefficient  $\chi_{mn}$ . If adjacent stopbands affect each other, however, they are shifted and constricted (see Section 7.3). The typical shape of the imaginary part of the propagation constant inside the stopband is indicated in Figure 7.4b by a dashed line. Usually, this curve has a maximum at the center of the stopband, i.e., the coupling of the optical field to the backward traveling modes is most efficient at this spectral position.

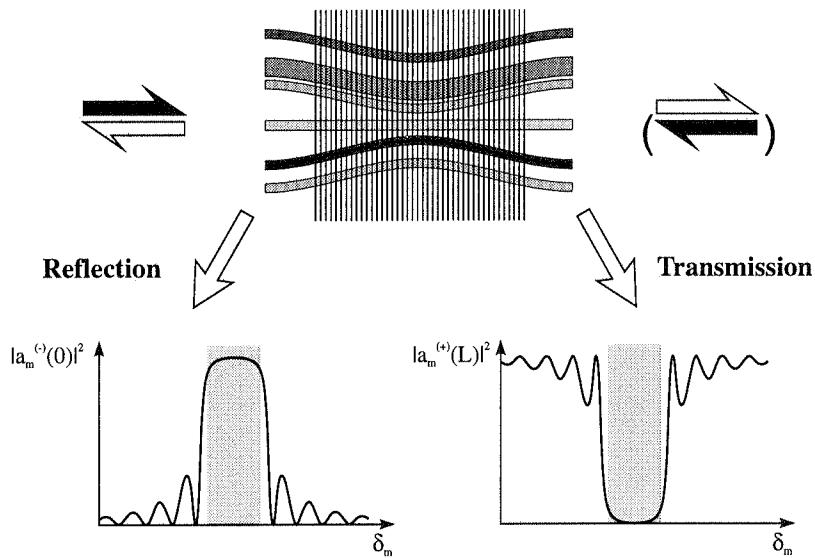


Figure 7.5. Transmission and reflection in a contradirectionally coupled waveguide array.

For calculating the response of the contradirectionally coupled waveguide array of a length  $L$ , we will now assume that the coupled mode equations are already solved, i.e.,

$$\begin{pmatrix} \mathbf{a}^{(+)}(L) \\ \mathbf{a}^{(-)}(L) \end{pmatrix} = \begin{pmatrix} \mathcal{U}_{++} & \mathcal{U}_{+-} \\ \mathcal{U}_{-+} & \mathcal{U}_{--} \end{pmatrix} \begin{pmatrix} \mathbf{a}^{(+)}(0) \\ \mathbf{a}^{(-)}(0) \end{pmatrix} \quad (7.17)$$

where  $\mathbf{a}^{(+)}$  and  $\mathbf{a}^{(-)}$  designate the rapidly varying field components and  $\mathcal{U}_{++}$ ,  $\mathcal{U}_{+-}$ ,  $\mathcal{U}_{-+}$ ,  $\mathcal{U}_{--}$  the constituent propagators. Obviously, they must satisfy the symmetry relations

$$\begin{aligned} \mathcal{U}_{--}(L) &= \mathcal{U}_{++}(-L) \\ \mathcal{U}_{-+}(L) &= \mathcal{U}_{+-}(-L). \end{aligned}$$

The diagram on the top of Figure 7.5 shows that  $\mathbf{a}^{(+)}(0)$  and  $\mathbf{a}^{(-)}(L)$  form the stimulus of the device. The response of the device is found at the vectors  $\mathbf{a}^{(+)}(L)$  and  $\mathbf{a}^{(-)}(0)$ . By resolving Equation (7.17) with respect to these two quantities we obtain

$$\begin{pmatrix} \mathbf{a}^{(+)}(L) \\ \mathbf{a}^{(-)}(0) \end{pmatrix} = \begin{pmatrix} \mathcal{U}_{++} + \mathcal{U}_{+-}\mathcal{U}_{--}^{-1}\mathcal{U}_{-+} & -\mathcal{U}_{+-}\mathcal{U}_{--}^{-1} \\ -\mathcal{U}_{--}^{-1}\mathcal{U}_{-+} & \mathcal{U}_{--}^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{a}^{(+)}(0) \\ \mathbf{a}^{(-)}(L) \end{pmatrix} \quad (7.18)$$

Usually, the stimulus is applied to the device only via its left-hand side ( $\mathbf{a}^{(-)}(L) = 0$ ). The vector  $\mathbf{a}^{(+)}(z)$  describes then the transmission behavior of the device, and  $\mathbf{a}^{(-)}(0)$

the reflection behavior. Figure 7.5 shows the transmission and reflection curves for the spectral vicinity of an isolated crossover. For the sake of clarity the stopband is indicated in the filter curve.

#### 7.1.4 Chirped and Tapered Coupling

At the end of this section we will discuss the filter response of contradirectionally coupled waveguide arrays which exhibit a quasi-adiabatic variation of the grating strength, i.e., a gradual change of the absolute value of the coupling coefficient  $\chi_{mn,l}$ , and of the grating constant  $\Lambda$  itself. The first effect is designated as *tapering*, and the second as *chirping*.

Let us start with a discussion of tapering and recall the results of Section 6.4.3 where we treated the tapered coupling of asymmetrical directional couplers. Within the framework of this discussion we have reformulated the coupled mode equations as a Riccati equation and have shown that “far away” from the resonance the upper envelope of the filter curve is given by the Fourier transform of the coupling coefficient. Since the coupled mode equations of the corresponding contradirectionally coupled device differ only by one sign, we can obviously apply the same procedure, and we obtain the same physical result. *Far away from the resonance, the relative optical power  $R$  reflected by a contradirectionally coupled device is given*

$$R = \left| \mathcal{F}[\chi_{mn}^{(\sigma)}]_{2q-Q_l} \right|^2 \quad (7.19)$$

*by the Fourier transform of the tapered coupling coefficient.* We should note, however, that from a fabrication point of view it is much more difficult to vary the grating strength of contradirectionally coupled devices (mostly Bragg gratings) than to taper codirectional couplers.

The second technique for influencing the filter curve is chirping, i.e., the gradual variation of the grating period  $\Lambda$ . Figure 7.6 is an illustrative example showing the most elementary chirped grating consisting of two grating sections  $A$  and  $B$ . The filter curves of the constituent gratings are mounted on top of the grating sections. In the spectral region of the stopband of the first grating, the optical power is reflected to a high degree by the first grating. Thus, we expect that the compound grating will exhibit the stopband of the first grating. For the same reasons, the chirped grating will exhibit the stopband of the second grating. Using our simple example we see that *it is possible to broaden the stopbands of contradirectionally coupled structures by chirping*. We should note, however, that broadening of stopbands can be achieved only at the expense of longer devices.

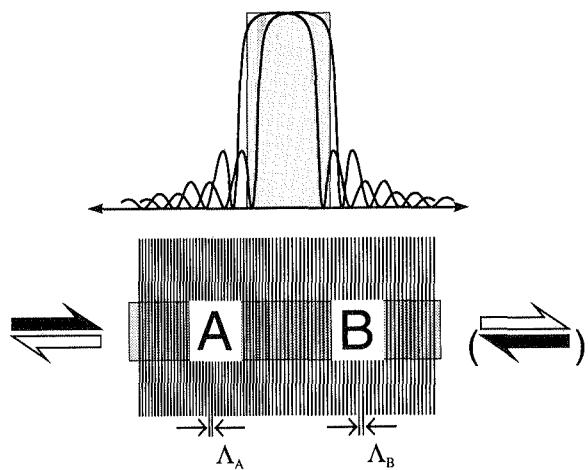


Figure 7.6. Bragg grating consisting of two grating sections.

### 7.1.5 Mathematical Assessment

We will now compare the coupled mode theories of codirectional coupling derived in the previous chapter and the contradirectional coupling discussed here from a more mathematical point of view. For this purpose we will restrict ourselves to lossless waveguide arrays.

For both codirectional and contradirectional coupling the evolution of expansion coefficients is described by a set of autonomous ordinary differential equations

$$-i \frac{d\tilde{\mathbf{a}}}{dz} = \mathcal{K}\tilde{\mathbf{a}}. \quad (7.20)$$

However, the conservation laws which hold for lossless waveguide arrays differ. The conservation law for codirectional coupling is namely given by

$$\frac{d}{dz} \sum_{n=1}^{\infty} |\tilde{a}_n|^2 = 0. \quad (7.21)$$

The corresponding theorem for contradirectional coupling is

$$\frac{d}{dz} \sum_{n=1}^{\infty} \left\{ |\tilde{a}_n^{(+)}|^2 - |\tilde{a}_n^{(-)}|^2 \right\} = 0 \quad (7.22)$$

in which the expansion coefficients for forward and backward traveling waves are designated as  $\tilde{a}_n^{(+)}$  and  $\tilde{a}_n^{(-)}$ , respectively. The different conservation laws will turn out to be crucial for the completely different behavior of devices which are based on co- and contradirectional coupling.

Let us start the discussion of the mathematical differences by briefly recalling the most important properties of codirectional couplers. We have seen in Section 6.1 that – as a consequence of the conservation law (7.21) – the coupling matrix  $\mathcal{K}$  is always Hermitian. We could conclude from this that the effective propagation constants of the coupled waveguide are always real, i.e., that the evolution of eigenmodes is described by pure phase factors. From that we could further conclude that the propagator  $\mathcal{U}_{\rightarrow}$  is a unitary operator. We have further seen that the propagator can be made unimodular by applying an appropriate phase factor such that the propagators become members of the unimodular unitary group  $SU(N)$ .

We have seen in the starting section of this chapter that as a consequence of the conservation law (7.22) the coupling matrix describing a contradirectionally coupled waveguide array is not Hermitian, and we can directly conclude that the corresponding propagator  $\mathcal{U}_{\leftrightarrow}$  will be not unitary. Nevertheless, the propagators  $\mathcal{U}_{\leftrightarrow}$  with the conservation law (7.22) will form a group. This group is designated as the group  $U(N, N)$  of  $(N, N)$ -unitary matrices where the doublet  $(N, M)$  stands for  $N$  positive and  $M$  negative squares in the conservation law (7.22). If the propagator  $\mathcal{U}_{\leftrightarrow}$  is made unimodular by introducing an appropriate phase factor, the propagators form the unimodular  $(N, N)$ -unitary group  $SU(N, N)$ .

From a more physical point of view  $U(N, N)$ -propagators are typical for reflecting devices or setups. Examples are the scattering at a potential barrier in quantum mechanics and the Bragg reflections in crystals and holograms. The most prominent example is the special theory of relativity where the group velocity of light in free space represents a  $(3, 1)$ -unitary form which is conserved.

As a consequence of their widespread applications, the properties of the unimodular  $(N, N)$ -unitary group  $U(N, N)$  – for example, its interdependence with hyperbolic geometry and its analysis of stability against parameter variations – are extensively discussed in the mathematical literature [1].

## 7.2 BRAGG GRATINGS

Bragg gratings, i.e., single-mode slab or stripe waveguides equipped with an appropriate periodic overlay, represent the most elementary and at the same time most important example of contradirectionally coupled devices in integrated optics. We will start our discussion of Bragg gratings with the derivation of the coupled mode propagator and derive from it the most important properties of uniform Bragg gratings. The discussion of these gratings operated at oblique incidence will yield a rich variety of phenomena such as TE-TM polarization conversion, the occurrence of a Brewster angle for TE polarization and the Goos-Hänchen shift of finite beams. At the end of this section we will show how to tackle compound Bragg gratings by stacking the propagators and add a few remarks on curved Bragg gratings.

### 7.2.1 Coupled Mode Propagator

In a first step we will now derive a closed expression for the coupled mode propagator  $\mathcal{U}_{\text{BG}}$  for a lossless first-order Bragg grating<sup>4</sup> of period  $\Lambda$ . The coupled mode equations (7.11) for the contradirectionally coupled waveguide array then simplify to

$$-i \frac{d}{dz} \begin{pmatrix} \tilde{a}^{(+)} \\ \tilde{a}^{(-)} \end{pmatrix} = \begin{pmatrix} \delta & \chi \\ -\chi^* & -\delta \end{pmatrix} \begin{pmatrix} \tilde{a}^{(+)} \\ \tilde{a}^{(-)} \end{pmatrix} \quad (7.23)$$

where

$$\delta = q - Q/2$$

designates the detuning of the affected eigenmode from half the grating constant  $Q = 2\pi/\Lambda$ , and  $\chi$  stands for the first Fourier coefficient of the periodic perturbation. The resonance condition for the Bragg grating

$$Q = 2q \quad (7.24)$$

is designated as the Bragg condition. The conservation law for the optical power inside the Bragg grating has the form

$$\frac{d}{dz} (|\tilde{a}^{(+)}|^2 - |\tilde{a}^{(-)}|^2) = 0 \quad (7.25)$$

The effective detuning, i.e., the detuning under the influence of the Bragg grating, is found by solving the characteristic equation

$$\begin{vmatrix} \delta & \chi \\ -\chi^* & -\delta \end{vmatrix} = 0.$$

Its solutions are given by

$$\delta_{\text{eff}} = \pm \sqrt{\delta^2 - |\chi|^2}. \quad (7.26)$$

In contrast to the corresponding expression (6.34) for the codirectional coupler the effective detuning of a (lossless) Bragg grating can become purely imaginary if  $\delta^2 < |\chi|^2$ .

On the basis of the eigenmode analysis, we can build up the fundamental solutions of the coupled mode equations (7.23), and from these the propagator. Thus the coupled mode propagator of a lossless first-order Bragg grating is given by

$$\mathcal{U}_{\text{BG}} = \begin{pmatrix} A^\Theta & A^\otimes \\ A^{\otimes*} & A^{\Theta*} \end{pmatrix}. \quad (7.27)$$

In the oscillatory regions ( $\delta^2 \geq |\chi|^2$ ), we obtain

$$\begin{aligned} A^\Theta &= \cos(\delta_{\text{eff}} z) - i\delta \sin(\delta_{\text{eff}} z)/\delta_{\text{eff}} \\ A^\otimes &= i\chi \sin(\delta_{\text{eff}} z) e^{iQz}/\delta_{\text{eff}} \end{aligned}$$

4. For higher order Bragg gratings, the additional coupling to the radiation modes (see Section 7.1.2) requires a more sophisticated analysis.

for the coefficients in Equation (7.27), and

$$\begin{aligned} A^\Theta &= \cosh(|\delta_{\text{eff}}|z) - i\delta \sinh(|\delta_{\text{eff}}|z)/|\delta_{\text{eff}}| \\ A^\otimes &= i\chi \sinh(|\delta_{\text{eff}}|z)/|\delta_{\text{eff}}| \end{aligned}$$

inside the stopbands ( $\delta^2 < |\chi|^2$ ).

The reflectivity of the Bragg grating is given by  $R = |A^\Theta/A^\otimes|^2$ . For a device of a length  $L$ , we obtain

$$R = \frac{|\chi|^2 \sin^2(\delta_{\text{eff}} L)}{\delta_{\text{eff}}^2 \cos^2(\delta_{\text{eff}} L) + \delta^2 \sin^2(\delta_{\text{eff}} L)} \quad (7.28)$$

inside the oscillatory regions  $\delta^2 \geq |\chi|^2$  and

$$R = \frac{|\chi|^2 \sinh^2(|\delta_{\text{eff}}|L)}{|\delta_{\text{eff}}|^2 \cosh^2(|\delta_{\text{eff}}|L) + \delta^2 \sinh^2(|\delta_{\text{eff}}|L)}$$

inside the stopbands  $\delta^2 < |\chi|^2$ . Obviously, the reflectivity is a symmetric function,  $R(\delta) = R(-\delta)$ . The transmittance of the Bragg grating is given by

$$T = 1 - R. \quad (7.29)$$

This formula can easily be proven by using the conservation law (7.25).

We will now briefly discuss some properties of both the dispersion curve (7.26) and filter characteristic (7.28). Figure 7.7 shows both curves on a common  $\delta/\chi$ -axis for this purpose. We see, roughly speaking, that the reflectivity inside the stopband is high but not perfect. By using the filter characteristic (7.28) we can easily see that the reflectivity at the center of the stopband is given by

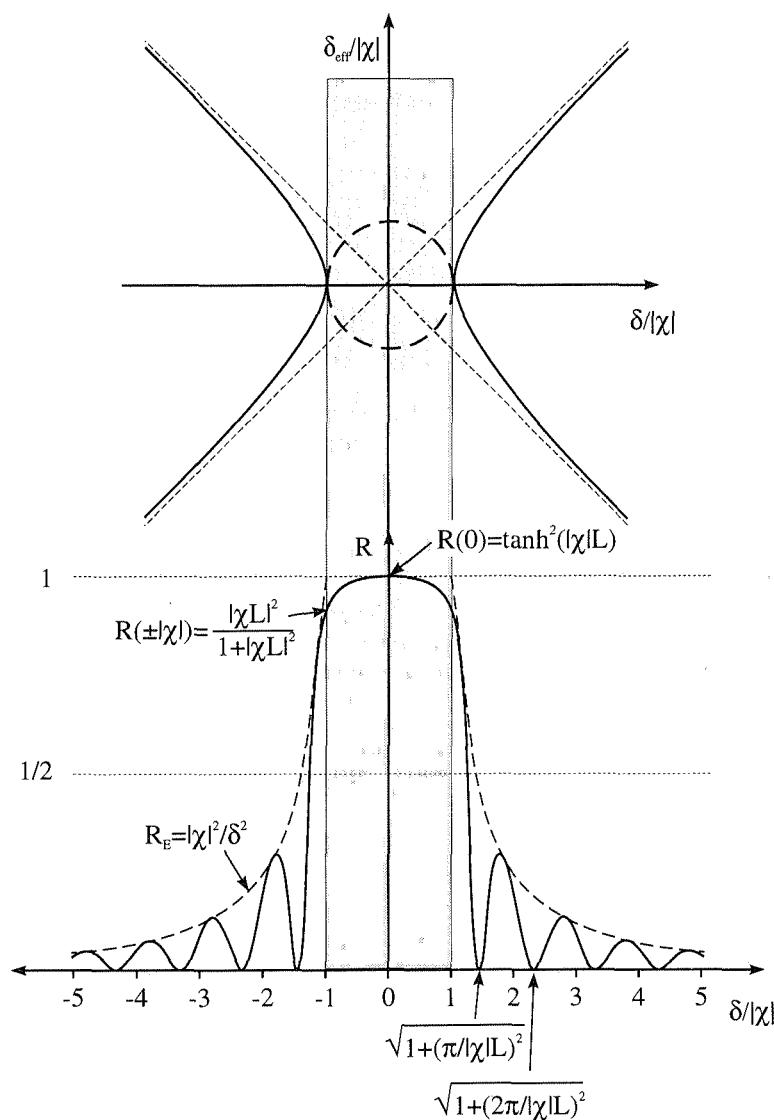
$$R(0) = \tanh^2(|\chi|L). \quad (7.30)$$

At the boundary of the stopband we obtain

$$R(|\chi|) = \frac{|\chi|L|^2}{1 + |\chi|L|^2}. \quad (7.31)$$

Obviously, the relative strength of a Bragg grating is characterized by the parameter  $\chi L$ . Figure 7.8 shows a plot of both functions. We see that the Bragg reflector works poorly for  $|\chi|L < 1$ . For  $|\chi|L > 10$  it acts as an excellent reflector over the full stopband. For practical devices, values of  $|\chi|L \approx 3$  represent a compromise between the requirements of small size and filter function. Figure 7.7 shows the filter characteristic for  $|\chi|L = 3$ . The shape of the imaginary part of the propagation constant (see Figure 7.7), which is indicated in the stopbands by dashed lines, shows why the reflectivity in the center of the stopband is always higher than at its borders. The zeros of the filter characteristic (7.28) are found at  $\delta_{\text{eff}} L = N\pi$ , i.e.,

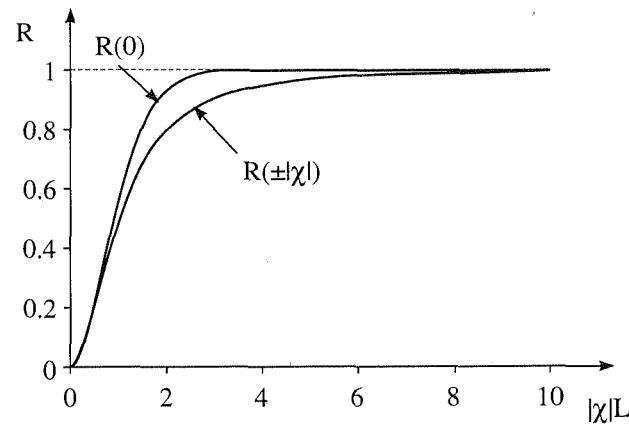
$$\delta_N = \sqrt{1 + \left(\frac{N\pi}{|\chi|L}\right)^2}. \quad (7.32)$$

Figure 7.7. Dispersion curve and filter characteristic of a Bragg grating with  $|\chi|L = 3$ .

We see that the density of zeros, i.e., the oscillation frequency of the filter curve, increases with increasing  $|\chi|L$ . The envelope of the filter curve is given by

$$R_E = \begin{cases} |\chi|^2/\delta^2 & \text{for } \delta^2 > |\chi|^2 \\ 1 & \text{otherwise.} \end{cases} \quad (7.33)$$

It is indicated in Figure 7.7 by a dashed line. We should note that it is independent of the parameter  $|\chi|L$ , i.e., all filter curves possess the same envelope.

Figure 7.8.  $R(|\chi|)$  and  $R(0)$  as a function of  $|\chi|L$ .

## 7.2.2 Operation at Oblique Incidence

At first glance, it may look as if the treatment of Bragg gratings at oblique incidence is a more or less straightforward generalization of the results obtained for their counterparts operated at normal incidence. However, we will see in the following that Bragg gratings operated at oblique incidence exhibit a couple of new phenomena such as a Brewster angle for TE polarization, TE-TM polarization conversion and not the least a Goos-Hänchen shift of finite beams.

### Coupling Coefficients

Let us start with a discussion of the coupling coefficient,<sup>5</sup> strictly speaking of its

5. The derivation of Equation (7.34) requires lengthy calculations. It is found in a paper of L. A. Weller-Brophy and D. G. Hall [2].

first Fourier coefficient

$$\chi_{mn} = \frac{\int_{-\infty}^{\infty} dx k_0 \epsilon_1 \mathbf{e}_m^{(-)*} \cdot \mathbf{e}_n^{(+)}}{\int_{-\infty}^{\infty} dx (\mathbf{e}_m^{(+)*} \times \mathbf{h}_m^{(-)}) \cdot \mathbf{i}_z}. \quad (7.34)$$

In Equation (7.34),  $\mathbf{e}_m^{(+)}$  and  $\mathbf{e}_m^{(-)}$  designate the slowly varying amplitudes of the electric field of the forward and backward traveling waves.  $\mathbf{h}_m^{(+)}$  and  $\mathbf{h}_m^{(-)}$  stand for the corresponding amplitudes of the magnetic field. The parameter  $\epsilon_1$  designates the first Fourier coefficient of the dielectric profile

$$\epsilon(z) = \sum_{m=-\infty}^{\infty} \epsilon_m \exp\left(i \frac{2m\pi}{\Lambda} z\right). \quad (7.35)$$

To derive expressions for the angular behavior of the coupling coefficient, let us now examine the electric and magnetic fields of the incident and reflected wave for a Bragg grating which is operated at oblique incidence. By using Figure 7.9 we see that the electric field of a transverse electric (TE) mode is given by

$$\mathbf{e}_{m,\text{TE}}^{(\pm)} \propto \begin{pmatrix} 0 \\ \cos \theta_m \\ \pm \sin \theta_m \end{pmatrix}. \quad (7.36)$$

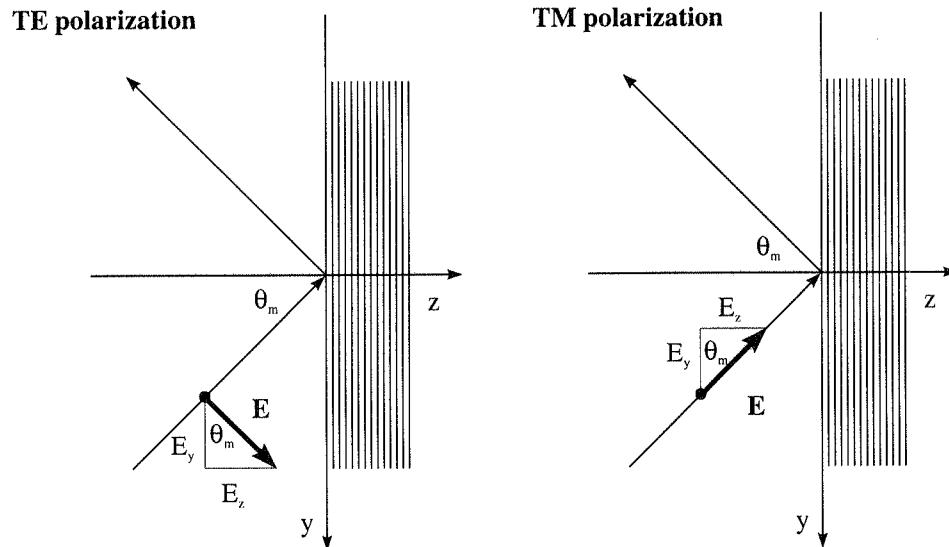


Figure 7.9. The electric field components  $E_y$  and  $E_z$  for a Bragg grating at oblique incidence.

By using the representation (3.105), we see that the electric field of a transverse magnetic (TM) mode is given by

$$\mathbf{e}_{m,\text{TM}}^{(\pm)} \propto \begin{pmatrix} \zeta_m^{(\pm)} \\ -\sin \theta_m \\ \pm \cos \theta_m \end{pmatrix} \quad (7.37)$$

with

$$\zeta_m^{(\pm)} = \frac{q_m h_{m,\parallel}^{(\pm)}}{\frac{\partial h_{m,\parallel}}{\partial x}}.$$

By using Equation (3.106) we see that the representation of the magnetic field of a TE mode is similar to that of the electric field of a TM mode, and vice versa.

Thus, we can conclude in a first step that the normalization integral forming the denominator of Equation (7.34) exhibits a cosine-like angular behavior for both polarizations, i.e.,

$$\int_{-\infty}^{\infty} dx (\mathbf{e}_m^{(+)*} \times \mathbf{h}_m^{(-)}) \cdot \mathbf{i}_z \propto \cos \theta_m \quad (7.38)$$

By calculating the scalar product  $\mathbf{e}_m^{(-)*} \cdot \mathbf{e}_n^{(+)}$  occurring within the nominator of Equation (7.34) in the same way, we obtain for the *TE-TE coupling*

$$\chi_{mn}^{(\text{TE-TE})} \propto \frac{\cos(\theta_m + \theta_n)}{\cos \theta_m}, \quad (7.39)$$

for the *TE-TM coupling*

$$\chi_{mn}^{(\text{TE-TM})} \propto \frac{\sin(\theta_m + \theta_n)}{\cos \theta_m}, \quad (7.40)$$

and for the *TM-TM coupling*

$$\chi_{mn}^{(\text{TM-TM})} \propto \frac{\zeta^2 - \cos(\theta_m + \theta_n)}{\cos \theta_m}. \quad (7.41)$$

Let us now elucidate the mode coupling for Bragg gratings operated at oblique incidence from a more physical point of view. For TE polarization, the electric field of the eigenmodes is parallel to the waveguide layers, i.e., parallel to the plane of incidence. In the notation we used for the treatment of reflection and refraction of plane waves at a dielectric interface (see Section 2.2.2) TE polarized light is designated as p-polarized light where *p* stands for parallel to the plane of incidence. According to our discussion in Section 2.2.2, we then expect that the reflectance will vanish for TE polarization at the Brewster angle which is found at

$$\tan \theta_B = \frac{\sqrt{\epsilon_{\text{out}}}}{\sqrt{\epsilon_{\text{in}}}} \quad (7.42)$$

If, at a Bragg grating, we replace the dielectric constants  $\epsilon_{\text{in}}$  and  $\epsilon_{\text{out}}$  by the effective dielectric constants of the input and output regions, we expect as a consequence of the small refractive index contrast inside the grating that the Brewster angle is found at

$$\theta_B \approx 45^\circ. \quad (7.43)$$

For pure Bragg reflection ( $m = n$ ), the TE-TE coupling coefficient (7.39) does indeed have a zero at this angular position. To summarize this discussion, *for TE polarization, Bragg gratings exhibit a Brewster angle (of vanishing reflectance) at 45°.*

Let us now discuss the other couplings. From Equation (7.40), we see directly that *for non-normal incidence Bragg gratings yield polarization conversion*, and we should note that the corresponding coupling coefficient is of the same order of magnitude as the TE-TE and TM-TM coupling coefficients. By using Equation (7.41) we see that the TM-TM coupling also shows a zero whose physical origin is not yet clear. Measurements reported in [2], however, indicate the existence of such a zero.

#### 4-wave coupling

The widths of the stopbands for TE-TE, TE-TM and TM-TM coupling are in most cases of the same order of magnitude as the spacing between the effective propagation constant between TE and TM modes. Therefore, we have to treat the TE-TM coupling within the framework of a 4-wave coupled mode theory. By using the general coupled mode Equation (7.11) we see that the coupled mode equations describing the interaction of forward and backward traveling TE and TM modes are given by

$$-i \frac{d}{dz} \begin{pmatrix} \tilde{a}_{\text{TE}}^{(+)} \\ \tilde{a}_{\text{TM}}^{(+)} \\ \tilde{a}_{\text{TE}}^{(-)} \\ \tilde{a}_{\text{TM}}^{(-)} \end{pmatrix} = \begin{pmatrix} \delta_{\text{TE}} & 0 & \chi_{\text{TE-TE}} & \chi_{\text{TE-TM}} \\ 0 & \delta_{\text{TM}} & \chi_{\text{TE-TM}}^* & \chi_{\text{TM-TM}} \\ -\chi_{\text{TE-TE}}^* & -\chi_{\text{TE-TM}} & -\delta_{\text{TE}} & 0 \\ -\chi_{\text{TE-TM}}^* & -\chi_{\text{TM-TM}}^* & 0 & -\delta_{\text{TM}} \end{pmatrix} \begin{pmatrix} \tilde{a}_{\text{TE}}^{(+)} \\ \tilde{a}_{\text{TM}}^{(+)} \\ \tilde{a}_{\text{TE}}^{(-)} \\ \tilde{a}_{\text{TM}}^{(-)} \end{pmatrix}$$

where  $\chi_{\text{TE-TE}}$ ,  $\chi_{\text{TE-TM}}$  and  $\chi_{\text{TM-TM}}$  designate the leading Fourier coefficients responsible for the TE-TE, TE-TM and TM-TM coupling. We see that the TE and TM modes are not codirectionally coupled.

The characteristic equation which yields the eigenmodes of the contradirectionally coupled system turns out to be biquadratic. It has four solutions

$$\delta_{\text{eff}} = \pm \sqrt{(A \pm \sqrt{B}) / 2} \quad (7.44)$$

two each with alternating signs. The two coefficients  $A$  and  $B$  occurring in Equation (7.44) are given by

$$A = \delta_{\text{TE}}^2 + \delta_{\text{TM}}^2 - |\chi_{\text{TE-TE}}|^2 - |\chi_{\text{TM-TM}}|^2 - 2|\chi_{\text{TE-TM}}|^2 \quad (7.45)$$

and

$$\begin{aligned} B = & [\delta_{\text{TE}}^2 - \delta_{\text{TM}}^2 - |\chi_{\text{TE-TE}}|^2 + |\chi_{\text{TM-TM}}|^2]^2 \\ & - 4|\chi_{\text{TE-TM}}|^2 [(\delta_{\text{TE}} - \delta_{\text{TM}})^2 - |\chi_{\text{TE-TE}}|^2 - |\chi_{\text{TM-TM}}|^2] \\ & + 4[\chi_{\text{TE-TE}} \chi_{\text{TM-TM}} \chi_{\text{TE-TM}}^* + \text{c.c.}] \end{aligned} \quad (7.46)$$

Figure 7.10 shows the dispersion curves for the forward and backward traveling TE and TM modes in the region of crossover as an illustration. The dispersion curves of the decoupled modes are indicated by dashed lines. Figure 7.10a shows a typical scenario for normal incidence ( $\theta = 0$ ). We see that the TE-TM coupling ( $\chi_{\text{TE-TM}} = 0$ ) vanishes at this angle. As a consequence, the TE-TE and TM-TM couplings become independent of each other. For the other couplings we have chosen typical values ( $\chi_{\text{TE-TE}} = (q_{\text{TE}} - q_{\text{TM}})/4$  and  $\chi_{\text{TM-TM}} = \chi_{\text{TE-TE}}/5$ ) in this example. Figure 7.10a shows the dispersion curves of the Bragg grating at the Brewster angle ( $\theta = 45^\circ$ , i.e.,  $\chi_{\text{TE-TE}} = 0$ ). We again chose typical values for the other couplings ( $\chi_{\text{TE-TM}} = \chi_{\text{TM-TM}} = (q_{\text{TE}} - q_{\text{TM}})/4$ ). By comparing dispersion curves with those of the decoupled modes we see that the stopbands of the TE-TM and the TM-TM couplings repel each other. By calculating the shift for weak coupling within the framework of a first-order Taylor expansion, we obtain the following expressions for the band shift in terms of equivalent wave numbers:

$$\begin{aligned} \Delta k_{\text{TE-TE}} &= -\frac{1}{\bar{n}} \frac{|\chi_{\text{TE-TM}}|^2}{\delta_{\text{TE}} - \delta_{\text{TM}}} \\ \Delta k_{\text{TM-TM}} &= -\Delta k_{\text{TE-TE}} \\ \Delta k_{\text{TE-TM}} &= -\frac{1}{2\bar{n}} \frac{|\chi_{\text{TE-TE}}|^2 - |\chi_{\text{TM-TM}}|^2}{\delta_{\text{TE}} - \delta_{\text{TM}}} \end{aligned} \quad (7.47)$$

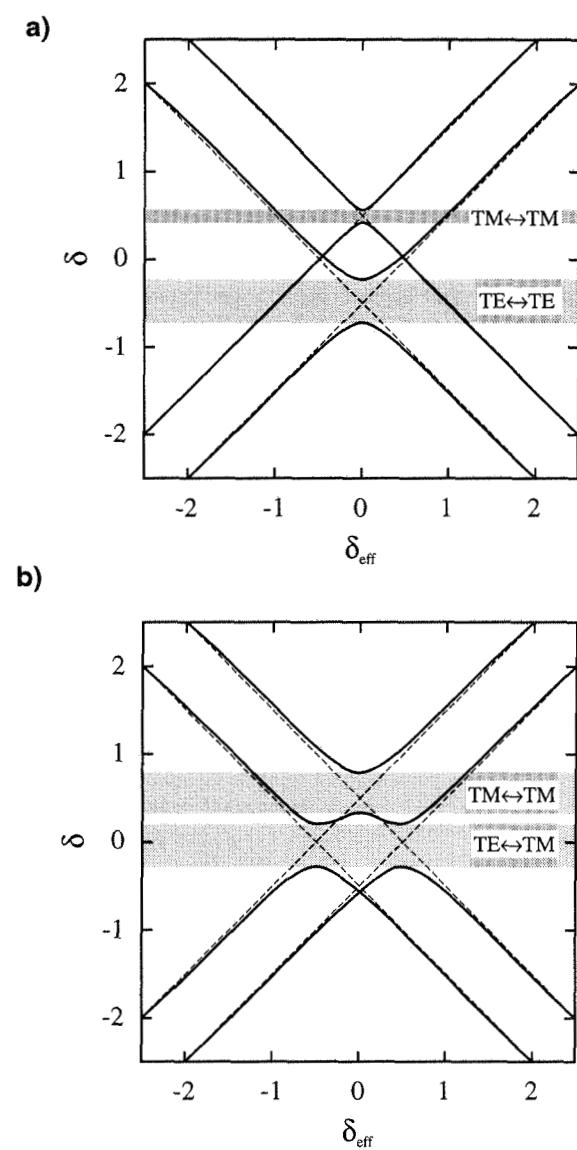
The mean effective refractive index is given by

$$\bar{n} = \frac{n_{\text{TE}} + n_{\text{TM}}}{2}$$

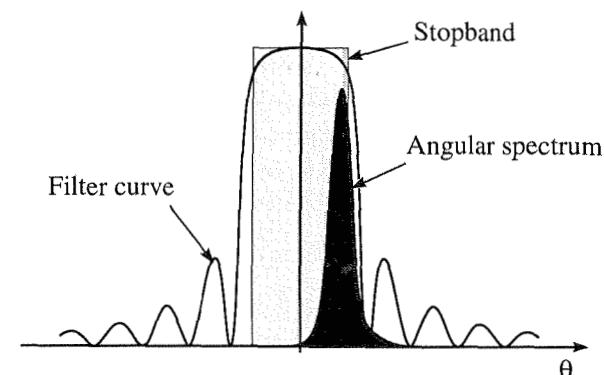
We see that the shift of the stopbands caused by the TE-TE and TM-TM couplings increases with increasing TE-TM coupling. The stopband caused by TE-TM coupling is shifted only if TE-TE and TM-TM coupling are unbalanced.

#### Goos-Hänchen Shift

As the last topic of this section let us discuss the handling of finite beams by Bragg gratings operated at oblique incidence. We have seen in Section 4.2 that a finite beam has an angular spectrum of finite extent. In physical terms, a finite beam contains plane waves traveling in different directions. Figure 7.11 shows the reflectance of a Bragg grating as a function of the angle of incidence as an illustration. In this diagram, the Bragg grating exhibits a stopband around the Bragg reflection



**Figure 7.10.** Dispersion curves for a Bragg grating at normal incidence (a) and at the Brewster angle (b).



**Figure 7.11.** Reflectance of a Bragg grating and angular spectrum of the incident beam vs. angle of incidence.

angle  $\theta_B$ . Figure 7.11 also shows the angular spectrum of a finite beam traveling to the Bragg grating under the angle  $\theta$  located inside the stopband of the Bragg grating. Nevertheless, the incident beam is not totally reflected by the Bragg grating since a part of its angular spectrum is located outside the stopband. Thus we can draw a first conclusion: *The operation of Bragg gratings with finite beams results in an effective constriction of their stopbands. In particular, if the angular width of a stopband becomes smaller than that of the incident beam, the Bragg grating will no longer exhibit a stopband.*

For a rough estimation of the Goos-Hänchen shift, i.e., the lateral shift of the beam by reflection, we will now assume that the Bragg grating is operated at the Bragg reflection angle and that the angular width of the incident beam is much smaller than the width of the stopband ( $\delta^2 \ll |\chi|^2$ ). We will further assume that the Bragg grating has so many periods that it exhibits a “perfect” stopband ( $|\chi|L > 10$ ) such that

$$\cosh(|\delta_{\text{eff}}|L) \approx \exp(|\delta_{\text{eff}}|L) \approx \sinh(|\delta_{\text{eff}}|L).$$

By using these approximations, the reflection coefficient

$$r = - \left( \frac{A^\otimes}{A^\ominus} \right)^* \approx \frac{i\chi^*}{\sqrt{|\chi|^2 - \delta^2} + i\delta} \propto e^{i\phi_R} \quad (7.48)$$

becomes a pure phase shift with

$$\phi_R = \tan^{-1} \left( \frac{\sqrt{|\chi|^2 - \delta^2}}{\delta} \right) \approx \tan^{-1} \left( \frac{|\chi|}{\delta} \right).$$

By using the expression (2.55) for the Goos-Hänchen shift which was derived in

Section 2.2.2 we obtain

$$d_{\text{GH}} = \left. \frac{\partial \phi_R}{\partial q_x} \right|_{q_B} = \left. \frac{\partial \phi_R}{\partial \delta} \frac{\partial \delta}{\partial q_x} \right|_{q_B}$$

where  $q_B$  designates the tangential vector for which Bragg reflection will occur. If we keep in mind that  $\delta$  designates the detuning of the  $z$ -component of the propagation constant, we can easily derive an expression for the second term of the right-hand side of this equation:

$$\left. \frac{\partial \delta}{\partial q_x} \right|_{q_B} = \left. \frac{\partial \sqrt{q^2 - q_x^2}}{\partial q_x} \right|_{q_B} = -\frac{q_B}{\sqrt{q^2 - q_B^2}} = -\tan \theta_B$$

By inserting all these, we obtain

$$d_{\text{GH}} = \frac{\tan \theta_B}{|\chi|}. \quad (7.49)$$

As expected, no Goos-Hänchen shift will occur for normal incidence ( $\theta_B = 0^\circ$ ). The Goos-Hänchen shift will then increase with increasing angle of incidence. We see furthermore that the Goos-Hänchen shift will successively increase with decreasing coupling coefficient  $|\chi|$ . This becomes clear if we keep in mind that the penetration depth of the optical field into the Bragg grating will increase with decreasing “strength” of the grating.

### 7.2.3 Compound Bragg Gratings

In the following we will briefly discuss the theory of compound Bragg gratings. In analogy to our description of codirectionally coupled networks (see Section 6.5) we intend to derive the propagators describing compound Bragg gratings by stacking their constituent propagators. However, we should note that it then becomes essential to keep careful track of any phase shifts, since any straight waveguide will form an effective Fabry-Perot resonator inside the optical path and thus affect device operation.

Most compound Bragg gratings can be described by stacking two types of elementary device. The uniform Bragg grating whose coupled mode propagator is given by

$$\mathcal{U}_{\text{BG}} = \begin{pmatrix} A^\ominus & A^\otimes \\ A^{\otimes*} & A^{\ominus*} \end{pmatrix}. \quad (7.50)$$

Its mathematical structure was analyzed in Section 7.2.1. The coefficients were given by

$$\begin{aligned} A^\ominus &= \cos(\delta_{\text{eff}} z) - i\delta \sin(\delta_{\text{eff}} z)/\delta_{\text{eff}} \\ A^\otimes &= i\chi \sin(\delta_{\text{eff}} z) e^{iQz}/\delta_{\text{eff}}, \end{aligned}$$

in the oscillatory regions ( $\delta^2 \geq |\chi|^2$ ) and by

$$\begin{aligned} A^\ominus &= \cosh(|\delta_{\text{eff}}|z) - i\delta \sinh(|\delta_{\text{eff}}|z)/|\delta_{\text{eff}}| \\ A^\otimes &= i\chi \sinh(|\delta_{\text{eff}}|z)/|\delta_{\text{eff}}|. \end{aligned}$$

inside the stopbands ( $\delta^2 < |\chi|^2$ ). The spacer, i.e., a waveguide of length  $L$  and effective propagation constant  $q$ , represents the second elementary device. Obviously, its propagator is given by

$$\mathcal{U}_- = \begin{pmatrix} e^{iqL} & 0 \\ 0 & e^{-iqL} \end{pmatrix}. \quad (7.51)$$

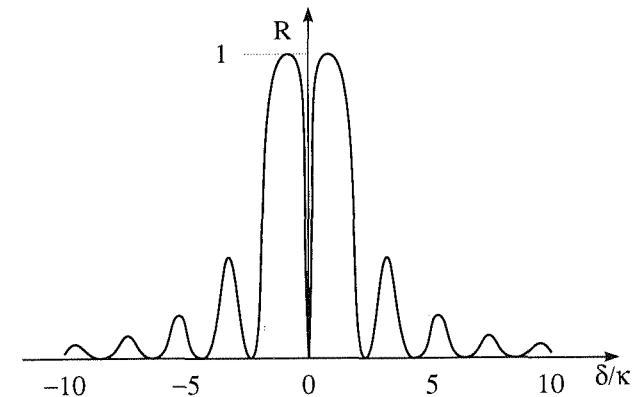


Figure 7.12. Quarter-wave shifted Bragg grating.

The propagator of a compound Bragg grating can now be calculated by stacking the constituent propagators. Let us examine the propagator of a quarter-wave shifted Bragg grating as an illustrative example containing both types of elementary device. This device consists of two Bragg gratings of the same kind which are separated by a spacer whose length represents exactly half the period of the Bragg gratings (see Figure 7.12). The propagator describing the compound device is then given by

$$\begin{pmatrix} \tilde{A}^\ominus & \tilde{A}^\otimes \\ \tilde{A}^{\otimes*} & \tilde{A}^{\ominus*} \end{pmatrix} = \begin{pmatrix} A^\ominus & A^\otimes \\ A^{\otimes*} & A^{\ominus*} \end{pmatrix} \begin{pmatrix} e^{iq\Lambda/2} & 0 \\ 0 & e^{-iq\Lambda/2} \end{pmatrix} \begin{pmatrix} A^\ominus & A^\otimes \\ A^{\otimes*} & A^{\ominus*} \end{pmatrix}. \quad (7.52)$$

where

$$\begin{aligned} \tilde{A}^\ominus &= A^\ominus e^{iq\Lambda/2} + |A^\otimes|^2 e^{-iq\Lambda/2} \\ \tilde{A}^\otimes &= A^\ominus A^\otimes e^{iq\Lambda/2} + A^{\otimes*} A^\ominus e^{-iq\Lambda/2} \end{aligned}$$

and

$$\frac{q\Lambda}{2} = \frac{\pi}{2} + \frac{q\delta}{2}.$$

In the center of the stopband ( $\delta = 0$ ) the coefficients are given by  $A^\Theta = \cosh(|\chi|L)$  and  $A^\otimes = i \sinh(|\chi|L)$  and thus  $\tilde{A}^\Theta = i$  and  $\tilde{A}^\otimes = 0$ . We can conclude that the compound device has a passband in the center of the stopband of the individual Bragg gratings. Figure 7.12 shows the filter curve  $R = |\tilde{A}^\otimes/\tilde{A}^\Theta|^2$  calculated on the basis of the propagator (7.52).

Obviously, it is easy to calculate the filter curve for chirped and tapered Bragg gratings (see Section 7.1.4) in the same way. It is also possible to analyze more complex networks as described in Section 6.5 by attaching the required backward traveling modes to the corresponding matrices. As pointed out before, it will then become necessary to keep track of any phase shifts.

#### 7.2.4 Curved Bragg Gratings

In Section 7.2.2 we examined the handling of finite beams by Bragg gratings, and we saw that the angular width of the incoming beam must be significantly smaller than that of the stopband of the Bragg grating to ensure proper operation of the device. In most cases, a lens will be required on the input side of the Bragg grating in order to transform the beam emitted from the input focus into a parallel beam, i.e., to reduce the width of the angular spectrum of the incoming beam. On the output side a second lens is usually needed to focus the outgoing beam onto a waveguide endface or a photodiode (see Figure 7.13a) at the output focus. In the following we will discuss curved Bragg gratings (see Figure 7.13b), i.e., Bragg gratings which incorporate the optical system described previously.

Let us now recall our discussion of the Fraunhofer diffraction integral in Section 4.3, where we saw that the phase fronts of the far-field of a finite beam propagating in a homogeneous space represent a sequence of equidistant spheres around its focus. A finite beam propagating in a homogeneous slab waveguide will exhibit a sequence of circles as phase fronts instead of spheres. We should note that the preceding representation applies to both the incoming and outgoing beams of the Bragg grating. As a consequence, all rays representing the optical field of the incoming beam start at the input focus  $I$ , and all rays describing the outgoing beam will end at the output focus  $R$ . In the following we will examine a family of rays which propagate from the input focus  $I$  to a certain line of the Bragg grating and further to the output focus  $R$  (see Figure 7.14). Obviously, the input field is mapped perfectly onto the output focus if the paths of all rays have the same length

$$r_I + r_R = \overline{IR} + \frac{m\lambda}{n} \quad (7.53)$$

except for an integral multiple of the wavelength  $\lambda/n$  of a guided mode of the slab waveguide. In Equation (7.53) the parameters  $r_I$  and  $r_R$  describe the distance

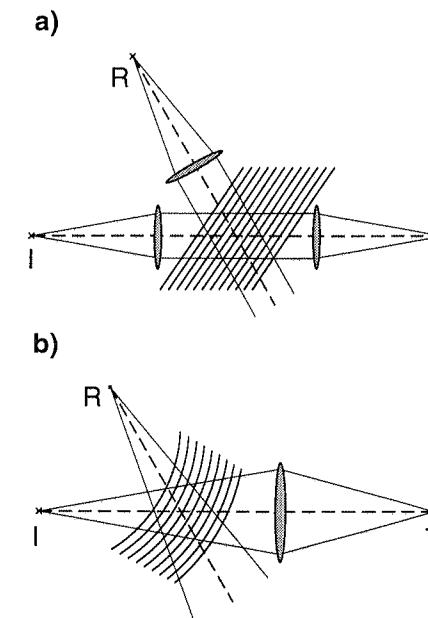


Figure 7.13. Straight Bragg grating with optical system (a) and curved Bragg grating (b).

between the input and the output focus and the reflection point.  $\overline{IR}$  stands for the spacing between the input focus  $I$  and output focus  $R$  and  $m$  for an integer specifying the lines of the Bragg grating. Figure 7.14 shows one representative of this family of rays. Obviously, a single grating line is described by

$$r_I + r_R = \text{const};$$

i.e., each grating line represents an ellipse with the two focal points  $I$  and  $R$ . By using Equation (7.53) we can further conclude that *a curved Bragg grating mapping the input focus  $I$  onto the output focus  $R$  is given by a family of confocal ellipses with the two focal points  $I$  and  $R$ . Along the long half axis, the spacing between two adjacent ellipses is half the wavelength  $\lambda/(2n)$  of a guided mode of the slab waveguide*. We can easily show by means of standard vector algebra that the rays which are reflected at the grating lines satisfy the law of reflection, i.e.,  $\alpha_I = \alpha_R$ .

Holographic construction represents an alternative way of obtaining the shape of a focussing Bragg grating. Within the framework of this construction the volume hologram, i.e., the curved Bragg grating, is described by the intensity distribution of an optical field which is given by the superposition of both the incoming and outgoing beams. If the Bragg grating is located in the far-field region of the two beams, we can again represent the beams by spherical waves, and we obtain

$$P \propto |e^{ik_0nr_I} + e^{-ik_0nr_R}|^2$$

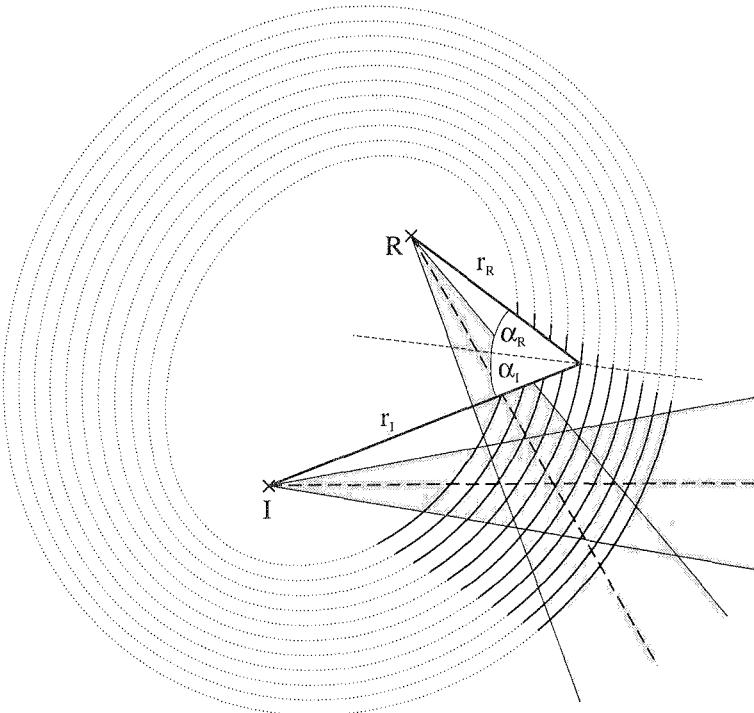


Figure 7.14. Construction of the curved Bragg grating.

$$\propto 2 \{1 + \cos(k_0 n(r_I + r_R))\}$$

The shape of the grating lines is then defined by contours of constant intensity, i.e., by

$$k_0 n(r_I + r_R) = k_0 \overline{IR} + 2\pi m.$$

This is exactly the condition defined by Equation (7.53) for the shape of the grating lines.

We have shown in this section how to design focussing Bragg gratings, i.e., Bragg gratings which incorporate the imaging system for mapping the input focus onto the output focus. We will add three remarks. First, our discussion of Bragg gratings at oblique incidence can serve as a guideline for estimating of the efficiencies of the TE-TE, TE-TM and TM-TM couplings. Second, curved Bragg gratings have rarely been realized up to now since their fabrication requires direct electron beam writing or complicated holographic exposure. Third, focussing grating couplers,

i.e., devices which launch light into or out of waveguides, are designed by applying similar methods.

### 7.3 CONTRADIRECTIONAL COUPLER

The contradirectional coupler represents the most complex example of contradirectional coupling that will be examined in this chapter. This device consists of a codirectional coupler equipped with a periodic overlay (see Figure 7.15) whose period

$$\Lambda = \frac{2\pi}{q_1(\lambda_{\leftrightarrow}) + q_2(\lambda_{\leftrightarrow})} \quad (7.54)$$

is chosen such<sup>6</sup> that the resonance condition for contradirectional coupling of the guided mode of waveguide 1 to the guided mode of waveguide 2 is satisfied at a design wavelength  $\lambda_{\leftrightarrow}$ . The effective propagation constants of these guided modes are designated as  $q_1$  and  $q_2$ , respectively. By using Figure 7.15 we see that the contradirectional coupler will exhibit not only the desired Bragg exchange coupling ( $1^{(+)} \leftrightarrow 2^{(-)}$ ) between the guided modes of waveguides 1 and 2 but also various types of other coupling, namely, Bragg reflections for both waveguides ( $1^{(+)} \leftrightarrow 1^{(-)}$ ,  $2^{(+)} \leftrightarrow 2^{(-)}$ ) and codirectional coupling ( $1^{(+)} \leftrightarrow 2^{(+)}$ ,  $1^{(-)} \leftrightarrow 2^{(-)}$ ) between waveguides 1 and 2. The expressions attached in parentheses indicate the kind of coupling involved.<sup>7</sup>

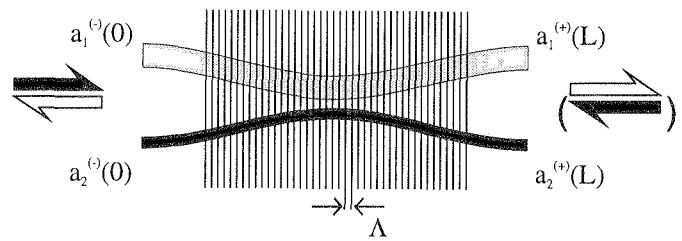


Figure 7.15. Contradirectional coupler.

The coupled mode equations describing the contradirectional coupler are then

6. We will restrict ourselves in the following to first-order contradirectional coupling since higher order contradirectional couplings will be affected by radiation (see Section 7.1.2).

7. For example, the expression  $1^{(+)} \leftrightarrow 2^{(-)}$  stands for guided mode of waveguide 1 propagating forwards (+) couples to guided mode of waveguide 2 propagating backwards (-).

given by

$$-i \frac{d}{dz} \begin{pmatrix} \tilde{a}_1^{(+)} \\ \tilde{a}_2^{(+)} \\ \tilde{a}_1^{(-)} \\ \tilde{a}_2^{(-)} \end{pmatrix} = \begin{pmatrix} \delta_1 & \kappa & \chi_{11} & \chi_{12} \\ \kappa & \delta_2 & \chi_{12}^* & \chi_{22} \\ -\chi_{11}^* & -\chi_{12} & -\delta_1 & -\kappa \\ -\chi_{12}^* & -\chi_{22}^* & -\kappa & -\delta_2 \end{pmatrix} \begin{pmatrix} \tilde{a}_1^{(+)} \\ \tilde{a}_2^{(+)} \\ \tilde{a}_1^{(-)} \\ \tilde{a}_2^{(-)} \end{pmatrix} \quad (7.55)$$

where  $\chi_{12}$  designates the leading Fourier coefficient describing the Bragg exchange coupling. The Fourier coefficients  $\chi_{11}$  and  $\chi_{22}$  describe the Bragg reflections and  $\kappa$  stands for the codirectional coupling of the two waveguides. The parameters

$$\delta_m = q_m - \frac{\pi}{\Lambda}$$

with  $m = 1, 2$  describe the detuning of the eigenmodes from half the grating constant.

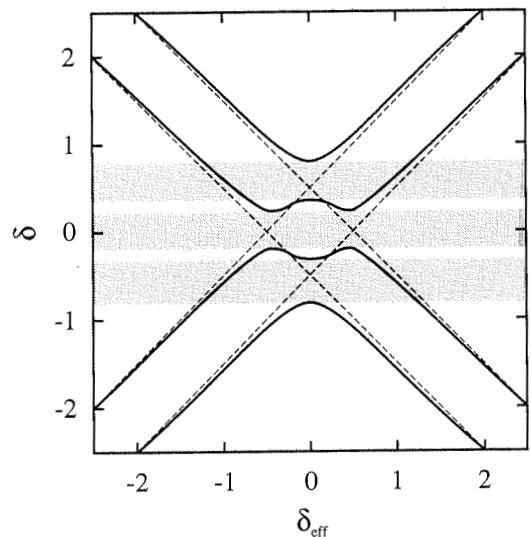


Figure 7.16. Dispersion curves for a strongly coupled contradirectional coupler.

Again, the characteristic equation which yields the eigenmodes of the contradirectionally coupled system turns out to be biquadratic. It has four solutions

$$\delta_{\text{eff}} = \pm \sqrt{(A \pm \sqrt{B})/2} \quad (7.56)$$

each two with alternating signs. The two coefficients  $A$  and  $B$  occurring in Equation (7.56) are given by

$$A = \delta_1^2 + \delta_2^2 - |\chi_{11}|^2 - |\chi_{22}|^2 - 2|\chi_{12}|^2 + 2\kappa^2 \quad (7.57)$$

and

$$\begin{aligned} B = & [\delta_1^2 - \delta_2^2 - |\chi_{11}|^2 + |\chi_{22}|^2]^2 \\ & - 4|\chi_{12}|^2 [(\delta_1 - \delta_2)^2 - |\chi_{11}|^2 - |\chi_{22}|^2] \\ & + 4\kappa^2 [(\delta_1 + \delta_2)^2 - |\chi_{11}|^2 - |\chi_{22}|^2] \\ & - 8\kappa [\chi_{12}^* (\delta_1 \chi_{22} + \delta_2 \chi_{11}) + \text{c.c.}] \\ & + 4\kappa^2 [\chi_{11}^* \chi_{22} + \text{c.c.}] + 4 [\chi_{11} \chi_{22} \chi_{12}^{*2} + \text{c.c.}] \end{aligned} \quad (7.58)$$

Figure 7.16 shows the dispersion curves for a strongly coupled contradirectional coupler. The parameters  $\kappa = |\chi_{11}| = |\chi_{12}| = |\chi_{22}| = |\chi_{11}| = (\delta_1 - \delta_2)/2$  of this numerical example are chosen such that the various effects become obvious. Nevertheless, the relative weighting of the parameters is realistic. We see that the stopbands of Bragg reflections are “repelled” by the stopband of Bragg exchange coupling, i.e., the crossover between the individual waveguides caused by contradirectional coupling. The formulas

$$\begin{aligned} \Delta k_{11} &= -\frac{1}{\bar{n}} \frac{|\chi_{12}|^2 + \kappa^2}{\delta_1 - \delta_2} \\ \Delta k_{22} &= -\Delta k_{11} \\ \Delta k_{12} &= -\frac{1}{2\bar{n}} \frac{(\delta_1 - \delta_2)^2}{(\delta_1 - \delta_2)^2 + 4\kappa^2} \frac{|\chi_{11}|^2 - |\chi_{22}|^2}{\delta_1 - \delta_2} \end{aligned} \quad (7.59)$$

with

$$\bar{n} = \frac{n_1 + n_2}{2}$$

describing the shift of the stopband in the weak-coupling limit are similar to those derived for the TE-TM coupling (see Section 7.2.2). However, the codirectional coupling magnifies the spectral shift. In addition to the shift of the stopbands, we observe a significant constriction of the stopband for Bragg exchange coupling.

Let us draw the most important conclusions for the practical design of contradirectional couplers:

- In order to keep the coupling coefficients for the Bragg reflections as small as possible, the periodic overlay must be restricted to the coupling region between the two waveguides forming the directional coupler.
- The directional coupler must be extremely asymmetric to minimize the influence of codirectional coupling on the operation of the device.

Obviously, the remarks on the technological problems mentioned at the beginning of this chapter also apply to contradirectional couplers.

## REFERENCES

- [1] Arnold, V. I., *Geometrical Methods in the Theory of Ordinary Differential Equations*, New York, Springer, 1983.
- [2] Weller-Brophy, L. A., and Hall, D. G., "Local Normal Mode Analysis of Guided Mode Interactions with Waveguide Gratings", *IEEE Journal of Lightwave Technology*, Vol. 6, 1988, pp. 1069–1082.
- [3] Yariv, A., "Coupled Mode Theory of Guided Wave Optics", *IEEE Journal of Quantum Electronics*, Vol. 9, 1973, pp. 919–933.
- [4] Kogelnik, H., "Filter Response of Nonuniform Almost-Periodic Structures", *The Bell System Technical Journal*, Vol. 55, 1976, pp. 109–126.
- [5] Wagatsuma, K., Sakaki, H., and Saito, S., "Mode Conversion and Optical Filtering of Obliquely Incident Waves in Corrugated Waveguide Filters", *IEEE Journal of Quantum Electronics*, Vol. 15, 1979, pp. 632–637.
- [6] Furuya, K., Suematsu, Y., and Shigeo, S., "Integrated Optical Branching Filter Consisting of Three-Dimensional Waveguide and Its Nonradiative Condition", *IEEE Transactions on Circuits and Systems*, Vol. 26, 1979, pp. 1049–1054.
- [7] Hardy, A., and Streifer, W., "Analysis of Waveguided Gaussian Beams Coupled by Misaligned or Curved Gratings", *Journal of the Optical Society of America*, Vol. 69, 1979, pp. 1235–1242.
- [8] Van Roey, J., and Lagasse, P. E., "Coupled Wave Analysis of Obliquely Incident Waves in Thin Film Gratings", *Applied Optics*, Vol. 20, 1981, pp. 423–429.
- [9] Oliner, A. A., and Peng, S. T., "New Physical Effects on Periodically Grooved Optical Planar Waveguides", *Applied Scientific Research*, Vol. 41, 1984, pp. 271–274.
- [10] Jacob, J., "A Goos-Haenchen Effect for Bragg Reflection", *AEÜ*, Vol. 39, 1985, pp. 69–72.
- [11] Suhara, T., and Nishihara, H., "Integrated Optical Components and Devices Using Periodic Structures", *IEEE Journal of Quantum Electronics*, Vol. 22, 1986, pp. 845–867.
- [12] März, R., and Nolting, H. P., "Spectral Properties of Asymmetrical Optical Directional Couplers with Periodic Structures", *Optical and Quantum Electronics*, Vol. 19, 1987, pp. 273–287.
- [13] Nishihara, H., Haruna, M., and Suhara, T., *Optical Integrated Circuits*, New York, McGraw-Hill, 1989.

## Chapter 8

### Planar Spectrographs

The history of spectrographs dates back to Fraunhofer, who used arrays of parallel wires as early diffraction gratings. At the end of the nineteenth century, Rowland developed the first concave reflection gratings which provided both diffraction and imaging within a single device. Incidentally, he constructed ruling machines for their fabrication. Marginal mechanical imperfections of these machines yield aberrations of the grating. In particular, periodic perturbations result in "ghosts" of the grating, i.e., in satellite peaks within their spectra. Nevertheless, most gratings manufactured today represent replicas of ruled master devices. By holographic exposure, however, it is possible at the expense of reduced efficiency to fabricate gratings which do not exhibit any ghosts. Spectrographs and spectrometers are applied to the spectral analysis of electromagnetic fields for a huge wavelength range from the deep-UV region up to microwaves. In addition, spectrographs and spectrometers can be used to provide wavelength selection within tunable sources such as external-cavity lasers.

Planar spectrographs play the corresponding role in integrated optics. They act as multi-channel filters for wavelength division multiplexing in optical communication systems, as sensor elements in integrated optical spectrum analyzers, and as cavities in integrated optical external cavity lasers. The technologies of integrated optics allow, in contrast to those for classical spectrographs, for an arbitrary shape and chirp of the gratings or grating-like devices. This increased flexibility is counterbalanced by restrictions on geometry and chip area which are characteristic for integrated optics.

This chapter deals with the design of focusing planar spectrographs, i.e., of diffracting devices having a fixed point of incidence. It starts with their general theory of imaging and aberration. An analysis of efficiency, spectral resolution and free spectral range follows. We will apply the results to analyze the properties of reflection gratings with one or two stigmatic, i.e., aberration-free, points. The end of this chapter is formed by a discussion of phased arrays.

## 8.1 THEORY OF PLANAR SPECTROGRAPHS

In the first section of this chapter we will derive some universal properties of planar spectrographs,<sup>1</sup> i.e., features which are common to reflection and transmission gratings operating inside slab waveguides and to phased arrays. We will start with an examination of focusing spectrographs. A brief discussion of their efficiency, spectral resolution and free spectral range will then follow.

### 8.1.1 Light-Path Function

The light-path function represents the crucial tool for the analysis of focusing gratings and grating-like devices. Before we start the discussion of this function, we will make a few comments on the underlying physical concept. We know from our discussion of the Fraunhofer diffraction integral (see Section 4.3) that the phase fronts of a finite beam propagating through the homogeneous space represent a family of equidistant spheres in the far-field. A finite beam propagating through a homogeneous slab waveguide will exhibit a family of circles as phase fronts. All these phase fronts are concentric, and their center is located at the waist of the beam. The phase difference between the beam waist and an arbitrary point in the far-field is given by

$$\Delta\varphi = 2\pi \frac{r}{\lambda}$$

where  $r$  designates the spacing between both points. Here and for the rest of this chapter, the parameter

$$\lambda = \frac{2\pi}{q} \quad (8.1)$$

represents the wavelength of a beam inside the material, i.e., the wavelength of an eigenmode propagating through a slab waveguide with an effective propagation constant  $q$ .

Obviously, the phase fronts of both incoming and outgoing beams<sup>2</sup> exhibit a circular shape as long as a grating is located in the far-field of the corresponding beams. The grating can then be analyzed by tracking the phase differences for a bundle of rays propagating from the point of incidence  $I$ , i.e., from the waist of the incoming beam, via the grating line  $z_G(y)$  to an observation point  $D$ . The technique described in the following can be regarded as a generalization of Huygen's principle for tackling focusing gratings.

For the following discussion let us consider a planar spectrograph which is illuminated by a finite beam emitted at the point of incidence  $I$ . Figure 8.1 shows

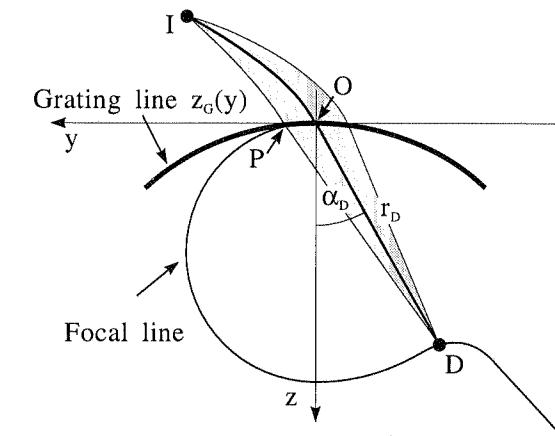


Figure 8.1. Planar spectrographs: coordinates

the output side of such a spectrograph. The grating line  $z_G(y)$  describes the position of the secondary wavelets according to Huygens's principle.<sup>3</sup> For illustration, Figure 8.1 shows also the focal line of the grating which will be calculated in the following. It describes the positions of the diffracted images.

The origin of the coordinate system used for our further calculations is located at the center  $O$  of the grating line (see Figure 8.1) where the incoming beam, i.e., its central ray, intersects the grating line  $z_G(y)$ . The coordinate system is oriented such that the grating line is tangential to its  $y$ -axis, i.e.,  $z'_G(0) = 0$ . The light-path function

$$F(y) = \widetilde{IP} + \widetilde{PD} - (\widetilde{IO} + \widetilde{OD}) + m\lambda G(y) \quad (8.2)$$

describes the effective path difference between a ray propagating from the point of incidence  $I$  to the observation point  $D = (y_D, z_D)$  via an intermediate point  $P = (y, z_G(y))$  at the grating line and the central ray of the beam propagating via the center  $O = (0, 0)$  of the grating line. The first two terms on the right-hand Equation (8.2) account for the physical path difference between both rays. The tildes above the terms  $\widetilde{IP}$  and  $\widetilde{IO}$  indicate that the light-path on the input side may be curved. The third term,  $m\lambda G(y)$ , describes the contributions of the (almost) periodic structure of the grating to the effective path difference.

To derive an expression for the groove function  $G(y)$ , we consider the phase difference

$$\Delta\varphi_{ID} = 2\pi \frac{F(y)}{\lambda}$$

1. For the sake of simplicity, we will often designate all types of planar spectrographs as gratings.  
2. With respect to the discussion of phased arrays (see Section 8.3) we should note that the phase difference caused by a strip waveguide can also be calculated from its length.

3. The detailed shape of the grating "grooves", i.e., the distribution of the secondary wavelets close to the grating line, affects the efficiency of the grating (see Section 8.1.4) but not the phase relations of the mounting.

between two different rays at the observation point  $D$ . Obviously, a grating with a constant period  $\Lambda$  along the  $y$ -axis operated in the  $m$ th diffraction order will allow for an additional phase shift

$$\Delta\varphi_\Lambda = m Q y$$

where

$$Q = \frac{2\pi}{\Lambda} \quad (8.3)$$

stands for its grating vector. Thus, the groove function for a grating with a constant period is given by  $G(y) = y/\Lambda$ .

By generalizing this expression to chirped gratings exhibiting a slowly varying grating vector  $Q(y)$  we see that the groove function occurring in Equation (8.2) is given by<sup>4</sup>

$$G(y) = \frac{1}{2\pi} \int_0^y d\tilde{y} Q(\tilde{y}). \quad (8.4)$$

For the further analysis we assume that only a “small” area around the point of intersection  $O$  is illuminated, i.e., that the functions describing the operation of the grating do not vary significantly over this region. We can then expand the light-path function<sup>5</sup>

$$F(y) = \sum_{\nu=1}^{\infty} \frac{1}{\nu!} F^{(\nu)}(0) y^\nu \quad (8.5)$$

and the groove function

$$G(y) = \frac{1}{2\pi} \sum_{\nu=0}^{\infty} \frac{1}{(\nu+1)!} Q^{(\nu)}(0) y^{\nu+1} \quad (8.6)$$

into rapidly converging Taylor series.

In the following we will derive expressions for the expansion coefficients  $F^{(\nu)}(0)$  of the light-path function. By using Figure 8.1 we see that the physical path difference

4. The parameter  $Q(y)$  in Equation (8.4) in fact represents the local grating vector  $Q(y) = 2\pi/\Lambda(y)$  of the chirped grating since

$$1 \stackrel{!}{=} G(y + \Lambda(y)/2) - G(y - \Lambda(y)/2) = \frac{1}{2\pi} \int_{y-\Lambda(y)/2}^{y+\Lambda(y)/2} d\tilde{y} Q(\tilde{y}) \approx \frac{1}{2\pi} Q(y) \Lambda(y).$$

5. To avoid clumsy notation, we will abbreviate the derivatives here and in the following by

$$\left. \frac{d^\nu f}{dy^\nu} \right|_{y=0} = f^{(\nu)}(0).$$

between the two rays interfering on the output side of the grating is given by

$$\begin{aligned} F_D(y) &= \overline{PD} - \overline{OD} \\ &= \sqrt{(y - y_D)^2 + (z_G(y) - z_D)^2} - r_D \end{aligned} \quad (8.7)$$

where

$$r_D = \sqrt{y_D^2 + z_D^2} \quad (8.8)$$

designates the spacing between the observation point and the origin of our coordinate system. By differentiating Equation (8.7) we arrive at

$$\begin{aligned} F'_D(0) &= \sin \alpha_D = -\frac{y_D}{r_D} \\ F''_D(0) &= T_D = \frac{\cos^2 \alpha_D}{r_D} - z_G''(0) \cos \alpha_D \\ F'''_D(0) &= U_D = -3 \frac{T_D \sin \alpha_D}{r_D} - z_G'''(0) \cos \alpha_D \\ F^{(IV)}_D(0) &= V_D = 3 \frac{z_G''(0)^2}{r_D} - 3 \frac{T_D^2}{r_D} - 4 \frac{U_D \sin \alpha_D}{r_D} - z_G^{(IV)}(0) \cos \alpha_D. \end{aligned} \quad (8.9)$$

The calculations leading to Equations (8.9) are elementary but somewhat lengthy.

We will designate the function describing the physical path difference on the input side of the grating by

$$F_I(y) = \overline{IP} - \overline{IO}. \quad (8.10)$$

This function  $F_I(y)$  will describe an arbitrary phase portrait along the grating line generated by purely geometrical means. We will avoid any further assumptions on this function here in order to keep the theory of imaging completely general. In fact, we will apply the results later on to various types of planar spectrographs including reflection gratings and even phased arrays.

The coefficients of the Taylor expansion (8.5) are then given by

$$\begin{aligned} F'(0) &= F'_I(0) + \sin \alpha_D + m \lambda G'(0) \\ F''(0) &= F''_I(0) + T_D + m \lambda G''(0) \\ F'''(0) &= F'''_I(0) + U_D + m \lambda G'''(0) \\ F^{(IV)}(0) &= F^{(IV)}_I(0) + V_D + m \lambda G^{(IV)}(0). \end{aligned} \quad (8.11)$$

### 8.1.2 Imaging by Focusing Spectrographs

On the basis of the Taylor expansion of the light-path function we will now establish the theory of imaging by focusing gratings. Obviously, a *focusing grating forms a perfect, completely aberration-free, image of the incoming beam if the light-path function*

$$F(y) \equiv 0 \quad (8.12)$$

*vanishes along the whole grating line.* We will see in the following that this rigid condition can at best be satisfied for only a few points in the observation plane. Such aberration-free observation points are designated as *stigmatic points of the mounting*. If only the leading coefficients of the rapidly converging Taylor expansion (8.5) vanish images exhibiting small aberrations will occur. By determining the observation point in the slab waveguide, we can usually make only the first two expansion coefficients vanish.

Fermat's principle ( $F'(0) = 0$ ) yields the diffraction angle  $\alpha_D$  via the grating equation

$$\sin \alpha_D = -F'_I(0) - m\lambda G'(0). \quad (8.13)$$

The image, i.e., the waist of the diffracted beam, is located

$$r_D = \frac{\cos^2 \alpha_D}{-m\lambda G''(0) - F''_I(0) + z''_G(0) \cos \alpha_D} \quad (8.14)$$

away from the center  $O$  of the grating line. The focal line<sup>6</sup> of the grating is then given by

$$\mathbf{r}_{\text{image}} = r_D \left( \begin{array}{c} -\sin \alpha_D \\ \cos \alpha_D \end{array} \right). \quad (8.15)$$

We should note that the grating will not always form a real image ( $|\alpha_D| < \pi/2$  and  $r_D > 0$ ). The discussion of the imaging condition requires an analysis of the inequality

$$G'(0) \frac{\cos^2 \alpha_D}{r_D} = G''(0) \sin \alpha_D + y''(0)G'(0) \cos \alpha_D + F'_I(0)G''(0) - F''_I(0)G'(0) > 0$$

which results from inserting Equation (8.13) into (8.14). We see that most gratings will satisfy the additional condition

$$y''(0)G'(0) \cos \alpha_D + F'_I(0)G''(0) - F''_I(0)G'(0) > 0$$

in order to provide a real image at the grating line. Figure 8.2 shows the focal line of two appropriately designed spectrographs as illustrative examples. All parameters of these devices are identical, only the first derivatives  $F'_I(0)$  of the light-path difference on the input side have opposite signs. The two different scenarios shown by Figures 8.2a and 8.2b become intuitively clear if we consider that, for reflection

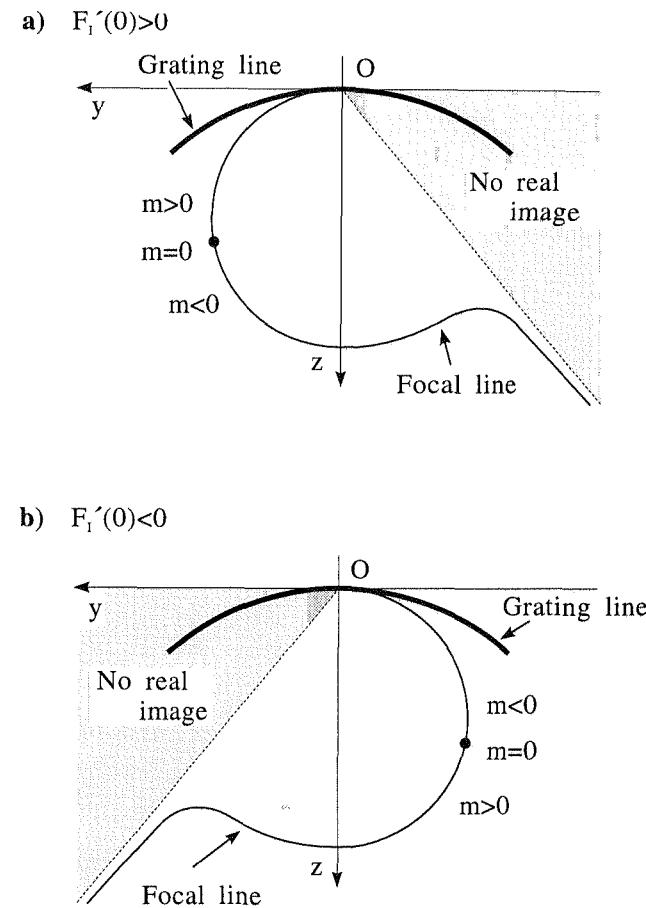


Figure 8.2. Focal curve for planar spectrographs with  $F'_I(0) > 0$  (a) and  $F'_I(0) < 0$  (b). The parameter  $m$  designates the diffraction order of the spectrograph.

gratings, the light-path difference on the input side of the grating depends on the angle of incidence, i.e.,  $F'_I(0) = \sin \alpha_I$ .

The appropriate choice of waveguides on the input and output sides must be based on the magnification of the optical system. Since the input side of the optical system is not yet defined, we cannot derive expressions for the spot magnification. However, the ratio of spot magnifications  $m_D$  and  $m_E$  for two arbitrarily chosen observation points  $D$  and  $E$  is given by

$$\frac{m_D}{m_E} = \frac{r_D \cos \alpha_E}{r_E \cos \alpha_D}. \quad (8.16)$$

6. Strictly speaking, the image curve of the grating.

The contours of constant spot magnification are formed by a family of circles which are tangential to the grating line at its center  $O$ . For a proof we consider Figure 8.3 and obtain

$$\frac{\sin d\theta_D}{dy} = \frac{\cos(\alpha_D - d\theta_D)}{r_D}.$$

For the paraxial limit ( $dy \rightarrow 0$ ) the ratio of angular magnifications  $m_{\theta,D}$  and  $m_{\theta,E}$  at the observation points  $D$  and  $E$  is then given by

$$\frac{m_{\theta,D}}{m_{\theta,E}} = \frac{d\theta_D}{d\theta_E} = \frac{\cos \alpha_D}{r_D} \frac{r_E}{\cos \alpha_E}.$$

Equation (8.16) follows directly if we consider (see Section 5.3.2) that the spot magnification represents the inverse of the angular magnification.

Further expansion coefficients can be made to vanish only for special mountings, i.e., by choosing the grating line  $z_G(y)$ , the groove function  $G(y)$  and the geometry of the fan-in region via  $F_I(y)$ . The expansion coefficients  $F^{(\nu)}(0)$  with  $\nu > 2$  represent the aberrations of the grating. The leading term  $F'''(0)$  designates the coma of the grating, and the following contribution  $F^{(IV)}(0)$  describes its spherical aberration. Astigmatism representing the most significant aberration for three-dimensional spectrographs cannot occur in planar devices. In order to obtain the maximum spectral resolution, designs usually aim to minimize the aberrations of the grating. In order to obtain the relevant magnitudes for the aberrations of a mounting, we consider the path difference between the central ray and the  $1/e$  ray of the outgoing beam (emitted at  $y_{1/e}$ ). A focusing spectrograph (and in fact an arbitrary optical system) will form a good image of the input beam if the path difference between both beams

$$F(y_{1/e}) < \frac{\lambda}{10}$$

remains smaller than a tenth of the wavelength (inside the slab waveguide), i.e.,  $F'''(0)y_{1/e}^3/6 < \lambda/10$  for the coma and  $F^{(IV)}(0)y_{1/e}^4/24 < \lambda/10$  for the spherical aberration. The following (universal) theorem shows the limits of aberration correction over an extended spectral range.

*It is impossible to design a planar spectrograph which exhibits neither a coma nor a spherical aberration ( $F'''(0) = 0$  and  $F^{(IV)}(0) = 0$ ) over an extended spectral range. Only Rowland-type mountings have no coma ( $F'''(0) = 0$ ).*

Before we start the proof of this theorem we will briefly examine the class of Rowland-type mountings emphasized by it. A Rowland mounting represents an unchirped grating

$$G^{(\nu)}(0) = 0, \nu > 1$$

mounted on a semicircle

$$z_G(y) = R - \sqrt{R^2 - y^2}$$

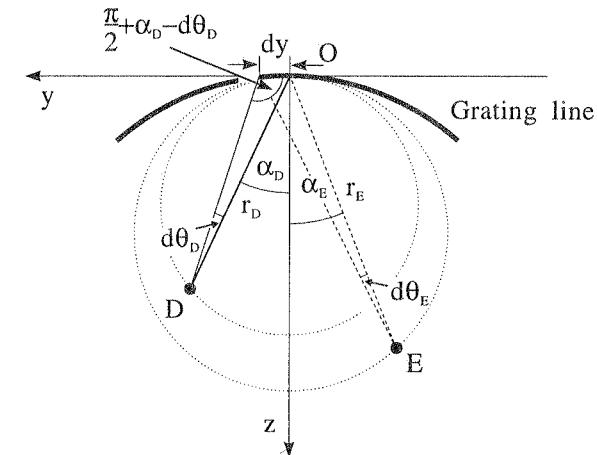


Figure 8.3. Angular magnification of a planar spectrographs.

with a radius  $R$ . The point of incidence  $I$  is chosen such that the phase portrait along the grating line becomes linear, i.e.,

$$F_I^{(\nu)}(0) = 0, \nu > 1.$$

Rowland-type mountings satisfy these conditions up to the third orders of the Taylor expansions of the functions  $G(y)$ ,  $z_G(y)$  and  $F_I(y)$ . By using Equation (8.14) we see that *the focal line of a Rowland-type mounting is given by*

$$r_D = R \cos \alpha_D \quad (8.17)$$

*the Rowland circle. This circle has a radius  $r = R/2$ . It is tangential to the grating line at the center  $O$  of the grating line. It has a radius  $r = R/2$  (see Figure 8.4). Rowland-type mountings exhibit a constant spot magnification (for any wavelength). At a few points on the Rowland circle, the spherical aberration will vanish if the condition*

$$F_I^{(IV)}(0) = -V_D \quad (8.18)$$

is satisfied. For a mounting with a linear phase portrait ( $F_I(y) \propto y$ ) and a circular grating line ( $z_G^{(IV)}(0) = 3/R^3$ ) the spherical aberration will vanish at the vertex of the Rowland circle ( $\alpha_V = 0$ ). By using the Equations (8.11), (8.2) and  $\overline{PV} - \overline{OV} \equiv 0$  we see that this point is in fact stigmatic.

For the interested reader we will now proof the theorem on aberration correction. We will first derive conditions for mountings which have no coma for any wavelength. By inserting the Equations (8.13) and (8.14) into the equation defining the coma  $F'''(0) = 0$  (see Equation (8.5)) we obtain

$$\sum_{i=0}^6 A_i(m\lambda)^i = \left[ \sum_{i=0}^3 S_i(m\lambda)^i \right]^2 - \left[ \sum_{i=0}^2 M_i(m\lambda)^i \right]^2 \cos^2 \alpha_D = 0 \quad (8.19)$$

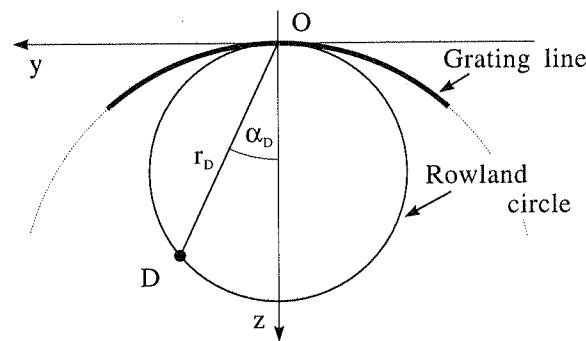


Figure 8.4. Rowland mounting.

with the abbreviations

$$\begin{aligned} M_0 &= 3F'_I(0)F''_I(0)z''_G(0) + [1 - F'_I(0)^2]z'''_G(0) \\ M_1 &= 3[F'_I(0)G''(0) + F''_I(0)G'(0)]z''_G(0) - 2F'_I(0)G'(0)z'''_G(0) \\ M_2 &= 3G'(0)G''(0)z''_G(0) - G'(0)^2z'''_G(0) \end{aligned}$$

and

$$\begin{aligned} S_0 &= 3F'_I(0)F''_I(0) + [1 - F'_I(0)^2]F'''_I(0) \\ S_1 &= [3F''_I(0)^2 - 2F'_I(0)F'''_I(0)]G'(0) + 6F'_I(0)F''_I(0)G''(0) + [1 - F'_I(0)^2]G'''(0) \\ S_2 &= -F'''_I(0)G'(0)^2 + 6F''_I(0)G'(0)G''(0) + 3F'_I(0)G''(0)^2 - 2F'_I(0)G'(0)G'''(0) \\ S_3 &= 3G'(0)G''(0)^2 - G'(0)^2G'''(0). \end{aligned}$$

The desired class of mountings must satisfy this equation for every power of the wavelength  $\lambda$ . The equation  $A_6 = 0$  leads to  $S_3^2 = -M_2^2G'(0)^2$ . If a grating exists ( $G'(0) \neq 0$ ) we can conclude that  $S_3 = 0$  and  $M_2 = 0$  must be satisfied. Successive solution of the equations  $A_i = 0$  under repeated application of the above argument yields the equations

$$\begin{aligned} S_i &= 0 \\ M_i &= 0. \end{aligned} \tag{8.20}$$

For  $F'_I(0) \neq 0$ , the equations  $S_3 = 0$ ,  $M_2 = 0$  and  $M_1 = 0$  can be formulated as

$$\begin{aligned} G'''(0) &= 3\frac{G''(0)^2}{G'(0)} \\ \frac{z'''_G(0)}{z''_G(0)} &= \frac{3G''(0)}{G'(0)} = \frac{3F''_I(0)}{F'_I(0)}. \end{aligned} \tag{8.21}$$

Inserting these results into  $M_0 = 0$  we obtain  $F''_I(0) = 0$ , and hence  $z'''_G(0) = 0$ ,  $G''(0) = 0$  and  $G'''(0) = 0$ . By solving the equation  $S_2 = 0$  we now find  $F'''_I(0) = 0$ . Inserting all results into Equations (8.9) finally yields  $T_D = 0$  and  $U_D = 0$ . The examination of the equations for the alternate case  $F'_I(0) = 0$  results in the same conditions.

By using Equation (8.14) we see that the observation point  $D$  will be found

$$r_D = \frac{\cos \alpha_D}{z''_G(0)}$$

away from the origin of the coordinate system. We can directly conclude that only concave mountings ( $z''_G(0) > 0$ ) represent appropriate candidates for mountings with low aberrations. A comparison of the conditions in fact shows that only Rowland-type mountings do not exhibit coma over an extended spectral range.

Next, we try to find Rowland-type mountings which have no spherical aberration for any wavelength  $\lambda$ . Inserting the preceding results into the definition for  $V_D$  (see Equation (8.9)), we obtain

$$V_D = 3 \frac{z''_G(0)^3}{\cos \alpha_D} - z_G^{(IV)}(0) \cos \alpha_D. \tag{8.22}$$

The equation  $F^{(IV)}(0) = 0$  can then be written as

$$\begin{aligned} &[F_I^{(IV)}(0) + (m\lambda)G^{(IV)}(0)]^2 \cos^2 \alpha_D \\ &= [3z''_G(0)^3 - z_G^{(IV)}(0) \cos^2 \alpha_D]^2. \end{aligned} \tag{8.23}$$

Again, it must be fulfilled for every power of the wavelength  $\lambda$ . A comparison of the coefficients belonging to  $\lambda^4$  yields  $G^{(IV)}(0)^2 = -z_G^{(IV)}(0)^2G'(0)^2$ , i.e.,  $G^{(IV)}(0) = 0$  and  $z_G^{(IV)}(0) = 0$ . Equation (8.23) now becomes

$$F_I^{(IV)}(0) \cos^2 \alpha_D = 9z''_G(0)^6.$$

This equation can be satisfied over an extended spectral range if and only if  $z''_G(0) = 0$ , i.e., if the image is located at infinity. However, this represents a discrepancy with respect to the conditions derived earlier.

### 8.1.3 Stigmatization

The technologies of integrated optics allow, in contrast to those for classical spectrographs, for an arbitrary shape and chirp of the grating line. We will show here how to make a spectrograph stigmatic at one freely chosen point by choosing an appropriate chirp. The groove function  $G(y)$  of a spectrograph exhibiting a stigmatic point  $D_1 = (y_1, z_1)$  at a wavelength  $\lambda_1$  is given by

$$G(y) = \frac{-1}{m\lambda_1} [F_I(y) + \overline{PD_1} - \overline{OD_1}] \tag{8.24}$$

At the center  $O$  of the grating line its derivatives are

$$\begin{aligned} G'(0) &= \frac{-1}{m\lambda_1} [F'_I(0) + \sin \alpha_1] \\ G''(0) &= \frac{-1}{m\lambda_1} [F''_I(0) + T_1] \\ G'''(0) &= \frac{-1}{m\lambda_1} [F'''_I(0) + U_1] \\ G^{(IV)}(0) &= \frac{-1}{m\lambda_1} [F_I^{(IV)}(0) + V_1]. \end{aligned} \quad (8.25)$$

The abbreviations are defined by Equation (8.9). It is obvious from Equation (8.2) that the light-path function  $F(y)$  vanishes at the stigmatic point  $D = D_1$ .

In order to reduce the chirp of the stigmatic grating it is useful to start the design process from an unchirped mounting and to correct the aberrations at a point  $D_1$  on its focal line in a second design step.

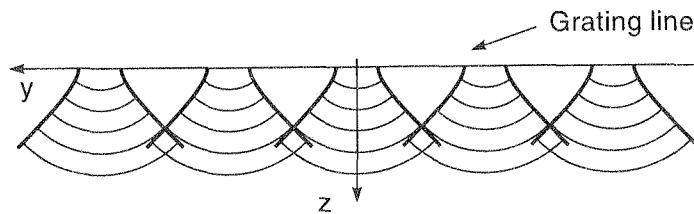


Figure 8.5. Diffraction from a straight grating.

#### 8.1.4 Efficiency

In the following we will derive a rough theory of efficiency for planar spectrographs. For this purpose we consider the optical field  $\Psi$  emitted from a straight grating line  $z_G(y)$  of an unchirped grating (see Figure 8.5). The (rapidly varying) optical far-field of the beam emitted from its central groove (located at  $y = 0$ ) is then given by<sup>7</sup>

$$\Psi_0(y, z) = A_0 \psi(q \sin \alpha_D) \frac{e^{iqr_D}}{r_D} \quad (8.26)$$

where  $\psi(q_y) = \mathcal{F}[\psi(y, z = 0)]$  designates the Fourier transform of the normalized optical near-field. The parameter  $A_0$  stands for the amplitude of the partial beam

and  $q$  for its propagation constant inside the underlying slab waveguide. The parameters  $\alpha_D$  and  $r_D$  describe the diffraction angle and the distance to the observation point  $D$ . The beams emitted from the other grooves of the grating exhibit the same near-field shifted by an integral multiple of the grating period, i.e., their centers are located at

$$y_n = n\Lambda.$$

By using the general formula

$$f(q) = \mathcal{F}[f(y)] = e^{iqa} \mathcal{F}[f(y - a)]$$

for the Fourier transform of a shifted function and by neglecting the shift of the circular waves  $e^{iqr_D}/r_D$  we obtain

$$\Psi_n(y, z) = A_n \exp(-in\Lambda \sin \alpha_D) \psi(q \sin \alpha_D) \frac{e^{iqr_D}}{r_D}. \quad (8.27)$$

For a grating with  $N$  grooves, the optical far-field is then described by the superposition of all partial beams

$$\Psi(y, z) = A_N(q \sin \alpha_D) \psi(q \sin \alpha_D) \frac{e^{iqr_D}}{r_D} \quad (8.28)$$

with

$$A_N(q \sin \alpha_D) = \sum_{n=-(N-1)/2}^{(N-1)/2} A_n \exp(-in\Lambda \sin \alpha_D). \quad (8.29)$$

It is assumed that all partial beams emitted from the grating line are in phase.

In the following, we will discuss the contributions on the right-hand side of Equation (8.28). Obviously,

$$A_N(q \sin \alpha_D \pm \frac{2m\pi}{\Lambda}) = A_N(q \sin \alpha_D) \quad (8.30)$$

represents a periodic function whose maximum values are found at

$$\sin \alpha_D = -\frac{m\lambda}{\Lambda}. \quad (8.31)$$

By using  $G'(0) = 1/\Lambda$  we see that Equation (8.31) represents the grating equation<sup>8</sup> for unchirped gratings illuminated with plane waves at normal incidence ( $F'_I(0) = 0$ ); i.e., the periods of the function  $A_N(q \sin \theta_D)$  represent the diffraction orders  $m$  of the grating. By recalling our discussion of the numerical Fourier transformation in Section 4.2.4 we see furthermore that the shape of each single diffraction peak is given approximately by

$$A_N(q \sin \alpha_D) \rightarrow A(q \sin \alpha_D \pm \frac{2m\pi}{\Lambda}) = \mathcal{F}[A(y)]$$

7. For a derivation of the Fraunhofer diffraction integral describing the optical field in the far-field limit see Section 4.3.

8. See Equation (8.13).

the Fourier transform of the envelope of the amplitudes  $A_m$  of the partial beams at the grating line.

The second contribution in Equation (8.28) – the Fourier transform of the near-field emitted from a single groove – acts as an envelope of the diffraction pattern. We can estimate the typical width of this envelope by considering a transmission grating consisting of an array of rectangular apertures. The envelope of the diffraction pattern will then be given by

$$\psi(q \sin \alpha_D) \propto \frac{\sin((q\Lambda_S \sin \alpha_D)/2)}{(q\Lambda_S \sin \alpha_D)/2}$$

a sinc-function ( $\text{sinc}(x) = \sin(x)/x$ ) whose first zero is found at

$$\sin \alpha_D^{(0)} = \frac{\lambda}{\Lambda_S},$$

i.e., a transmission grating with a typical on-off ratio of 1:1 ( $\Lambda_S = \Lambda/2$ ) will exhibit zero efficiency at the  $\pm 2$ nd diffraction order of the grating. In general, the envelope  $\psi$  will “allow” for one or a few diffraction orders. This means that the relative optical power diffracted into the other diffraction orders is rather small.

Figure 8.6 summarizes the results obtained so far: *The shape of a single diffraction peak represents the Fourier transform of the envelope of the optical field emitted from the grating line. The Fourier transform of the near-field of a single grating groove forms the envelope of the diffraction pattern.*

For the moment, these results apply only to unchirped gratings emitting an array of beams of equal phase from a straight grating line. Such simple gratings, however, exhibit high insertion losses since most of the optical power is launched into the (useless) 0th diffraction order and each half of the remaining optical power is launched into the  $m$ th and  $-m$ th diffraction orders. In addition, only the low diffraction orders providing low spectral resolution are available.

By using Equation (8.28) we see, however, that a tilt of the phase fronts of the partial beams<sup>9</sup> will result in a shift of the envelope  $\psi$  of the diffraction patterns. This technique is designated as “blazing” the grating. A constant phase shift from one partial beam to another, in contrast, will result in a shift of the diffraction pattern  $A_N$  itself. Although blazing can be analyzed on the basis of Equation (8.28) we prefer a less abstract approach in order to describe this technique. We have already seen that the envelope of the diffraction pattern is given by the Fourier transform of the near-field of a beam emitted from a single groove. Thus, the envelope can be rotated by rotating the grooves relative to the grating line. Furthermore, the central rays of the partial beams<sup>10</sup> propagating in the direction defined by the blaze angle  $\alpha_B$  will point to the maximum value of the envelope  $\psi$ . By choosing an appropriate blaze angle  $\alpha_B$  we can launch most of the optical power into a high diffraction

order of a grating. *For curved gratings, blazing becomes more complicated. The central rays emitted from all partial beams must intersect each other at a single point which can be regarded as the “blaze point” of the mounting.* This point must obviously be located on the focal line (see Figure 8.7).

Finally, we will make a few remarks on the electromagnetic theory of reflection gratings. Up to now, we have analyzed the properties of an array of beams propagating through a homogeneous half-space. For transmission gratings and phased arrays this approach will yield accurate results. For reflection gratings, however, we have to take the boundary conditions at the corrugated surface of the grating into

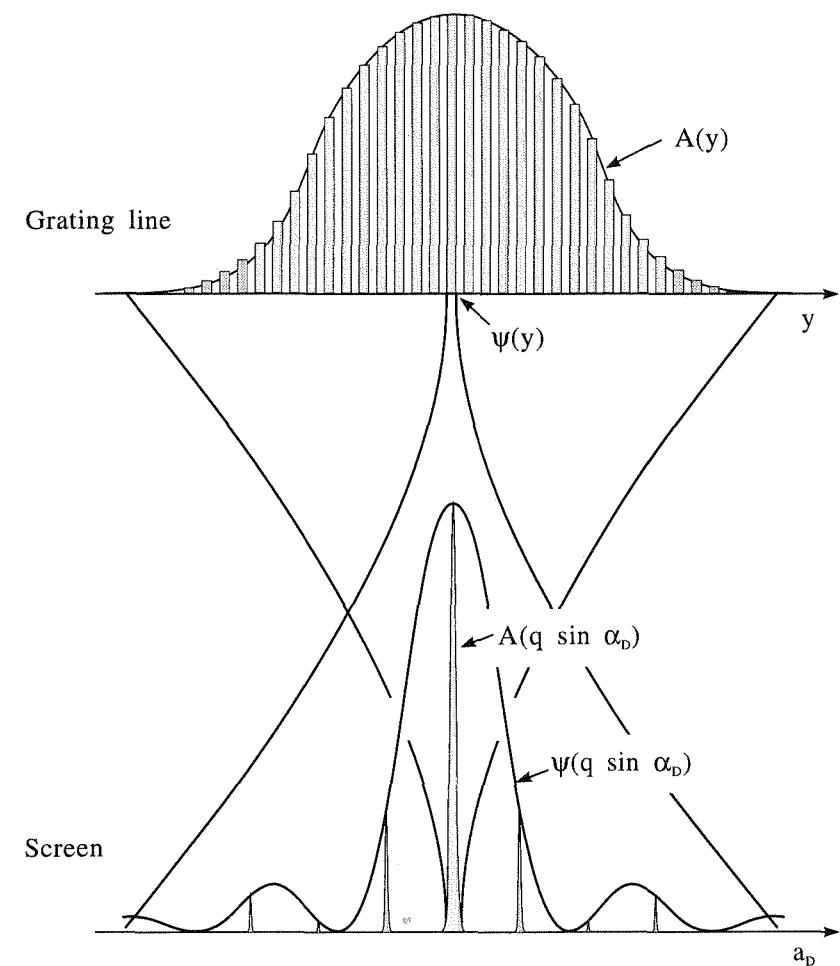


Figure 8.6. Diffraction efficiency of spectrographs.

9. This tilt is usually caused by a tilt of the grooves themselves.

10. We should note that the propagation direction of the central ray of a beam includes the influence of tilted near-fields.

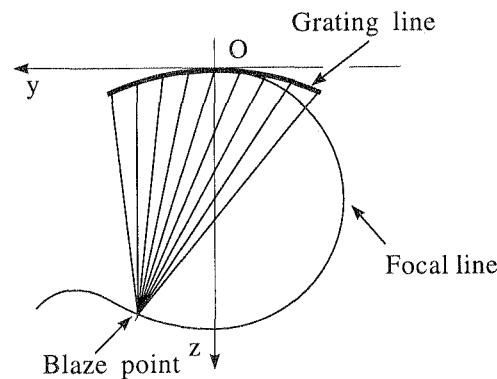


Figure 8.7. Blazing of focusing spectrographs.

account. Since the boundary conditions differ for the different vector components of the electric and magnetic fields the efficiency curves of the reflection gratings will be affected by the vectorial character of the incoming beam. In consequence, the envelope of the diffraction pattern of a reflection grating will always exhibit polarization-dependent behavior. The theory of plane reflection gratings for optical instruments (i.e., of gratings mounted on a plane surface) has been extensively examined [1]. The efficiency curves of such gratings exhibit a rich variety of physical effects. The Wood anomaly caused by “vanishing” diffraction orders represents the most prominent candidate for such effects. The grating coupler briefly discussed in the previous chapter constitutes an example of its technical use. Metal cladded gratings exhibit additional anomalies due to the excitation of surface waves (surface plasmons). Obviously, the same effects will also occur for reflection gratings of planar spectrographs. However, the electromagnetic theory of reflection gratings for planar spectrographs, i.e., of gratings which form the endface of a slab waveguide, has not been elaborated. The electromagnetic theory of curved gratings is equally poorly established. Experimental work, however, has shown that high-order reflection gratings with reasonable efficiencies for both TE and TM polarization can be realized.

### 8.1.5 Spectral Resolution and Free Spectral Range

We have seen in the previous paragraph that the shape of a single diffraction pattern is determined by the Fourier transform of the envelope of the optical field emitted from the grating line. We can directly conclude that the width of a diffraction peak will increase with decreasing width of the optical power distribution along the grating line; i.e., the spectral resolution of a spectrograph depends on the number of

illuminated grating grooves. By differentiating Equation (8.13) we can relate

$$\frac{d(\sin \alpha_D)}{d\lambda} = -\frac{m}{\Lambda} \quad (8.32)$$

the deviations of the diffraction angle  $\alpha_D$  and the wavelength  $\lambda$ .

For a rectangular power distribution along the grating line, i.e., for a grating with  $N$  grooves all emitting the same optical power, we obtain for the shape of a single diffraction peak

$$A(q d(\sin \alpha_D)) \propto \frac{\sin(N q \Lambda d(\sin \alpha_D)/2)}{N q \Lambda d(\sin \alpha_D)/2}. \quad (8.33)$$

According to Rayleigh's criterion, two diffraction peaks described by a sinc-function can be just resolved if the center of one is located at the first zero of the other. By using Equations (8.32) and (8.33) we obtain the spectral resolution

$$d\lambda_R = \frac{\lambda}{|m|N} \quad (8.34)$$

according to Rayleigh's criterion.

However, the focusing reflection grating and the phased array discussed in the following two sections are illuminated by finite beams emitted from a waveguide endface. Analogously to our treatment of mode conversion (see Chapter 5) we will use the Gaussian beam as a model describing the fundamental mode of a single-mode waveguide.<sup>11</sup> The grating line is then illuminated by the far-field of a Gaussian beam; i.e., the optical power distribution along the grating line will exhibit a Gaussian shape. In consequence, the shape of the peaks on the output side of the spectrograph

$$A(q d(\sin \alpha_D)) \propto \exp\left(-\left[\frac{N q \Lambda d(\sin \alpha_D)}{4}\right]^2\right) \quad (8.35)$$

will be also Gaussian provided the spectrograph itself is aberration-free. The integral number  $N$  now designates the number of grooves for which the optical power exceeds  $1/e^2$  of the maximum value. By using the derivative of the grating Equation (8.32) we obtain

$$d\lambda(\rho) = \frac{2\rho}{\pi} d\lambda_R \quad (8.36)$$

a new criterion for the spectral resolution. The free parameter  $\rho = d_s/r_0$  describes the spacing  $d_s$  between two diffraction peaks in terms of their field radius  $r_0$ .

In integrated optics, the choice of the relative spacing  $\rho$  of the diffraction peaks depends on the underlying application, essentially on the required crosstalk attenuation between adjacent wavelength channels. By applying Equation (5.27) which

11. This model tends to slightly overestimate the spectral resolution since Gaussian modes decay more rapidly than the eigenmodes of usual waveguides

describes the coupling loss of two transversely misaligned beams we obtain an equation

$$\gamma(\rho) = 10 \log(e) \rho^2 \quad (8.37)$$

for the free parameter  $\rho$ . By using Equations (8.34) and (8.36) we can relate this parameter to the spectral resolution  $d\lambda(\rho)$ . Figure 8.8 shows the graph corresponding to Equation (8.37). We should note that the final extinction ratio between signal and crosstalk will decrease if the crosstalk is caused by two signals on the left-hand and right-hand sides, the power levels of signal and crosstalk are not equalized or if the requirement specifications allow for reasonable wavelength tolerances.

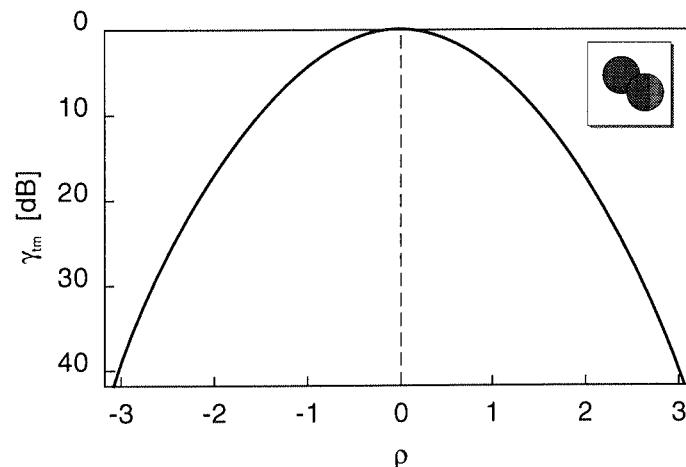


Figure 8.8. Crosstalk attenuation vs. relative transverse shift.

We have seen that the spectral resolution of a grating depends essentially on the product  $|m|N$  the diffraction order  $m$  and the number of illuminated grooves  $N$ . We will now examine how to choose the appropriate diffraction order. For this purpose, we consider a fixed diffraction angle  $\alpha_D$  which is chosen such that the  $m$ th-order diffraction peak is located close to this angle. The diffraction peaks of the adjacent diffraction orders will be found at  $\alpha_D$  if the conditions

$$-(m-1) \frac{\lambda + \Delta\lambda_{FSR}}{\Lambda} = \sin \alpha_D = -(m+1) \frac{\lambda - \Delta\lambda_{FSR}}{\Lambda}$$

are satisfied. The parameter  $\Delta\lambda_{FSR}$  is designated as the free spectral range between adjacent diffraction orders (at  $\lambda$ ). A spectrograph obviously will provide a free spectral range  $\Delta\lambda_{FSR}$  if its diffraction order  $m$  satisfies the condition

$$|m| \leq \left| \frac{\lambda}{\Delta\lambda_{FSR}} \right|; \quad (8.38)$$

i.e., the maximum diffraction order is determined by the free spectral range of the underlying application.

The period

$$\Lambda_m = |m|\Lambda_1$$

of an  $m$ th-order grating represents a multiple of the period of the equivalent 1st order grating. Thus, the fabrication of planar spectrographs can be simplified at the expense of a reduced spectral range. In fact, planar spectrographs often exhibit the maximum diffraction order allowed by inequality (8.38). Devices with  $|m| > 100$  have been realized. We should note that the size of spectrographs is not affected by the diffraction order since the extent of the illuminated area  $N|m|\Lambda$  at the grating line remains unchanged.

The number of accessible wavelength channels  $N_{ch}$  is given by the quotient of free spectral range and spectral resolution. For the Rayleigh criterion (8.34) the number of accessible wavelength channels

$$N_{ch} = N \quad (8.39)$$

is given by the number of grooves  $N$  of the grating or grating-like device. For the criterion (8.36), which is adapted to the requirements of integrated optics, we obtain

$$N_{ch}(\rho) = \frac{\pi}{2\rho} N, \quad (8.40)$$

i.e., usually a slightly smaller number of wavelength channels.

## 8.2 FOCUSING REFLECTION GRATINGS

A focusing reflection grating represents a folded optical system; i.e., both input and output beams propagate through a slab waveguide in front of the grating. We can draw a number of conclusions from that property.

- The physical path difference on the input side is given by

$$F_I(y) = \overline{IP} - \overline{IO} \quad (8.41)$$

where  $I$  represents the point of incidence and  $P$  an arbitrary point on the grating line. Its derivatives at the center  $O$  of the grating line are given by  $F'_I(0) = \sin \alpha_I$ ,  $F''_I(0) = T_I$ ,  $F'''_I(0) = U_I$  and  $F^{(IV)}_I(0) = V_I$ . Explicit expressions for the parameters  $T_I$ ,  $U_I$  and  $V_I$  are given by Equation (8.9).

- The spot magnification factor of a focusing reflection grating is given by

$$m_o = \frac{r_D \cos \alpha_I}{r_I \cos \alpha_D} \quad (8.42)$$

where  $\alpha_D$  designates the diffraction angle and  $r_D$  the distance between the center of the grating line and the focal point  $D$ . The parameters  $\alpha_I$  and  $r_I$  describe the corresponding parameters of the input beam.

- The point of incidence  $I$  of a Rowland-type mounting must be located on the Rowland circle.
- Reflection gratings are usually smaller than equivalent transmission gratings or phased arrays.

Focusing reflection gratings are blazed by orienting the grating grooves such that all reflected rays intersect at the blaze point.

The diffraction order of a reflection grating also represents a free parameter. Although a high diffraction order appears attractive for technological reasons, it might be necessary to adapt it in order to avoid the efficiency problems described in Section 8.1.4. We should note in addition that the efficiency of reflection gratings is drastically affected by the tilt of their grooves with respect to the vertical, since any tilt of a diffracted beam results in a coupling loss to the slab waveguide.<sup>12</sup>

### 8.2.1 Stigmatic Mountings

Reflection gratings for planar spectrographs are, in contrast to their equivalents for three-dimensional spectrographs, fabricated by photolithography and dry etching. These fabrication technologies allow for gratings of arbitrary shape and chirp. We will show in the following how these degrees of freedom can be used in order to design mountings with extremely low aberrations by choosing one or two stigmatic, i.e., aberration-free, points along the focal line.

In the following, we will examine a method of constructing mountings having a fixed point of incidence  $I$  and two freely chosen stigmatic points  $D_1$  and  $D_2$  on the focal line. In contrast to the usual procedure, the construction of stigmatic mountings does not start with a given grating line, but with a choice of the point of incidence  $I$  and two stigmatic points  $D_1$  and  $D_2$ . The grating line is then found at the intersection points of two families ( $i = 1, 2$ ) of confocal ellipses with the parameter  $t$ . These two families of ellipses have one common focus at the point of incidence  $I$  and two more focal points at the stigmatic points  $D_1$  and  $D_2$  (see Figure 8.9).

The distance between the focal points  $2c_i$  of the ellipses, their long axes  $2a_i(t)$  and short axes  $2b_i(t)$  are given by

$$\begin{aligned} 2c_i &= \overline{ID}_i \\ 2a_i(t) &= (r_I + r_i) + \lambda_i t \\ b_i(t) &= \sqrt{a_i(t)^2 - c_i^2}. \end{aligned} \quad (8.43)$$

The following equations describe the grating line of a mounting with two stigmatic points on the basis of a polar coordinate system  $(r, \varphi)$  whose origin is located at

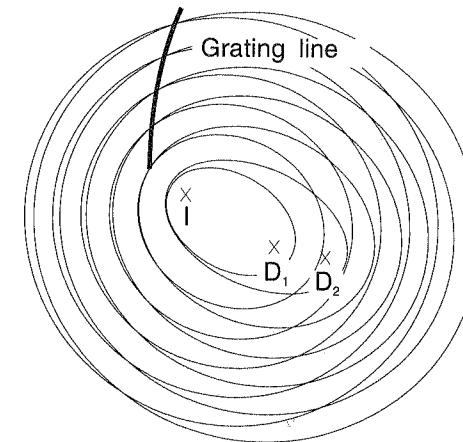


Figure 8.9. Construction of reflection gratings with two stigmatic points.

the point of incidence  $I$ :

$$\begin{aligned} \cos \varphi(t) &= \frac{SQ \pm R\sqrt{Q^2 + R^2 - S^2}}{Q^2 + R^2} \\ \sin \varphi(t) &= \frac{SR \mp Q\sqrt{Q^2 + R^2 - S^2}}{Q^2 + R^2} \\ r(\varphi(t), t) &= \frac{b_1(t)^2}{a_1(t) - c_1 \cos \varphi(t)}. \end{aligned} \quad (8.44)$$

The reference axis of this coordinate system is defined by the vector  $\overrightarrow{ID}_1$  (see Figure 8.10). The parameter  $\varphi_2 = \angle(\overrightarrow{ID}_1, \overrightarrow{ID}_2)$  designates the angle between the principal axes of the two families of ellipses. The other parameters are given by

$$\begin{aligned} Q &= b_1(t)^2 c_2 \cos \varphi_2 - b_2(t)^2 c_1 \\ R &= b_1(t)^2 c_2 \sin \varphi_2 \\ S &= b_1(t)^2 a_2(t) - b_2(t)^2 a_1(t). \end{aligned}$$

To analyze the performance of a reflection grating with two stigmatic points, i.e., to derive the focal line of the mounting and the most significant aberrations along its track, we must calculate the derivatives of the grating line  $z_G^{(\nu)}(0)$  at its center. Let us now recall that all derivatives of the light-path function  $F^{(\nu)}(0)$  must vanish for both stigmatic points. Inserting this condition into the equation defining

12. See Section 5.2 for an estimation of the coupling losses for tilted beams.

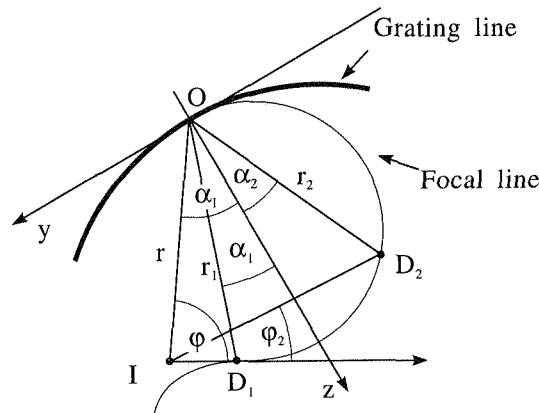


Figure 8.10. Reflection gratings with two stigmatic points: coordinates.

the first derivative  $F'(0) = 0$  yields

$$\sin \alpha_I = -\frac{\sigma(\lambda_1 \cos^2(\chi_2/2) - \lambda_2 \cos^2(\chi_1/2)) (\lambda_1 \sin \chi_2 - \lambda_2 \sin \chi_1)}{\lambda_1 \sqrt{(\lambda_1 \cos^2(\chi_2/2) - \lambda_2 \cos^2(\chi_1/2))^2 + (\lambda_1 \sin \chi_2 - \lambda_2 \sin \chi_1)^2}}, \quad (8.45)$$

where  $\chi_i = \angle(\overrightarrow{OI}, \overrightarrow{OD_i})$  designate the angles between the incoming and the diffracted rays for both stigmatic points. We should note that these angles are independent of the grating layout. Here  $\sigma(s)$  represents the signum function, i.e.,  $\sigma(s \geq 0) = 1$  and  $\sigma(s < 0) = -1$ . By using the higher order derivatives  $F^{(v)}(0)$  for the stigmatic wavelengths, we obtain the following expressions for the higher order derivatives of the grating line

$$\begin{aligned} z_G''(0) &= \frac{1}{\Gamma} \left\{ \frac{\cos^2 \alpha_I}{r_I} + \frac{\cos^2 \alpha_2}{r_2} - \frac{\lambda_2}{\lambda_1} \left( \frac{\cos^2 \alpha_I}{r_I} + \frac{\cos^2 \alpha_1}{r_1} \right) \right\} \\ z_G'''(0) &= \frac{-3}{\Gamma} \left\{ \frac{T_I \sin \alpha_I}{r_I} + \frac{T_2 \sin \alpha_2}{r_2} - \frac{\lambda_2}{\lambda_1} \left( \frac{T_I \sin \alpha_I}{r_I} + \frac{T_1 \sin \alpha_1}{r_1} \right) \right\} \\ z_G^{(IV)}(0) &= \frac{1}{\Gamma} \left\{ V_I + V_2 - \frac{\lambda_2}{\lambda_1} (V_I + V_1) \right\}. \end{aligned} \quad (8.46)$$

with

$$\Gamma = \cos \alpha_I + \cos \alpha_2 - \frac{\lambda_2}{\lambda_1} (\cos \alpha_I + \cos \alpha_1).$$

The other abbreviations were defined in Equation (8.9). Numerical studies will sometimes show a “low aberration point” beside the two stigmatic points forming the basis of the mounting. Unfortunately, no systematic procedure is known which would allow such points to be created.

In general, mountings exhibiting more than two stigmatic points would be attractive for correcting the aberrations of the grating over a more extended spectral range. In fact, mountings with three stigmatic points have been reported in the literature [2]. However, as a consequence of their holographic construction, the point of incidence  $I$  already represents one of the stigmatic points. A brief glance at the construction described here shows that mountings exhibiting three stigmatic points can be realized only at the expense of a substantially reduced flexibility. It is unknown whether mountings with more than three stigmatic points exist.

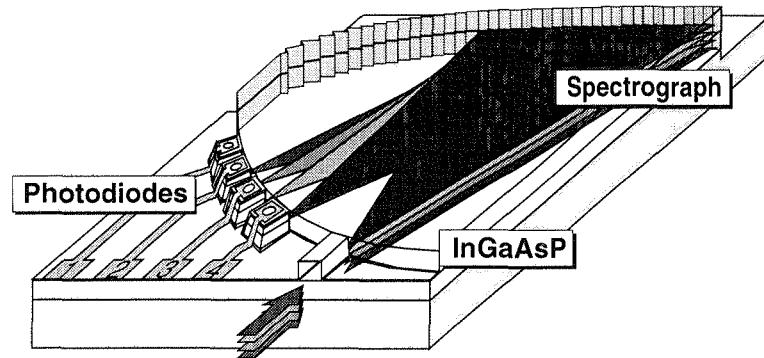


Figure 8.11. Rowland spectrograph in InGaAsP/InP.

## 8.2.2 Improved Rowland Mountings

As a first numerical example we will examine a selection of spectrographs in the InGaAsP/InP material system which are based on Rowland-type and very similar mountings. Such spectrographs might be regarded as a subsystem of a more complex OEIC comprising photodiodes and/or other functional units (see Figure 8.11). In view of the fabrication technologies of integrated optics, the detailed shape of the focal line is of minor importance. For our numerical example we will consider the following device specifications:

Point of incidence:	$I = (0, 0)$
Grating line:	$O = (0, 2.5 \text{ mm}), R = 1/z_G''(0) = 3.6 \text{ mm}$
Channel range:	$\lambda = 1300 - 1600 \text{ nm}$
Effective refractive index:	$n = 3.22$
Angular dispersion:	$ d\vartheta/d\lambda  \approx 1.5 \cdot 10^{-3}/\text{nm}$

These specifications lead to a  $-1^{\text{st}}$  order grating with a period of  $\Lambda = 0.3 \mu\text{m}$ . In the following, we will compare three different mountings. The first candidate is a classical Rowland mounting with a constant period  $\Lambda$ , and the second one, a stigmatic Rowland mounting with a stigmatic wavelength  $\lambda_1 = 1580 \text{ nm}$ . The other parameters coincide with those of the classical Rowland mounting. The third mounting is based on two stigmatic points which are located at the focal points of the classical Rowland mounting for the wavelengths  $\lambda_1 = 1580 \text{ nm}$  and  $\lambda_2 = 1547 \text{ nm}$ . The points of incidence  $I$  and the center  $O$  of the grating line coincide with those of the other mountings.

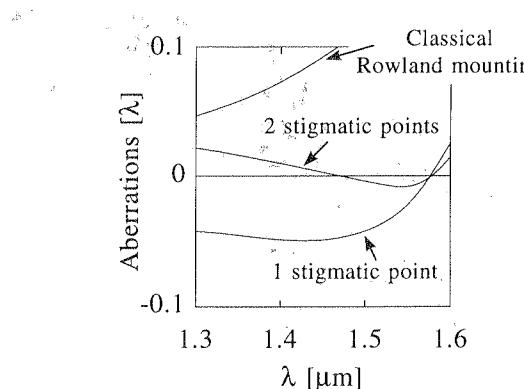


Figure 8.12. Spherical aberration for some mountings in InGaAsP/InP.

Figure 8.12 shows the spherical aberrations between the central ray and the  $1/e^2$  ray of a Gaussian beam having a field radius of  $1 \mu\text{m}$ . The aberrations are given as usual in units of the wavelength  $\lambda$  inside the slab waveguide. The classical Rowland mounting exhibits severe aberrations  $F > \lambda/2$  that prevents a practical use of the device in the spectral range of operation. The stigmatic Rowland mounting has only half as many aberrations as the classical Rowland mounting. The third mounting, based on two stigmatic points, exhibits a further significant reduction of the aberrations. This numerical example shows clearly that even standard mountings such as the classical Rowland mounting can be substantially improved by using stigmatic points.

### 8.2.3 Flat-Field Spectrographs

By defining the position and wavelength of mountings with the two stigmatic points, we can tune the most important properties of the planar spectrograph, i.e., its spectral resolution, dispersion and the position and shape of its focal line. In addition,

the aberrations of such devices will usually be small over the spectral range of operation. Thus, they represent interesting candidates for non-standard layouts which usually tend to exhibit high aberrations.

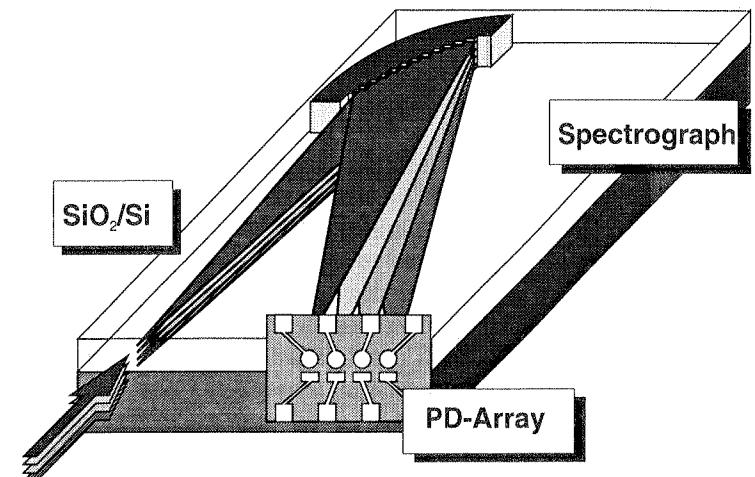


Figure 8.13. Flat-field spectrograph.

As an illustrative example we will now examine a planar spectrograph (see Figure 8.13) which does not require any strip waveguides. The point of incidence  $I$  and the focal line (at operating wavelengths) must then be located along the chip endface where the diffracted light can be coupled directly into an array of photodiodes or fibers. Such devices are designated as *flat-field spectrographs*. For our numerical example we will consider a device in the material system  $\text{SiO}_2/\text{Si}$  with the following specifications:

Point of incidence:	$I = (0, 0)$
Observation point at 1530 nm:	$D \approx (2 \text{ mm}, 0)$
Focal line:	$\vec{r}_F \approx (s, 0)$
Grating line:	$O = (0, 20 \text{ mm})$
Channel range:	$\lambda = 1500 - 1560 \text{ nm}$
Effective refractive index:	$n = 1.446$
Linear dispersion:	$ ds/d\lambda  \approx 50 \mu\text{m}/\text{nm}$

We should note that it is impossible to design a Rowland-type mounting that meets these specifications. We will now compare two appropriate designs based on stig-

matic gratings. The first mounting is constructed by choosing two stigmatic points at  $D_1 = (1 \text{ mm}, 0)$  and  $D_2 = (3 \text{ mm}, 0)$ . The corresponding wavelengths are given by  $\lambda_1 = 1510 \text{ nm}$  and  $\lambda_2 = 1550 \text{ nm}$ . The second mounting represents a stigmatic grating based on a circular grating line. It has one stigmatic point located at  $D_1 = (2 \text{ mm}, 0)$ . The corresponding wavelength  $\lambda_1 = 1530 \text{ nm}$  represents the center of the channel range. The orientation and the radius of curvature of its grating line coincide with the values obtained for the mounting I with two stigmatic points. Thus the second layout might be regarded as the result of a (numerical) optimization process.

Figure 8.14a shows the defocusing of both mountings, i.e., the distance between real and desired focus measured along the central ray of the beam. The defocusing curves of both mountings are shifted slightly relative to each other. This shift is caused by the adjustment of the focal line to the two stigmatic points  $D_1$  and  $D_2$  of the first mounting and to the single stigmatic point  $D$  of the second one. Clearly, it can be easily corrected. Figure 8.14b compares the most significant aberrations, i.e., coma (solid lines) and spherical aberration (dashed lines), between the central ray and the  $1/e^2$  ray of a Gaussian beam which matches to the near-field of a standard single-mode fiber. Both coma and spherical aberration are significantly higher for the mounting based on one stigmatic point than for the first mounting having two stigmatic points.

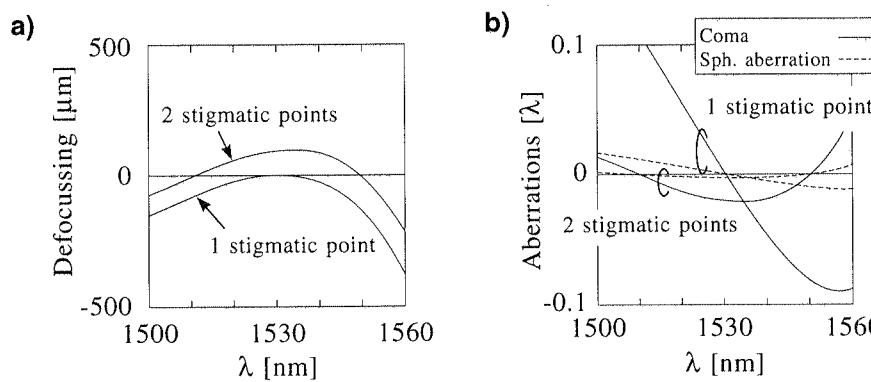


Figure 8.14. Defocusing (a) and aberrations (b) for a flat-field spectrograph.

This example demonstrates that non-standard layouts can also be used to design spectrographs with low aberrations. We should note, however, that the significantly smaller aberrations  $F < \lambda/10$  (in comparison to the Rowland mountings treated before) are a consequence of the less extended spectral range of operation and not of the mountings themselves.

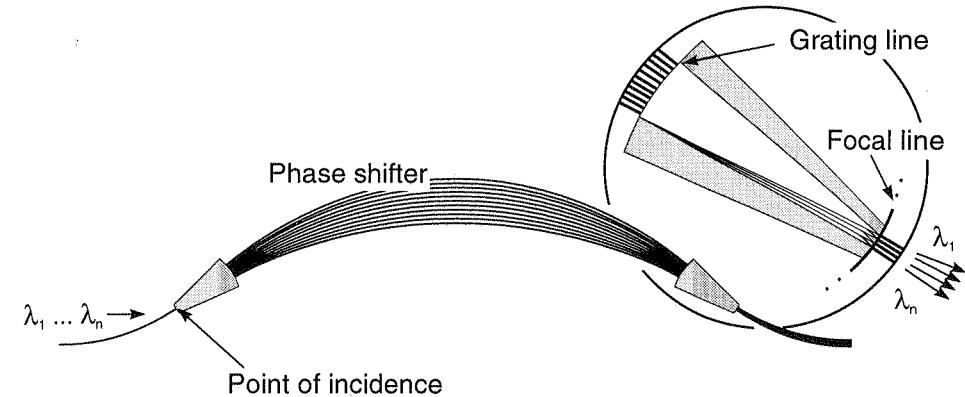


Figure 8.15. Phased array: principle of operation.

### 8.3 PHASED ARRAYS

From an optical point of view, a phased array represents a phased transmission grating. In contrast to more conventional devices of this kind, it allows for a huge phase shift and thus for operation at extremely high diffraction orders. In this section, we will examine the design of phased arrays, in particular the construction of imaging devices and their aberration analysis. A brief examination of Rowland and stigmatic mountings will be included.

#### 8.3.1 Principle of Operation

Figure 8.15 shows a schematic drawing of a phased array. In this diagram the light propagates, as usual, from the left to the right. The point of incidence  $I$  of the phased array is located at the beginning of the left slab waveguide. Starting from this point, the input beam propagates under the influence of diffraction through the homogeneous slab waveguide. At the front end of the bundle of strip waveguides forming the phase shifter, the optical far-field represents the Fourier transform of the near-field of the input beam.<sup>13</sup> The beam is now divided into  $N$  partial beams, each propagating separately through one of the strip waveguides forming the phase shifter. Obviously, the phase shifter allows the phase positions of the partial beams to be adjusted relative to each other, i.e., the phase portrait at the grating line, which is located at the input side of the second slab waveguide, to be tuned. We see furthermore that the endfaces of the strip waveguides at the grating line replace the grating grooves of a conventional grating. If the front and back of the phase

13. See Section 4.3 for a discussion of the Fraunhofer diffraction integral.

shifter have the same layout, the optical field at the grating line will represent the Fourier transform of the near-field at the input side modified by the phase portrait due to the phase shifter. We have seen in Section 8.1.4 that a constant phase shift from one waveguide to another will result in a rotation of the diffraction pattern while the envelope of the peaks formed by the Fourier transform of the near-field of a single waveguide remains unchanged. The near-field at the point of incidence will be reconstructed at the focal curve provided the optical system of the phased array does not exhibit any aberrations.

Thus the optical system of a phased array can be regarded as an extended lens in which the left-hand slab waveguide acts as the input side and the right-hand one as the output side, respectively. If we replace the slab waveguide on the left-hand side by a circuit of strip waveguides (e.g., a tree of  $y$ -branches) we will affect the shape of the power distribution at the grating line and thus the shape of the peaks at the focal curve. The blazing of phased arrays is straightforward; i.e., the axes of the waveguides ending at the grating line must intersect the focal curve at a single point<sup>14</sup>.

The period of the phased array is given by the spacing of the strip waveguides at the back of the phase shifter (projected to the tangent at the center of the grating line). The required decoupling of the strip waveguides forming the phase shifter will set a lower bound to the period. On the other hand, the period of the phased array should be minimized in order to increase the on-off ratio along the grating line and thus the relative width of the envelope of the diffracted pattern. This means that we will usually choose the minimum period according to the requirements of the phase shifter. Due to the restrictions of the period, practical designs tend to be based on the maximum diffraction order allowed by the free spectral range of the underlying application (see inequality (8.38)).

We will now end our more qualitative discussion and start the quantitative analysis. *For a phased array, the light-path difference between the point of incidence  $I$  and the grating line  $z_G(y)$  is given by*

$$F_I(y) = \overline{IP_I} - \overline{IO_I} + \frac{n_P}{n_S} (\widetilde{P_I P} - \widetilde{O_I O}) \quad (8.47)$$

where  $O$  represents the center of the grating line and  $P$  an arbitrary point on it. The other two points  $O_I$  and  $P_I$  are located at the input side of the phase shifter, connected with their counterparts  $O$  and  $P$  by the strip waveguides of the phase shifter. The tilde on top of the symbols on the right-hand side of Equation (8.47) designates the arc length along the corresponding strip waveguides. The ratio  $n_P/n_S$  accounts for the phase difference caused by the different effective propagation constants of the strip ( $n_P$ ) and slab ( $n_S$ ) waveguides. The front side of the phase shifter usually represents a circle whose center is located at the point of incidence. The first term on the right-hand side of Equation (8.47) will then vanish.

14. See Section 8.1.4 for a more detailed discussion of the blazing.

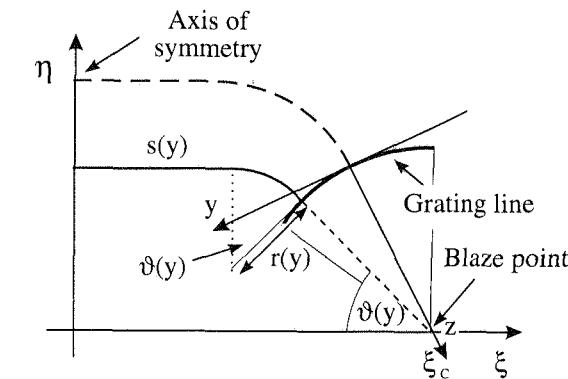


Figure 8.16. Phased array: Rowland mounting.

### 8.3.2 Rowland Mountings

In this paragraph we will discuss Rowland mountings, that represent the most popular layout for phased arrays. In Section 8.1.2 we have seen that the groove function of a Rowland mounting must exhibit linear behavior, i.e.,

$$G(y) = \frac{y}{\Lambda} \quad (8.48)$$

where  $\Lambda$  designates the period of the phased array (projected onto the tangential coordinate axis  $y$ ). The grating line is given by a semicircle of a radius  $R$ . The fan-in of the phased array must be assembled such that the phase portrait becomes linear along the grating line. The light-path difference on the input side is then given by

$$F_I(y) = \left( -m \frac{\lambda_C}{\Lambda} - \sin \alpha_C \right) y \quad (8.49)$$

in which  $\lambda_C$  designates a design wavelength (inside the slab waveguide) and  $\alpha_C$  the diffraction angle for this wavelength.

As discussed in Section 8.1.2, Rowland mountings have no coma. We have seen furthermore that the Rowland mounting for the phased array has a stigmatic point at the vertex of the Rowland circle ( $\alpha_D = 0$ ), i.e., such mountings exhibit extremely small aberrations close to this point.

In the following we will discuss a symmetrical layout of a Rowland phased array [3] in more detail. Each strip waveguide of the phase shifter of this layout consists of a straight section embedded between two circular arcs. The radius of curvature  $r(y)$  and the length of the straight section  $s(y)$  vary from one waveguide to another. With respect to the further analysis, we use the tangential coordinate  $y$  for the parameter representation of the strip waveguides. Figure 8.16 shows half of the symmetrical layout. We will now introduce a new coordinate system with the

coordinates  $\xi$  and  $\eta$  for the discussion of the layout. The vertex of the Rowland circle (or the point of incidence on the input side) is then located at  $C = (\xi_C, 0)$ . The  $\eta$ -axis of the coordinate system represents the axis of symmetry. The local angle of inclination is then given by

$$\vartheta(y) = \vartheta_C - \sin^{-1} \left( \frac{y}{R} \right) \quad (8.50)$$

where  $\vartheta_C$  designates the angle of inclination of the center of the grating line. By using standard analytical geometry we obtain the following expression for the length of the straight section:

$$s(y) = 2 [\xi_C - R \cos \vartheta(y) - r(y) \sin \vartheta(y)]. \quad (8.51)$$

The arc of the curved sections  $r(y)$  is chosen such that all strip waveguides point to the vertex  $C$  of the Rowland circle. Thus the phased array will provide perfect blazing at this point. The parameter  $\xi_C$  must be so large that the straight sections of all strip waveguides remain positive. By using

$$F_I(y) = s(y) + 2\vartheta(y)r(y) - s(0) - 2\vartheta_0r_0 \quad (8.52)$$

we can express the radius of the curved waveguide section  $r(y)$  in terms of the required light-path difference (see Equation (8.49)). In fact, the layout described here has two free parameters  $r_0$  and  $\vartheta_0$  which must be determined such that the waveguides forming the phase shifter remain decoupled and that radiation losses in the curved section are avoided.

### 8.3.3 Stigmatic Mountings

Analogously to the reflection gratings discussed in the last section, we can also construct  $m$ th order phased arrays having two stigmatic points  $D_1$  and  $D_2$  at the wavelengths  $\lambda_1$  and  $\lambda_2$ . By using the Equation (8.2) we see directly that the groove function  $G(y)$  and the light-path difference on the input side must satisfy the following equations.

$$\begin{aligned} -mG(y) &= \frac{(\overline{PD}_1 - \overline{OD}_1) - (\overline{PD}_2 - \overline{OD}_2)}{\lambda_1 - \lambda_2} \\ F_I(y) &= \frac{\lambda_2(\overline{PD}_1 - \overline{OD}_1) - \lambda_1(\overline{PD}_2 - \overline{OD}_2)}{\lambda_1 - \lambda_2} \end{aligned} \quad (8.53)$$

Since the layout of the phase shifter depends on the location of the blaze point, it is impossible to derive here a general procedure for the construction of stigmatic mountings.

## REFERENCES

- [1] Petit, R. (ed.), *Electromagnetic Theory of Gratings*, New York, Springer, 1980.
- [2] Hutley, M. C., *Diffraction Gratings*, New York, Academic Press, 1982.
- [3] Dragone, C., "Optimum Design of a Planar Array of Tapered Waveguides", *Journal of the Optical Society of America*, Pt. A, Vol. 7, 1990, pp. 2081–2093.
- [4] Noda, H., Namioka, T., and Seya, M., "Geometric Theory of the Grating", *Journal of the Optical Society of America*, Vol. 64, 1974, pp. 1031–1042.
- [5] Velzel, C. H. F., "A General Theory of the Aberrations of Diffraction Gratings and Grating Like Optical Instruments", *Journal of the Optical Society of America*, Vol. 66, 1976, pp. 346–353.
- [6] Smit, M., "New Focussing and Dispersive Planar Component Based on an Optical Phased Array", *Electronics Letters*, Vol. 24, 1988, pp. 385–386.
- [7] März, R., and Cremer, C., "On the Theory of Planar Spectrographs", *IEEE Journal of Lightwave Technology*, Vol. 10, 1992, pp. 2017–2022.

## Appendix A

### Application of Photonics

In discussing the applicability of integrated optical circuits, it is useful to start with an analysis of the applications of photonics in general. Photonics covers, in broad terms, the techniques of photonic signaling, detecting and actuating. *Photonic components often substitute their electronic equivalents.* Thus, they must offer benefits in cost or performance if they are to be successful on the market. In most cases, photonic substitution is in fact driven by performance arguments. Figure A.1 shows the arguments in favor of photonics for a collection of application areas. These applications can be divided into two more or less independent classes: one class requiring an *increase of performance*, i.e., for

- high bandwidth,
- low insertion loss and
- low crosstalk;

and another class which tries to *avoid problems of electrical signaling* such as

- electromagnetic interference,
- bias problems,
- wire tapping,
- explosion risk,
- weight and/or size.

The first class of applications, which in fact represents the communications technologies (telecommunications, data transmission, and optical interconnections), calls explicitly for the performance parameters of photonic components. The second class makes use of the inherent properties of photonics. We can conclude that *communication systems constitute the technological driver of photonics.* The complexity of

optical circuits, and in particular the potential for topologically complex routing, is poor in comparison to electronic circuits<sup>1</sup>.

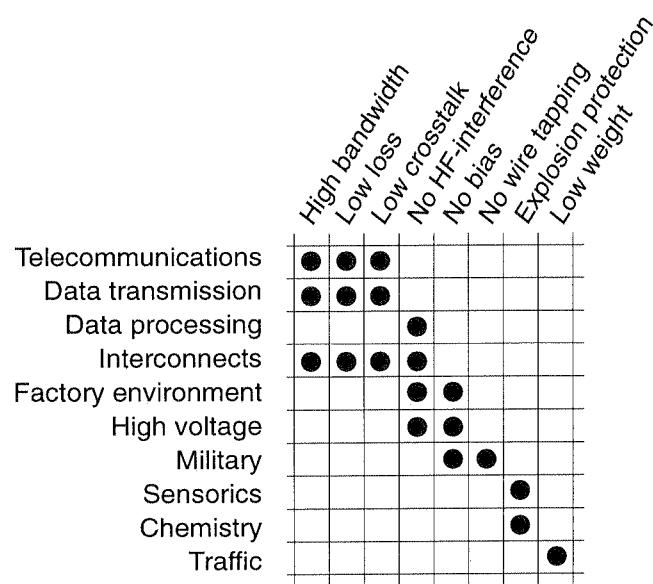


Figure A.1. Arguments for photonics.

This scenario is often formulated by the following rule of thumb: *Let the photons transmit and the electrons switch*. Although this key phrase represents a crude simplification, it can serve as a general guideline through the photonic applications shown in Figure A.1.

### Fiber-Optical Transmission Systems

Optical fibers offer an increase in transmission capacity of many orders of magnitude over conventional transmission lines. The spacing between repeaters can be significantly increased at the same time. This dramatic upgrade in performance allows for the implementation of a series of new services such as HDTV distribution, broadband data highways and video telephony.

Worldwide communication and data transmission networks exhibit a hierarchical topology. Thus, increasing communications and data traffic will result in a successive spread of the optical part of the network from the trunk lines forming its “wide area” level down to its “access” levels.

The optical transmission is usually carried out in the  $1.3/1.5 \mu\text{m}$  wavelength region<sup>2</sup> where optical fibers exhibit minimal transmission losses. The active components (mainly laser diodes and photodiodes) for this wavelength region are usually realized in the InGaAsP/InP material system. The overall bandwidth that can be used for optical transmission is in the order of THz ( $1 \text{ THz} = 10^{15} \text{ Hz}$ ). We have access to this enormous capacity by three essentially different techniques:

- Wavelength Division Multiplex (WDM)

The overall data stream is divided into a set of substreams each having a different wavelength. The optical signals are multiplexed by an optical power combiner or filter. They are demultiplexed by optical filters. Signal reception is by direct detection.

- Optical Frequency Division Multiplex (OFDM)

Again, the overall data stream is divided into a set of substreams each with a different wavelength. In contrast to wavelength division multiplex (WDM) transmission, the signals are detected by a coherent detection scheme which is based on a tunable laser acting as a local oscillator on the receiver side of the transmission line.

- Single-Channel Transmission

In principle, it is possible to use a single optical channel having a huge bandwidth. However, transmission systems of this type cause severe problems on the electronic side of the network. In addition, even small fiber dispersion will limit the possible link length.

Communications networks provide three different classes of service:

- *Distributive services* aim at broadcasting information to all participants of a communications network. They obviously require a minimal amount of switching. TV, radio and electronic newspapers are typical examples of distributive services.
- *Mailing services* support the unidirectional transmission of information from one participant of the communications system to another. In contrast to distributive services, the postal service must establish an (at least virtual) connection for each of its jobs. EMail is typical example of a postal service, fax and “file transfer” services could be organized in the same way.
- *Interactive services* allow bidirectional transmission between two (or more) participants. In addition to the need for switching, their interactivity results in stringent requirements on the tolerated delays for the underlying communications channels. Telephony, video conferences and “remote login” to computer systems represent typical examples of interactive services.

1. For special purpose processors, however, it is possible to design compact massively parallel optical units based on arrays of identical devices such as “smart pixel” or SEED arrays.

2. Small size local area networks can also be operated with GaAlAs/GaAs sources ( $0.8 \mu\text{m}$  wavelength region).

Within an evolutionary scenario future optical networks will act as a neutral transport medium for such a mix of incompatible services. This means that *optical networks require transparency with respect to changes in protocols, bit rates and modulation formats*. Obviously, this required specification can best be obtained by photonic signaling in which a single photonic channel acts as a container for a single service or a collection of compatible services. A cross connect in the network can then be based on an optical  $N \times N$  space switch. Wavelength routing, i.e., the addressing of output ports by wavelengths within the framework of a “broadcast and select” strategy, represents an alternative realization of such cross connects.

Upgraded versions of transparent networks will use wavelength channels as containers in order to utilize the bandwidth of the fiber more efficiently. To conserve the full connectivity within such networks, the wavelength channels must be switched just like fiber channels (wavelength conversion), i.e., the networks must be equipped with wavelength converters.

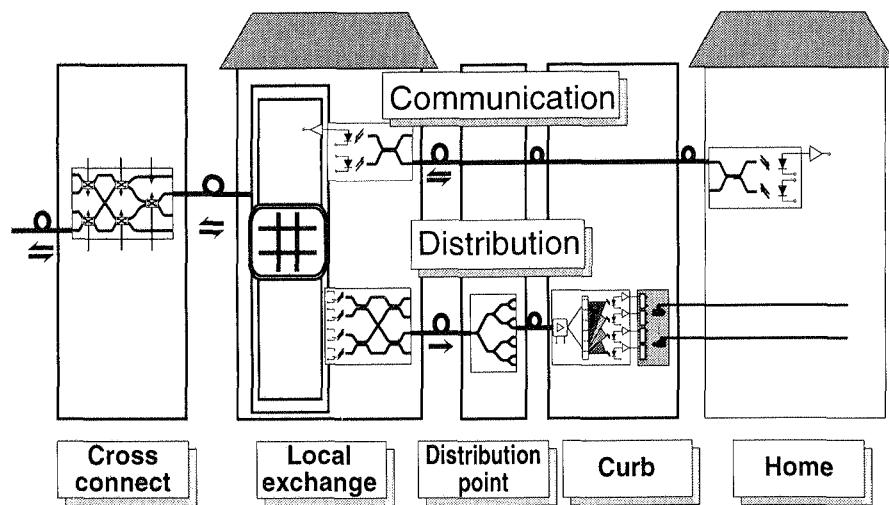


Figure A.2. Public communications network.

Figure A.2 illustrates a part of a public communications network and some optical components – power splitters, optical filters and switches – that can be used within such an environment. Today, large optical communications networks are usually based on single-mode fibers. The required specifications for optical devices inside optical networks can be quite different. As a rule of thumb, *the wide area levels of a communication network call primarily for high performance, and its access levels for low cost*. In addition, we should note that long term stability represents an essential boundary condition for any components used in such networks. Integrated optical components for communications networks – such as power splitters and

external modulators – are commercially available today.

### Optical Interconnections

As data rates per channel increase, highly parallel electrical interconnections become significantly affected by bias problems and electromagnetic interference between adjacent data streams. The degradation will grow with increasing link length.

In consequence, the electrical cabling inside electronic systems exhibiting large internal data streams is being successively replaced by optical interconnections that avoid both problems. The substitution always starts from the frame-to-frame and cabinet-to-cabinet interconnections which are usually realized by fiber ribbon cables (long distance). It is continued by implementing board-to-board interconnections via an optical backplane (medium distance). In extreme cases even chip-to-chip interconnections will be required (short distance).

Large switching fabrics inside local exchanges and high performance computer systems are typical candidates for internal cabling by optical interconnects.

Optical interconnects are operated in both the  $0.8 \mu\text{m}$  and the  $1.3/1.5 \mu\text{m}$  wavelength regions. Since they represent typical low cost devices, many optical interconnection systems are still based on multi-mode waveguides or fibers. In order to minimize delays, the number of electro-optical conversions is usually minimized. Optical backplanes have been realized by using the technologies of integrated optics.

### Sensors and Sensor Systems

Sensors are part of most systems today, and they will play an increasing role with the growing possibilities offered by signal processing. Photonics can be used within sensor systems for detection, communication and power supply. Competitive advantages of optical communications and power supply include: no electromagnetic interference, electrical isolation and high explosion safety. The photonic sensors themselves can be divided into two classes: true photonic and photon-assisted sensors. The former detect and/or analyze the optical signals. Examples of this type of sensing are

- barrier or flame detection,
- absorption analysis and
- spectroscopy.

Photon assisted sensors detect the influence of an acoustical, electrical or chemical signal on the behavior of an optical wave. Examples are

- interferometry,
- optical time domain reflectrometry (OTDR),
- gyroscopy,
- hydrophony,

- acceleration sensing,
- temperature sensing,
- vapor sensing (e.g., SO<sub>2</sub> and NH<sub>3</sub>) and
- voltage and current detection.

Photonic sensors find applications in industrial and military environments (e.g., chemical plants, oil platforms), transportation (e.g., aircrafts), electrical power plants and distribution systems, robotics and machine control. High costs often prevent more extensive use of photonic sensors. The materials used for these sensors are as diverse as their applications. A part of the sensors – such as interferometers, gyroscopes, acceleration and vapor sensors – were realized by using the technologies of integrated optics.

#### **Others**

The applications listed up to now represent the most important areas in which integrated optics are used. Nevertheless, photonics in general has a number of other applications in

- consumer photonics (CD players, laser pointers),
- peripherals (flat panel displays, laser printers, CD-ROMs and MO discs),
- industrial environment (control and communication systems because of no electromagnetic interference and high explosion safety, sensors and high-power lasers),
- medical environment (sensors, lasers and tomography) and
- military environment (sensors, filters, control and communication systems to avoid wire tapping).

Only a few integrated optical components are relevant to these application fields. Examples of this kind are disk-pickups for CD players and CD-ROMs and high-power lasers based on laterally coupled laser diodes.

## **Appendix B**

### **Computer Aided Engineering**

Within the framework of this book we have so far analyzed the design and modeling of integrated optical devices. We will now examine the complete scenario of computer aided engineering (CAE) in order to find out how the methods presented here fit into this environment. We will also discuss some aspects of the implementation of a powerful BPM software. At the end of this section we will tackle some problems by means of rapid prototyping.

#### **B.1 SOFTWARE ENVIRONMENT**

Figure B.1 shows the fully developed CAE environment for a chip technology. The two essential tasks of computer aided engineering are

- modeling (thick arrows), i.e., process and device modeling,
- preparing layouts (thin arrows), covering both circuit simulation and computer aided design.

The modeling path starts from a physical description of a component, consisting of a set of reticles for its fabrication. In a first step, “process modeling” will yield the device description (e.g., a directional coupler) providing its geometry and materials (e.g., the dielectric profile). The subsequent “device modeling” (e.g., the beam propagation method) analyzes the functional behavior of the device starting from this description. It delivers a set of device parameters (e.g., the SPICE parameters of microelectronics) describing the operation of the device within a circuit or any other composite structure. The layout path, in contrast, starts from the system design. The physical description, i.e., the true mask data of the chip, is worked out from the system design by using one (or more) intermediate levels of circuit descriptions. We should note that these two paths of computer aided engineering are to a large extent independent of each other.

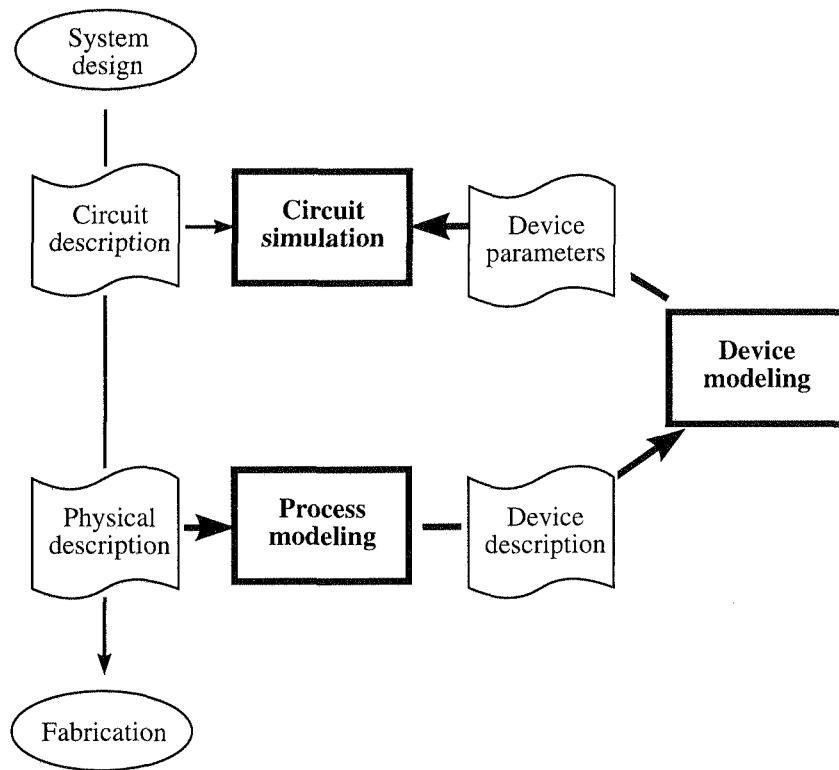


Figure B.1. Software environment for computer aided engineering.

Only microelectronics, the most established chip technology, exhibits a fully developed CAE scenario.

### Process Modeling

Process modeling supports the development of technology by computer aided engineering. It provides tools for two essentially different tasks: device fabrication and equipment simulation. The simulation tools used here are based on rigorous calculations and often also on heuristic models which must be fitted to the equipment and process conditions. Commercial software packages simulating lithography, dry etching (e.g., SAMPLE), implantation, diffusion and oxidation (e.g., MIMAS and SUPREM) are available to support the device fabrication,. For optimizing the processes themselves, simulators such as PHOENICS can be used to analyze the gas flow inside the process chambers.

All the tools used for process modeling have been developed for microelectronics. We should note that process modeling is so far rarely applied within the framework of integrated optics.

### Device Modeling

Originally, the development of new devices was accompanied by theoretical investigations aimed at obtaining a deeper understanding of their behavior and thus at deriving strategies for their layout. In fact, this objective represents the essential motivation for this book. With growing technological expenditures, an increasing demand can be observed for quantitative predictions of device behavior since precise layouts reduce the number of development cycles, and thus time and cost. This evolution has been accompanied by a trend from more qualitative modeling to numerically oriented predictions of device behavior.

Device modeling covers two essentially different methods: behavioral simulation and device analysis. In broad terms, behavioral simulation represents a computer experiment, i.e., it shows the response of a device to a certain stimulus. The beam propagation method (BPM) derived in Chapter 4 as well as ray tracing are the most important tools in this field applied to integrated optics.<sup>1</sup> The term *device analysis* covers direct design methods starting from first principles or from phenomenological descriptions. For integrated optics, they are usually based on a knowledge of the eigenmodes.

Software packages for beam propagation, ray tracing and eigenmode analysis are commercially available today.

### Circuit Simulation

Within microelectronics, the field of very large scale integration (VLSI) is characterized by great circuit complexity. Consequently, the corresponding CAE scenario exhibits several levels of circuit description. Integrated optical circuits, in contrast, are of such low complexity (< 100 devices/circuit) that no circuit description has yet been established. Nevertheless, the transfer matrix description derived in Chapters 6 and 7 can be regarded as a precursor of a description for certain classes of integrated optical devices.

### Computer Aided Design

Computer aided design (CAD) systems are used to translate the system design into a physical description, i.e., into a set of masks for the fabrication of a chip. CAD systems are intended to speed up the design process. They should also support the realization of non-erroneous layouts by providing an appropriate design strategy (“correctness by construction”) or at least (a posteriori) error indicators.

Integrated optical circuits are formed from a small number of complex-shaped basic structures (“primitives”). Examples of such primitives are straight and curved waveguides, Y-branches, directional couplers, and Bragg gratings. The CAD system should handle an integrated optical design in terms of its physical parameters (e.g., waveguide width, radius of curvature) and not of the polygons occurring on the masks. Since kink angles of a few degrees<sup>2</sup> between its strip waveguides will

1. Analog tools are used in microelectronics to predict the SPICE parameters for a new generation of chips.

2. Such weak kinks often cannot be visually detected.

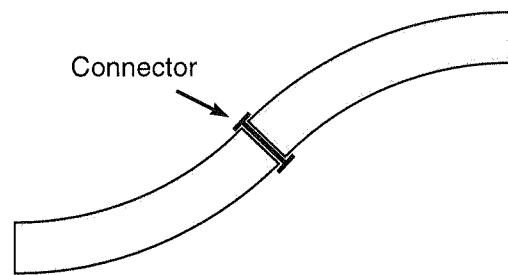


Figure B.2. CAD for integrated optics: design strategies.

significantly affect the operation of an integrated optical circuit an appropriate CAD-system must also support the kink-free construction of integrated optical circuits. This can be done by introducing a “connector” which acts as an auxiliary primitive inside the CAD system (see Figure B.2). The SIGGRAPH-OPTIK CAD system,<sup>3</sup> for example, is based on such a design strategy.

The software tools for integrated optics are developed by independent groups in a multi-site environment on mainframes, workstations and personal computers.

Powerful, flexible and user-friendly standard interface definitions are required to connect the CAD/CAE tools under such circumstances. The PIF (profile interchange format) [1] represents an attempt to establish a standard interface format for the exchange of data between device and process modeling tools for microelectronics (see Figure B.3). There are still no well-established standards for integrated optics. Some BPM packages provide interfaces to the SIGGRAPH-OPTIK CAD system.

## B.2 BEAM PROPAGATION METHOD

We will now discuss some aspects of the implementation of a BPM software by way of an illustrative example. A discussion of the underlying algorithms is found in Chapter 4.

### Device Definition

To meet the requirements of flexibility, a description must be provided of an arbitrary integrated optical device, i.e., of the entire refractive index profile. The device description should be generated outside the main BPM software. This, in turn, requires the support of interface formats for data exchange. If these formats are easy to handle, a device description can be implemented by using a set of device

3. SIGGRAPH is a registered trademark of Siemens Nixdorf Informationssysteme AG.

```
(PIF
***** geometry from Section IV.C.5 *****
***** grid from Section IV.C.5 *****
***** boron profile from Section IV.E.1 *****
(materialType
  (specificationName "MaterialTypeList")
  (table
    (objectName "Segment")
    (data "silicon" "siliconDioxide" "polysilicon")
    (dielectricConstant 11.9 3.9 11.9)
    (electronMobility 1000. ? 1000.)
    (holeMobility 500. ? 500.))
  (snapshot
    (snapshotName "TimeStep1")
    (units "microns")
    (usesGeometry "FloatingGateStructure")
    (usesGrids "CoarseTriangularGrid")
    (usesAttributes "BoronProfile" "MaterialTypeList")
    (attribute
      (attributeName "TimeStep")
      (valueType shortReal)
      (units "ns")
      (value .1))
    (TerminalBias(value(objectName "SourceContact") (data(0.0))))
    (TerminalBias(value(objectName "DrainContact") (data(1.0))))
    (TerminalBias(value(objectName "GateContact") (data(1.0))))
    (TerminalBias(value(objectName "BaseContact") (data(0.0))))
    (TerminalCurrent 1.E-7)))
  )
```

Figure B.3. PIF example.

generators, each describing one class of devices. For each new class of devices, only a new device generator needs be implemented while the rest of the software remains unchanged. A more sophisticated attempt to provide a flexible device description is to use a CAD system. However, process modeling tools to describe the refractive index profile in terms of the entire mask data are then necessary.

### Stimulus Definition

A flexible stimulus definition is as important as a flexible device description. This should include the ability to excite the device with eigenmodes of the input cross section or of a part of it. If the input of eigenmodes is not supported, the BPM results are always accompanied by a considerable level of radiation which makes it

hard to derive extinction ratios or crosstalk attenuations.

### Memory Management

The beam propagation method, like all computer simulations dealing with partial differential equations, produces an overwhelming amount of data which tends to exceed every mass storage device independent of its size. So every BPM program takes advantage of an efficient mass storage management system which reduces the amount of output at the earliest possible stage.

### Error Indication

In the previous paragraphs, we studied the influence of both grid and step size on the performance of different BPM algorithms. For realistic devices – an optical taper is a good illustration – it is difficult to keep track of the corresponding conditions without additional software support. This is a particular problem for “error-tolerant” algorithms such as FE-BPM, which do not indicate erroneous results by their noisy shape.

### Phase Fronts

One of the most attractive options of the beam propagation method is to observe the phase fronts within a given integrated optical device. A straightforward calculation of the phase fronts results in a non-physical representation, since the optical phase is determined only up to modulo  $2\pi$ . The reconstruction of the phase fronts by continuous assembly runs into problems since the procedure is sensitive to phase noise occurring in regions with practically no optical power. It is therefore the best way to use either the real or the imaginary parts of the optical field to represent the phase fronts.

### Tracing of Power Flux

In addition to the straightforward features just listed, a fully developed BPM program should be able to track the optical power along an arbitrary trace within an integrated optical device. A trace  $\mathbf{r}_{\text{trace}}(z)$  may be predefined externally as the center line of a waveguide, for example. Alternatively, the trace may be derived from the results of the current BPM run; for example, it could be defined by the locations of the intensity maxima. Both strategies for defining a trace have advantages. We can now observe the evolution of optical power inside the device along such traces. First of all, we can calculate the confinement factor, i.e., the total optical power within a subwindow of the computational window. The subwindow of a size  $(\Delta x, \Delta y)$  itself is adjusted to the previously defined trace  $\mathbf{r}_{\text{trace}}(z)$ . The confinement factor is then given by

$$S_{\text{tot}}(\mathbf{r}_{\text{trace}}(z), \Delta r) = \frac{1}{2} \operatorname{Re} \left( \int_{x_{\text{trace}}-\Delta x}^{x_{\text{trace}}+\Delta x} dx \int_{y_{\text{trace}}-\Delta y}^{y_{\text{trace}}+\Delta y} dy (\mathbf{E} \times \mathbf{H}^*)_z \right).$$

With a little extra effort, we can obtain the optical power stored in a predefined or actual eigenmode of the device by projecting the optical field  $\Psi$  calculated using

the BPM onto that eigenmode  $\Psi_{\text{test}}$  which is also adjusted to the trace, i.e.,

$$P(\mathbf{r}_{\text{trace}}(z)) = |\langle \Psi(\mathbf{r}) | \Psi_{\text{test}}(\mathbf{r} - \mathbf{r}_{\text{trace}}(z)) \rangle|^2.$$

Figure B.4 shows a possible BPM software environment. In this diagram, the rectangular boxes stand for executable programs that can run on different computer systems. We recognize four different programs, the BPM providing the core algorithm, the EIGENMODE GENERATOR providing stimuli, the DEVICE GENERATOR providing device-dependent data and the BPM ANALYSIS providing the power flux analysis. Within the framework of a fully developed CAE scenario, the process modeling program will act as a device generator.

The different tools interact with each other via data files represented in this diagram by stylized papers. The data files are formatted using one or more data interchange formats. The arrows between the programs used in the diagram indicate the direction of data flow. As expected, we must first run the DEVICE GENERATOR. It produces the refractive index profile  $n(\mathbf{r}_t, z)$  for the BPM and some refractive index profiles  $n_j(\mathbf{r}_t)$  for the EIGENMODE GENERATOR. If the device is not excited by an eigenmode, the stimulus may also be generated directly. The DEVICE GENERATOR may then produce some traces  $\mathbf{r}_{\text{trace}}(z)$  for the BPM ANALYSIS. In the second step, we have to start the EIGENMODE GENERATOR to produce the stimulus  $\Psi(\mathbf{r}_t, z_{\text{initial}})$  for the BPM and test functions  $\Psi_{\text{test}}(\mathbf{r}_t)$  for the BPM ANALYSIS. All data for starting the BPM are now available. The results of the BPM will be the optical field within the integrated optical device  $\Psi(\mathbf{r}_t, z)$  and the optical field at the end of the device  $\Psi(\mathbf{r}_t, z_{\text{end}})$ , which may be used as the input for a subsequent run of the BPM. As the last of our four programs, we start the BPM ANALYSIS to obtain the derived properties such of the device as insertion loss or crosstalk attenuation. Obviously, the whole procedure can be carried out in both interactive and batch mode.

### B.3 RAPID PROTOTYPING

In the previous section we elaborated the structure of a typical BPM software. The development of such a platform represents clearly a reasonable software project. It is therefore useful to test the underlying algorithms within a rudimentary prototype before implementing the complete platform. For small design and modeling problems, the implementation of such prototypes will often turn out to be sufficient.

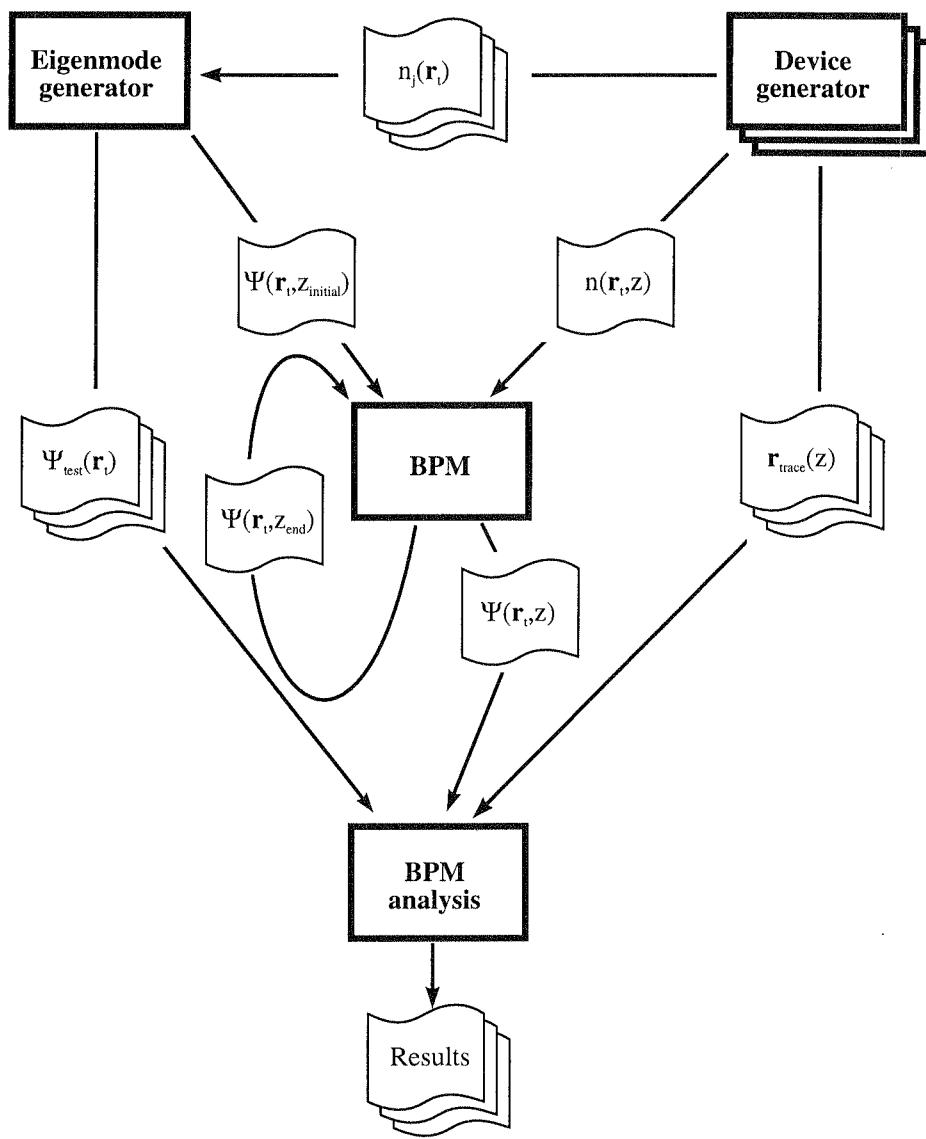


Figure B.4. Possible BPM software environment.

Rapid prototyping, i.e., the implementation of rudimentary prototypes, is supported by several software systems. All these systems provide a toolbox of mathematical and graphical routines. We will now present prototypes for three completely integrated optical problems which were discussed within the framework of this book. The programs were realized by using the MATLAB and MATHEMATICA programming systems.<sup>4</sup>

#### Free Space Beam Propagation Method

As our first example we will discuss a MATLAB program which simulates the propagation of a Gaussian beam in a slab waveguide on the basis of the classical beam propagation method:

```

% ***** CLASSICAL FREE SPACE BPM *****
x0=12;                                % transverse boundary [um]
dx=2*x0/256;                            % sampling interval, N=256
dz=1;                                    % stepsize [um]
zSteps=10;                               % number of steps
q=2*pi;                                  % wavenumber [1/um]
xPts=-x0:dx:x0-dx;                      % sampling: position space
qPts=-pi/dx:pi/x0:(1-dx/x0)*pi/dx;     % sampling: Fourier space
field=fftshift(exp(-xPts.*xPts));        % stimulus (std. Gaussian)

for m=1:2*x0/dx                         % sqrt-propagator
  bpmProp(m)=exp(-i*dz*qPts(m)^2/(q+sqrt(q^2-qPts(m)^2)));
end
bpmProp=fftshift(bpmProp);

for z=dz:dz:zSteps*dz
  field=fft(field);                      % FFT
  field=field.*bpmProp;                  % propagation
  field=ifft(field);                   % inverse FFT

  plot(xPts,abs(fftshift(field)));    % display
  xlabel('x [um]');
  ylabel(sprintf('abs(field) at z=%5.3f um',z));
  pause(0.1);
end
  
```

4. MATLAB is a trademark of The MathWorks Inc.. It represents a bundling of the "Los Alamos" algorithms (LINPACK, EISPACK, ...) and graphical routines for numerical computation. MATHEMATICA is a registered trademark of Wolfram Research Inc. It represents a general software system for both numerical and symbolic computation.

The simulator starts with a definition of the computational window, the propagation steps and the stimulus. What follows is a definition of the square root propagator `bpmProp`. The three lines of code

```
field=fft(field); % FFT
field=field.*bpmProp; % propagation
field=ifft(field); % inverse FFT
```

represent the kernel of the classical BPM algorithm (see Section 4.2 for a more detailed discussion of the classical beam propagation method).

### $\Delta\beta$ -Coupler

This sample MATHEMATICA program provides the switching curves of a  $\Delta\beta$ -coupler consisting of two periods.

```
(* ***** DELTA BETA COUPLER ***** *)
Np=2; (* number of periods of the coupler *)

DeltaKappa[m_,t_] := (* RETURN: {delta(t),kappa(t)}*Lambda *)
Module[{mM=Mod[m,4*Np],
      mD=Quotient[m,4*Np],rho,sChi},
      If[mM == 0, (* ?? m=4*Np,8*Np,... ?? *)
          {2*mD*Cos[(t+1)*Pi/4], (* yes! *)
           2*mD*Sin[(t+1)*Pi/4]},
          rho=2*(mD+.5-t*(mM-2*Np)/(4*Np)); (* no! *)
          sChi= Abs[Sin[m*Pi/(4*Np)]/Sin[rho*Pi/2]];
          {rho*Sqrt[Abs[1-sChi^2]],rho*sChi}]; (* ENDIF *)];

ParametricPlot[(* ** SWITCHING DIAGRAM ** *)
{DeltaKappa[1,t], (* = DeltaKappa[7,t] *)
 DeltaKappa[2,t], (* = DeltaKappa[6,t] *)
 DeltaKappa[3,t], (* = DeltaKappa[5,t] *)
 DeltaKappa[8,t], (* ***** circle ***** *)
 DeltaKappa[9,t], (* = DeltaKappa[15,t] *)
 DeltaKappa[10,t], (* = DeltaKappa[14,t] *)
 DeltaKappa[11,t]}, (* = DeltaKappa[13,t] *)
 {t,-1,1}, AspectRatio->1, Frame->True,
 FrameLabel->{"delta*Lambda/Pi","kappa*Lambda/Pi"}];
```

It has two parts. The module `DeltaKappa[m,t]` provides a parameter representation

$$r_m(t) = \begin{pmatrix} \delta_m(t)\Lambda \\ \kappa_m(t)\Lambda \end{pmatrix}$$

of an individual switching curve. The parameter  $t$  is chosen such that the switching curves intersect the  $\kappa$ -axis at  $t = \pm 1$ . The second part provides the graphical representation of the switching curves. Since some of these curves (e.g., `DeltaKappa[1,t]` and `DeltaKappa[7,t]`) are identical, only a few curves need be plotted.

### Light-Path Function of Planar Spectrographs

The first two examples provide numerical results for a given theory. We will now use MATHEMATICA to verify the expressions (8.9) for the physical light path difference between two rays, i.e., perform an analytical calculation.

```
(* ***** PHYSICAL PATH DIFFERENCE ***** *)
FD[y_] := Sqrt[(y-yD)^2+(zG[y]-zD)^2]-rD; (* path difference *)

zD := Sqrt[rD^2-yD^2]; (* introduce parameters *)
cosD := zD/rD;
sinD := -yD/rD;
TD := cosD^2/rD-zG'',[0]*cosD;
UD := -3*TD*sinD/rD-zG''',[0]*cosD
VD := 3*zG'',[0]^2/rD-3*TD^2/rD-4*UD*sinD/rD-zG''''[0]*cosD;

zG[0_] := 0; (* extra rules for the grating line zG[y] *)
zG',[0_] := 0;

IsZero[expr_] := (* helps to check zG^\nu[0] *)
Module[
{FR=ExpandAll[PowerExpand[ToExpression[expr]]]},
If[ToString[FR]==="0",
Print[expr," == 0, PASSED !"],
Print["ERROR: ",expr," = ",FR," != 0"]]]; (* ENDIF *)

(* Expressions for FD^\nu[y] at y==0 *)
IsZero["FD[0]"];
IsZero["FD'[0]-sinD"];
IsZero["FD''[0]-TD"];
IsZero["FD'''[0]-UD"];
IsZero["FD''''[0]-VD"];
```

At the beginning of this program we set up the physical light path difference  $F_D(y)$ , the polar coordinates  $r_D$  and  $\alpha_D$  and the other parameters  $T_D$ ,  $U_D$  and  $V_D$  defined within the Equations (8.9). We should note that the grating line  $z_G(y)$  has an arbitrary shape which is restricted only by the two extra rules  $z_G(0) = 0$

and  $z'_G(0) = 0$ . The module `IsZero["Expr"]` checks whether the fully expanded expression "Expr" equals zero.

We see that none of these examples is trivial. Nevertheless, each of the corresponding prototypes can be realized within a few hours.<sup>5</sup>

## REFERENCES

- [1] Duvall S. G., "An Interchange Format for Process and Device Simulation", *IEEE Transactions on CAD*, Vol. 7, 1988, pp. 741–754.
- [2] Selberherr, S., *Analysis and Simulation of Semiconductor Devices*, New York, Springer, 1984.
- [3] Müller-Nawratil, R., and März, R., "A Layout Editor for Integrated Optics", *Proc. 7th Conference on Integrated Optics and Optical Fiber Communication (IOOC)*, Vol. 2, Kobe, July 18-21, 1989, pp. 152–153.
- [4] März, R., "Computer Aided Engineering for Microsystems – Status and Trends", *Proc. 1st Conference on Micro System Technologies (MST)*, Berlin, Sept. 10-13, 1990, pp. 31–37.
- [5] Wolfram, S., *Mathematica*, 2 ed., New York, Addison Wesley, 1991.

## Appendix C Component Cost Modeling

Within this appendix, we will examine a phenomenological yield and cost model based on easily obtainable data that allows a rough cost forecast for future chips manufactured in today's technology. Within this model, the cost evolution is (formally) divided into fixed costs (cost evolution at 100% yield) and yield evolution. However, both fixed costs and yields are not calculated in absolute terms but relative to production figures for closely related chips. For example, a laser diode based on a one-step epitaxy may be used as a reference chip for monolithically integrated optical chips in the InGaAsP material system. In selecting this reference chip, it is merely important that both the new and reference chips are manufactured by the same technological process. This procedure offers decisive benefits for cost forecasting at the research stage, as it is extremely inexpensive and also includes future production conditions in the forecast by transferring costs and yields from the production process.

Statistical models like the one described here imply a fully "run-in" production line. Specific difficulties – such as manipulation problems and confusions that are largely impervious to modeling – arise in the starting phase of a line or in laboratory-scale production (i.e., only a few wafers/week). The yields calculated here can thus be understood to represent an optimistic estimate which is, however, based on today's technology.

From the system supplier's point of view, not the chip costs but the module costs, that largely stem from the construction and qualification of a module and its subunits, are of importance. It is therefore in general no use to optimize chip costs independently of module costs. Rather will it be necessary to define the boundary between integration and hybrid design a new after each overall optimization. To do so, typical assembly techniques (flip chip, solder bump technique etc.) will in future have to be modeled with respect to the resulting costs.

5. In fact, many diagrams shown in this book are "quick and dirty" prototypes.

### C.1 YIELD MODELING

The phenomenological yield model developed in the following describes the statistical occurrence of manufacturing faults during processing. It cannot be used to describe systematical defects (as in the edge areas of a wafer). Also, the technological causes of yield problems can be derived from the results only to a very limited extent.

For the following discussion, we will replace the large number of process steps by a quasi-continuous process time  $t$ . Obviously, this process time is not a time term in the usual sense, as the process time is not continued when the wafer is not being processed. The yield evolution  $\eta(t)$  is then governed by the rate equation

$$\dot{\eta} = -\alpha(t)\eta \quad (\text{C.1})$$

with the initial condition  $\eta(0) = 1$ . The destruction rate

$$\alpha(t) = \dot{n}_D(t)A_c \quad (\text{C.2})$$

is determined by the density  $\dot{n}_D(t)$  of the defects newly appearing in the time window  $[t, t + dt]$  and by the weighted critical chip area  $A_c$ . We should note that  $\alpha(t)$  describes the chip destruction rate in this time window, irrespective of whether the chips were previously still intact or not. The critical areas are divided into different classes with different weights on the basis of their sensitivity to defects. For the monolithic integrated optics in the InGaAsP/InP material system, a suitable division into three classes is based on the internal current density taking into account possible threshold behavior:

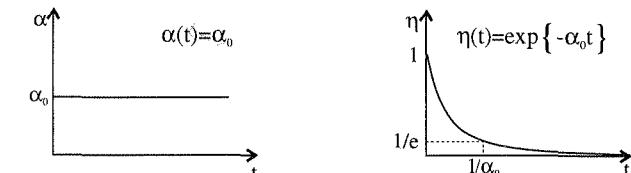
	Weight	Margin
Lasers, amplifiers	1	25 $\mu\text{m}$
Photodiodes, switches	1/2	25 $\mu\text{m}$
Passive waveguides, filters	1/10	5 $\mu\text{m}$

This division into classes reflects the technological experience gained in recent years: (1) lasers do not tolerate fabrication faults at all; (2) photodiodes, only to a modest degree; and (3) passive waveguides, to a considerable degree. For passive components, the component itself and a 5  $\mu\text{m}$  margin around the component was declared a critical area, for active components the margin was widened to 25  $\mu\text{m}$  in view of the danger of shunts.

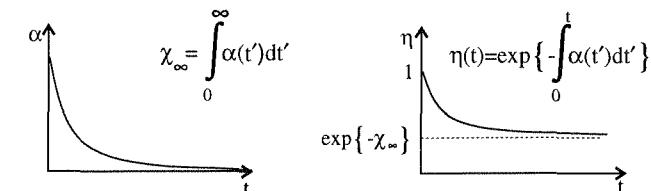
The yield evolution as a solution of Equation (C.1) is given by

$$\eta(t) = \exp(-\chi(t)), \quad (\text{C.3})$$

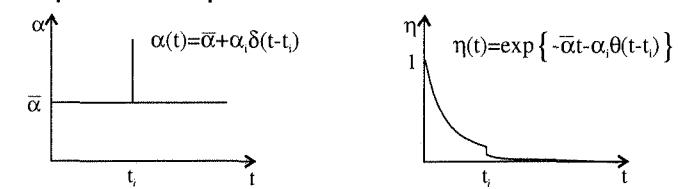
### Continuous process



### Hidden defects



### Critical process steps



### Morphological defects

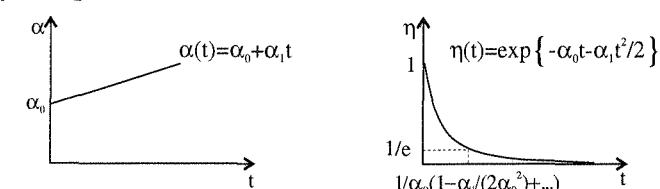


Figure C.1. Different damage mechanisms.

where

$$\chi(t) = \int_0^t \alpha(t')dt'. \quad (\text{C.4})$$

describes the accumulated number of defects. The chip yield  $\eta(t_E)$  then represents the yield of the chips at the end of the processing stage, i.e., at a point in time  $t_E$ .

The following four examples show how situations typical for the development of the technology can be modeled. Figure C.1 shows an overview of the associated profiles of  $\alpha(t)$  and  $\eta(t)$ .

### Continuous Process

The simplest nontrivial assumption for this yield model is a constant destruction rate  $\alpha(t) = \alpha_0$  during the entire process stage. The yield then obeys an exponential law

$$\eta(t) = \exp(-\alpha_0 t) \quad (\text{C.5})$$

with respect to both the total process time  $t$  and the critical area  $A_c$ . The statistical yield models known from the literature [1] provide this result for a Poisson distribution of defects.

### Hidden Defects

Another important defect mechanism is the successive appearance of hidden defects (e.g., dislocations on InP wafers). These defects are by definition initially invisible and appear only in the course of processing. In mathematical terms this represents a (finite) reservoir of defects, i.e.,

$$\chi_\infty = \int_0^\infty \alpha(t') dt' < \infty. \quad (\text{C.6})$$

The yield saturates with loss mechanisms of this kind with a non-negligible value

$$\eta(\infty) = \exp(-\chi_\infty). \quad (\text{C.7})$$

When using a suitable measurement technique, it would of course have been obvious from the start that the process yield cannot exceed  $\eta(\infty)$ .

### Critical Process Steps

A particularly critical process step  $i$  (e.g., selective area epitaxy) at the point  $t_i$  with a destruction rate  $\alpha_i$

$$\alpha(t) = \bar{\alpha} + \alpha_i \delta(t - t_i) \quad (\text{C.8})$$

leads in an otherwise continuous process to an abrupt yield loss

$$\eta(t) = \exp(-\bar{\alpha}t - \alpha_i \theta(t - t_i)). \quad (\text{C.9})$$

This trick easily allows a classical yield model with discrete process steps to be recovered from the model discussed here. For the yield of the  $i$ th process step we simply set  $\eta_i = \exp(-\alpha_i \theta(t - t_i))$  and thus obtain the following for the total yield after  $N$  process steps

$$\eta = \prod_{i=1}^N \eta_i. \quad (\text{C.10})$$

### Morphological Defects

The increasing defect rate in the course of processing  $\alpha(t) = \alpha_0 + \alpha_1 t$ , that may be understood as a consequence of increasingly structured wafer surfaces in the course of the process, represents another defect mechanism of interest. Effects of this type

lead (in comparison to a continuous process) to an increasing destruction rate as the process proceeds. In quantitative terms, we get

$$\eta(t) = \exp\left(-\alpha_0 t - \alpha_1 \frac{t^2}{2}\right). \quad (\text{C.11})$$

The yield model discussed hitherto is based on a destruction rate independent of the position. Experience from silicon technology shows, however, that defects tend to form clusters. Thus the destruction rate depends generally on the position, defect density and process. In chips that are small compared with the mean cluster size, this effect is of minor importance for the chip yield, whereas the cluster formation is clearly noticeable in chips with a large continuous critical area that are manufactured with low yield. The actual yield may then be up to twice as large as that predicted from a position-independent model.

## C.2 YIELD EXTRAPOLATION

We will now specify the yields of future integrated optical chips in terms of the yield of currently manufactured reference chips. This approach offers the following practical advantages:

- The yield evolution under real production conditions is included in the forecast.
- The estimates for the number of required process steps and for the expected critical area are transparent. Incompatible assumptions and consequently incompatible forecasts are therefore largely avoided.
- The entire procedure is simple. Thus a yield estimate can be performed for a project at the planning stage within a few hours. The results of such estimates can serve as a "figure of merit" in the preliminary phase of technological projects.

To ensure that the outlined process is worthwhile, the technology for the future chip and the reference chip (subsequently denoted by the subscript REF), i.e., lithography, etching technique and epitaxy, must be largely compatible. If this condition is satisfied, it can be assumed that temporal averaging over the rate of defect density  $\dot{n}_D(t)$ , i.e.,

$$\int_0^{t_E} \dot{n}_D(t') dt' = \langle \dot{n}_D \rangle t_E \quad (\text{C.12})$$

supplies the same mean value  $\langle \dot{n}_D \rangle$  for both chips. By using Equation (C.2) we obtain

$$\eta = \eta_{\text{REF}}^{a_c \tau}, \quad (\text{C.13})$$

for the yield where  $\eta_{\text{REF}}$  represents the yield of the reference chip,  $a_c = A_c/A_{c,\text{REF}}$  is the normalized critical area and  $\tau = t_E/t_{E,\text{REF}}$  the relative number of required process steps for implementing the new chip.

Obviously, the yield is reduced as the chip area and duration of processing increase. Beyond that, Equation (C.13) shows that the yield evolution proceeds exponentially with both quantities, i.e., dramatic yield reductions are to be expected for  $a_c \tau \gg 1$  (i.e., large critical chip area and/or long process time) or  $\eta_{\text{REF}} \ll 1$  (i.e., low yield of the reference chip). In practice, a dramatic yield reduction must be expected with a large increase of the critical chip area and/or the duration of processing.

### C.3 COST EXTRAPOLATION

The cost of manufacturing a working chip is given by the cost  $C_0$  of a fully processed wafer, the area yield (chip area  $A$ /utilized wafer area  $A_W$ ) and the processing yield  $\eta$ :

$$C = C_0 \frac{A}{A_W} / \eta. \quad (\text{C.14})$$

We will divide the cost  $C_0$  of the processed wafer into the cost of the substrate alone and the processing cost. It should be noted that the cost  $C$  does not include the cost for qualification, i.e., for chip selection. As substrate cost is low compared with processing cost in current integrated optics technology, the former will be neglected.

Using the reference chip from the preceding paragraph again, the following result is obtained for the cost ratio of a working chip:

$$C/C_{\text{REF}} = c_0 a / \eta_{\text{REF}}^{a_c \tau - 1}. \quad (\text{C.15})$$

All parameters in Equation (C.15) – the cost ratio for the fully processed wafer  $c_0 = C_0/C_{0,\text{REF}}$ , the normalized chip area  $a = A/A_{\text{REF}}$ , the critical part of the chip area  $a_c = A_c/A_{c,\text{REF}}$  and the normalized processing time for fully processed chips  $\tau_E = t_E/t_{E,\text{REF}}$  – are normalized to the corresponding values of the reference chip. The numerator in Equation (C.15) is the processing cost ratio per chip, whereas the denominator is strongly yield-dependent. It is clear from Equation (C.15) that the costs are governed essentially by the yield.

Two extremes are considered in estimating the cost ratio  $c_0$  for fully processed wafers:

- An existing production line is not fully utilized. Hence it follows that the new chip can be manufactured at a low additional cost, i.e.  $c_0 \approx 1$ .
- The existing production line is fully utilized. In this case the cost per wafer increases with increasing number of processing steps, i.e.,  $c_0 \approx \tau$ .

In the following, the average value will be used, i.e., partial synergy will be assumed:

$$c_0 = \frac{1 + \tau}{2}. \quad (\text{C.16})$$

### C.4 EXAMPLE: TRANSCEIVER CHIP

For monolithically integrated optics, let us consider a transceiver chip in the InGaAsP/InP material system comprising a laser, monitor, detector and waveguide elements which are distributed over an area of  $A = 2 \text{ mm}^2$  as an illustrative example.

The reference chip for the yield and cost forecast of the transceiver chip is a discrete laser diode (single epitaxy structure) as this chip is already produced in large quantities with high yield. The process time  $\tau$  will be characterized by the number of epitaxial steps. This is considered to be a good approximation, because each epitaxial step is accompanied by a similar set of lithography, etching and cleaning procedures. Hence the number of epitaxial steps governs the complexity and length of the overall process. A more detailed evaluation is of course possible if all the individual processing steps of both the new and reference chip (laser diode) are taken into consideration.

The chip-area or weighted critical chip-area ratio is assumed as  $a = 22$  and  $a_c = 2.5$  respectively. In Figure C.2a, the yield  $\eta$  of the new chip as a function of the yield  $\eta_{\text{REF}}$  of the laser-diode reference chip is shown for different processing time ratios  $\tau$ . The parameter  $\tau = 2, 3, 4, 5$  refers to different processing versions of the new chip with 2,3,4 or 5 epitaxial steps. For example, a yield of  $\eta = 0.45$  is predicted for the new chip with  $\tau = 3$  epitaxial growths if the reference chip is produced with a yield  $\eta_{\text{REF}} = 0.9$  by one epitaxial step.

The cost ratio  $C/C_{\text{REF}}$  of the new chip is plotted in Figure C.2b as a function of the same parameters as in Figure C.2a. Using the example in the preceding paragraph, a cost ratio  $C/C_{\text{REF}} = 87$  is obtained for  $\tau = 3$  and  $\eta_{\text{REF}} = 0.9$ .

### C.5 ARRAYS OF IDENTICAL DEVICES

As a second example, let us consider the yield for arrays of identical devices, e.g., arrays of lasers, detectors or switches. This case is of great importance for a communications technology such as integrated optics where throughput increases with the number of parallel channels.

From the preceding yield model (Equation (C.13)), taking the single device with yield  $\eta$  as the reference chip and setting size  $N$  of the desired array equal to the normalized critical area  $a_c$ , one immediately obtains an estimate for the array yield

$$\eta_N = \eta^N. \quad (\text{C.17})$$

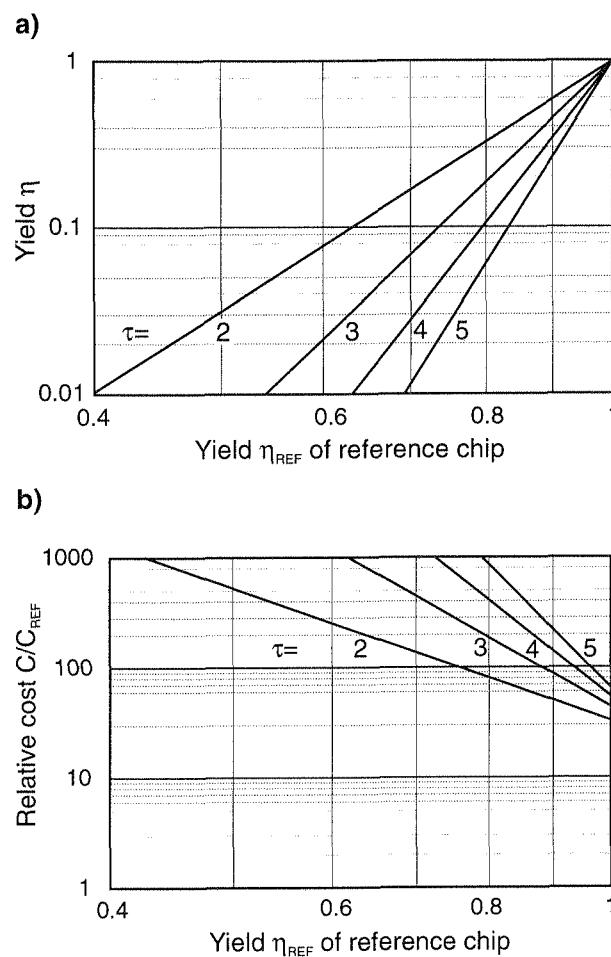


Figure C.2. (a) Yield  $\eta$  of transceiver chip ( $a_c = 2.5$ ) versus yield  $\eta_{\text{REF}}$  of reference laser chip for different process time ratios  $\tau$ . (b) Relative cost  $C/C_{\text{REF}}$  of transceiver chip ( $a = 22$ ,  $a_c = 2.5$ ) compared to reference laser chip (yield  $\eta_{\text{REF}}$ ) for different process time ratios  $\tau$ .

This simple result illustrates that a reasonable array yield  $\eta_N > 1 - \Delta$  can be achieved only if the yield for the individual device is sufficiently high  $\eta \geq 1 - \Delta/N$ .

This yield can be increased by including some redundancy in the design as it is common practice in microelectronics (e.g., in DRAM manufacturing). In the case of arrays of identical devices one can make use of their translation invariance. Thus, if a single row of  $M$  devices, larger than the required array size ( $M > N$ ) is fabricated, arrays can be extracted (diced) from any succession of at least  $N$  functioning devices (see Figure C.3).

The yield for linear arrays with  $N$  devices obtained from a redundant row of  $M$  elements is

$$\eta_N^{(M)} = \frac{1}{[M/N]} \sum_{n=N}^M [n/N] \sum_{w=n}^M (M+1-w) \times \binom{M-1-n}{w-n} \eta^w (1-\eta)^{M-w}. \quad (\text{C.18})$$

where  $[M/N]$  designates the integer division of  $M$  by  $N$ . The statistics of working and faulty devices is described by a binomial distribution. There are  $\binom{M}{w}$  configurations with a probability  $\eta^w (1-\eta)^{M-w}$  with  $w$  working devices out of  $M$ . The result in Equation (C.18) can be obtained by counting all successions of  $n$  working devices which exceed the required array size ( $n > N$ ). The explicit calculation leading to Equation (C.18) shall not be reproduced here.

Figure C.4 shows the yield, according to Equation (C.18), for linear arrays of 8 devices selected from a row of 32 devices as a function of the yield  $\eta$  of the individual device. For comparison, the dashed curve depicts the yield according to Equation (C.17), if individual arrays of 8 elements are fabricated. The yield im-

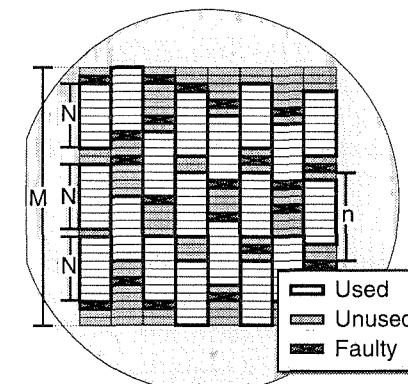


Figure C.3. Arrays of identical devices: layout.

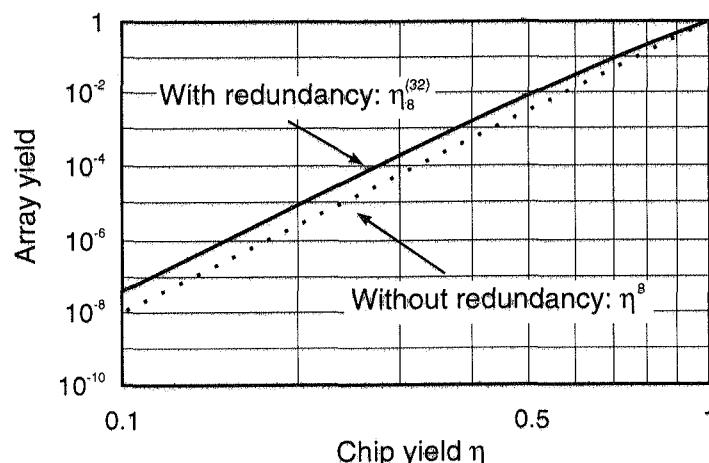


Figure C.4. Yield for an array of 8 devices vs. the single element yield  $\eta$ , as obtained from a redundant array of 32 elements (solid line), and from 4 individual arrays of 8 elements (dashed line).

provement by making use of the translational invariance of arrays is rather modest. In particular, the improvement does not exceed a factor of 2 in the case of high individual yield  $\eta$ , which is indispensable to obtain significant array yield. Thus, the simple model leading Equation (C.17) gives a valuable first approximation for the more laborious statistical evaluation.

## C.6 INTEGRATED OPTICS VS. MICROELECTRONICS

The question inevitably arises as to the essential differences between microelectronics and integrated optics, for in this context each new chip generation in microelectronics has hitherto paid off in terms of cost and yield. However, the less optimistic results obtained here become clear when the following differences are again considered.

- Microelectronics is a digital technology, whereas integrated optics is analog in nature. Accordingly, in most cases narrow tolerances must be maintained and the design of tolerance-insensitive structures will become important.
- In microelectronics, the degree of integration was largely increased by raising the packing density while the chip dimensions remained essentially constant. In integrated optics, it is not expected that the circuits used can be significantly compressed – if only because of the fiber coupling and the input leads. As a

rule, therefore, a greater degree of integration must be traded off by a greater chip area.

- There is still no standard process for integrated optics that could be compared with the CMOS, NMOS or other standard processes in microelectronics. The process technologies currently used in integrated optics are still strongly affected by the requirements of mono-functional discrete components. Yields will probably increase further when the technology is standardized.
- Integrated optics is still a young technology compared with microelectronics, and further significant improvements of processes and structures are to be expected.
- In integrated optics, component concepts are required to reduce the component costs that reduce the chip area (e.g., by three-dimensional structures) and/or simplify the manufacturing technology. It should also be discussed how far reductions in component performance can be tolerated in favor of reduced production costs.

## REFERENCES

- [1] Stapper, C. H., "Large-Area Fault Clusters and Fault Tolerance in VLSI Circuits: A Review", *IBM Journal on Research and Development*, Vol. 33, 1989, pp. 162–173.
- [2] Hu, S. M., "Some Considerations in the Formulation of IC Yield Statistics", *Solid State Electronics*, Vol. 22, 1979, pp. 205–211.
- [3] Ramacher, U., "A Cost Oriented Redundancy Model for Defect-Tolerant VLSI/WSI Systems", in *Yield Modelling and Statistics*, Bristol, Hilger, 1987, pp. 233–245.
- [4] De Bortoli, M., De Padova, M., Lord, R. H., and Moncalvo, A., "Economic Evaluation of Optoelectronic Integration for Optical Broadband Subscriber Loop", *Proc. 15th European Conference on Optical Communication (ECOC)*, Vol. 3, 1989, pp. 120–127.
- [5] Schuelke, R. J., and Pande, K. P., "Manufacturing Cost Analysis of Optoelectronic Integrated Circuits", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 2, 1989, pp. 29–30.

## Appendix D

### Mathematical Background

We will concentrate here on the mathematical background of Chapters 2 and 3, i.e., vector algebra and analysis and on the theory of linear operators.

The calculus of variations is treated in the textbooks by Courant and Hilbert [1] and Morse and Feshbach [2]. Information on the higher functions is available in the handbooks by Gradshteyn and Ryzhik [3] and Abramowitz and Stegun [4].

#### D.1 VECTOR ALGEBRA AND ANALYSIS

We will now introduce the coordinate representations of vectors and the most important vector operators in the three-dimensional space. Since we never use non-orthogonal coordinates here, we will restrict our discussion to orthogonal coordinate systems.<sup>1</sup> The unit vectors  $\mathbf{i}_n$  are then always of unit length, i.e., the scalar product

$$\mathbf{i}_m \cdot \mathbf{i}_n = \delta_{mn}$$

where

$$\delta_{mn} = \begin{cases} 1 & \text{for } m = n \\ 0 & \text{else} \end{cases}$$

denotes Kronecker's  $\delta$ . A vector  $\mathbf{F}$  can be represented as

$$\mathbf{F} = \sum_{n=1}^3 F_n \mathbf{i}_n.$$

The vector product of two vectors is defined by

$$\mathbf{F} \times \mathbf{G} = \sum_{l,m,n=1}^3 \epsilon_{lmn} F_m G_n \mathbf{i}_l$$

---

1. Thus, we can avoid having to introduce the reciprocal coordinate system and having to handle upper and lower indices.

where

$$\epsilon_{lmn} = \begin{cases} 1 & \text{for the even permutations } (lmn) = (123), (231), (312) \\ -1 & \text{for the odd permutations } (lmn) = (321), (213), (132) \\ 0 & \text{otherwise} \end{cases} \quad (\text{D.1})$$

is the skew-symmetric unity tensor. This tensor satisfies the following useful relations

$$\epsilon_{lmn}\epsilon_{nop} = (1 - \delta_{lm})(1 - \delta_{ln})(1 - \delta_{mn})(\delta_{lo}\delta_{mp} - \delta_{lp}\delta_{mo}) \quad (\text{D.2})$$

and

$$\sum_{n=1}^3 \epsilon_{lmn}\epsilon_{nop} = \delta_{lo}\delta_{mp} - \delta_{lp}\delta_{mo}. \quad (\text{D.3})$$

By using these representations, we can easily derive the following formulas for vector algebra

$$\begin{aligned} \mathbf{F} \cdot (\mathbf{G} \times \mathbf{H}) &= \mathbf{G} \cdot (\mathbf{H} \times \mathbf{F}) = \mathbf{H} \cdot (\mathbf{F} \times \mathbf{G}) \\ \mathbf{F} \times (\mathbf{G} \times \mathbf{H}) &= (\mathbf{F} \cdot \mathbf{H})\mathbf{G} - (\mathbf{F} \cdot \mathbf{G})\mathbf{H} \\ (\mathbf{F} \times \mathbf{G}) \times (\mathbf{H} \times \mathbf{I}) &= (\mathbf{F} \cdot \mathbf{H})(\mathbf{G} \cdot \mathbf{I}) - (\mathbf{F} \cdot \mathbf{I})(\mathbf{G} \cdot \mathbf{H}) \end{aligned} \quad (\text{D.4})$$

and vector analysis

$$\begin{aligned} \nabla \times (\nabla \phi) &= \mathbf{0} \\ \nabla \cdot (\nabla \times \mathbf{F}) &= 0 \\ \nabla \times (\nabla \times \mathbf{F}) &= \nabla(\nabla \cdot \mathbf{F}) - \Delta \mathbf{F} \\ \nabla \cdot (\phi \mathbf{F}) &= \mathbf{F} \cdot \nabla \phi + \phi \nabla \cdot \mathbf{F} \\ \nabla \times (\phi \mathbf{F}) &= \nabla \phi \times \mathbf{F} + \phi \nabla \times \mathbf{F} \\ \nabla \cdot (\mathbf{F} \times \mathbf{G}) &= \mathbf{G} \cdot (\nabla \times \mathbf{F}) - \mathbf{F} \cdot (\nabla \times \mathbf{G}) \\ \nabla \times (\mathbf{F} \times \mathbf{G}) &= \mathbf{F}(\nabla \cdot \mathbf{G}) - \mathbf{G}(\nabla \cdot \mathbf{F}) + (\mathbf{G} \cdot \nabla)\mathbf{F} - (\mathbf{F} \cdot \nabla)\mathbf{G}. \end{aligned} \quad (\text{D.5})$$

For the position vector  $\mathbf{r}$  having the length  $r = |\mathbf{r}|$  we find in particular

$$\begin{aligned} \nabla \cdot \mathbf{r} &= 3 \\ \nabla \times \mathbf{r} &= \mathbf{0}, \end{aligned} \quad (\text{D.6})$$

the corresponding equations for the unit radial vector  $\mathbf{r}/r$  are

$$\begin{aligned} \nabla \cdot \frac{\mathbf{r}}{r} &= \frac{2}{r} \\ \nabla \times \frac{\mathbf{r}}{r} &= \mathbf{0}. \end{aligned} \quad (\text{D.7})$$

The most general integral theorems of vector analysis relate integrals over a  $n$ -dimensional volume  $V_n$  to integrals over its  $(n-1)$ -dimensional surface  $\partial V_n$ . We will restrict ourselves to the three-dimensional case. Stokes's theorem

$$\int_A da \mathbf{n} \cdot \nabla \times \mathbf{F} = \oint_{\partial A} dl \cdot \mathbf{F} \quad (\text{D.8})$$

and a related formula

$$\int_A da \mathbf{n} \times \nabla \phi = \oint_{\partial A} dl \phi \quad (\text{D.9})$$

express integrals over an area  $A$  in terms of integrals over the bounding contour  $\partial A$ . The line element  $dl$  and the unit normal vector on the surface  $\mathbf{n}$  are oriented according to the right-hand rule. The following formulas allow to transform the integral over a volume  $V$  into an integral over its surface  $\partial V$ . These are the divergence theorem

$$\int_V dv \nabla \cdot \mathbf{F} = \int_{\partial V} da \mathbf{n} \cdot \mathbf{F} \quad (\text{D.10})$$

and two related formulas

$$\int_V dv \nabla \phi = \int_{\partial V} da \mathbf{n} \phi \quad (\text{D.11})$$

$$\int_V dv \nabla \times \mathbf{F} = \int_{\partial V} da \mathbf{n} \times \mathbf{F} \quad (\text{D.12})$$

as well as Green's theorems

$$\int_V dv (\phi \nabla^2 \psi - \psi \nabla^2 \phi) = \int_{\partial V} da \mathbf{n} \cdot (\phi \nabla \psi - \psi \nabla \phi) \quad (\text{D.13})$$

$$\begin{aligned} \int_V dv [\mathbf{F} \cdot \nabla \times \phi (\nabla \times \mathbf{G}) - \mathbf{G} \cdot \nabla \times \phi (\nabla \times \mathbf{F})] \\ = \int_{\partial V} da \mathbf{n} \cdot [\mathbf{G} \times \phi (\nabla \times \mathbf{F}) - \mathbf{F} \times \phi (\nabla \times \mathbf{G})] \end{aligned} \quad (\text{D.14})$$

in which the unit normal vector  $\mathbf{n}$  points out of the integration volume  $V$ .

### D.1.1 Coordinate Systems

One essential advantage of vector calculus is its coordinate-free formulation, which allows (see Chapter 2) fundamental properties of the electromagnetic field to be expressed without any reference to a coordinate system. However, to obtain numerical results we must introduce appropriate coordinates. Let

$$u_m = u_m(x, y, z) \quad (\text{D.15})$$

represent three orthogonal coordinates, i.e., continuous differentiable functions which may be resolved with respect to the original Cartesian coordinates  $x, y, z$ . Any position vector

$$\mathbf{r} = \mathbf{r}(u_1, u_2, u_3) \quad (\text{D.16})$$

can then be expressed in terms of new coordinates. For orthogonal coordinate systems, the metrical tensor  $\mathcal{G}$  with the tensor components

$$g_{mn} = \frac{\partial \mathbf{r}}{\partial u_m} \cdot \frac{\partial \mathbf{r}}{\partial u_n} \quad (\text{D.17})$$

becomes diagonal ( $g_{mn} = 0$  if  $m \neq n$ ). By using the abbreviations

$$h_m = \sqrt{g_{mm}} = \sqrt{\left(\frac{\partial x}{\partial u_m}\right)^2 + \left(\frac{\partial y}{\partial u_m}\right)^2 + \left(\frac{\partial z}{\partial u_m}\right)^2} \quad (\text{D.18})$$

for the square root of the diagonal elements of  $\mathcal{G}$  and

$$\sqrt{|g|} = h_1 h_2 h_3 \quad (\text{D.19})$$

for the square root of its determinant, we can express the fundamental operators in terms of the new coordinates  $u_m$ . The gradient of a scalar function  $\phi(\mathbf{r})$  is given by

$$\nabla \phi = \sum_{n=1}^3 \frac{1}{h_n} \frac{\partial \phi}{\partial u_n} \mathbf{i}_n, \quad (\text{D.20})$$

the scalar Laplacian<sup>2</sup> of  $\phi(\mathbf{r})$  is

$$\nabla^2 \phi = \frac{1}{\sqrt{|g|}} \sum_{n=1}^3 \frac{\partial}{\partial u_n} \left( \frac{\sqrt{|g|}}{h_n} \frac{\partial \phi}{\partial u_n} \right). \quad (\text{D.21})$$

The divergence of a vector field  $\mathbf{F}(\mathbf{r})$  is given by

$$\nabla \cdot \mathbf{F} = \frac{1}{\sqrt{|g|}} \sum_{n=1}^3 \frac{\partial}{\partial u_n} \left( \frac{\sqrt{|g|} F_n}{h_n} \right) \quad (\text{D.22})$$

and the curl of  $\mathbf{F}(\mathbf{r})$  is

$$\nabla \times \mathbf{F} = \frac{1}{\sqrt{|g|}} \sum_{l,m,n=1}^3 \epsilon_{lmn} \frac{\partial(h_n F_n)}{\partial u_m} h_l \mathbf{i}_l. \quad (\text{D.23})$$

2. To ensure that the vectorial and scalar Laplacians cannot be mixed up, we will always write  $\nabla^2$  for the scalar and  $\Delta$  for the vectorial Laplacian.

It is important to point out here that the Laplacian of a vector field is defined via the vector relation (see Equation (D.5))

$$\Delta \mathbf{F} = \nabla(\nabla \cdot \mathbf{F}) - \nabla \times (\nabla \times \mathbf{F}), \quad (\text{D.24})$$

where the first term can be calculated using (D.22) and (D.20). The second term is given by

$$\nabla \times (\nabla \times \mathbf{F}) = \frac{1}{\sqrt{|g|}} \sum_{l,m,n,o,p=1}^3 \epsilon_{lmn} \epsilon_{nop} h_l \frac{\partial}{\partial u_m} \left[ \frac{h_n^2}{\sqrt{|g|}} \frac{\partial(h_p F_p)}{\partial u_o} \right] \mathbf{i}_l. \quad (\text{D.25})$$

The line element for a path parallel to  $\mathbf{i}_l$  is

$$dl_l = h_l du_l \mathbf{i}_l \quad (\text{D.26})$$

and the surface element for a normal unit vector parallel to  $\mathbf{i}_l$  is

$$da_l = h_m h_n du_m du_n. \quad (\text{D.27})$$

The volume element is given by

$$dv = h_1 h_2 h_3 du_1 du_2 du_3. \quad (\text{D.28})$$

### D.1.2 Generalized Cylindrical Coordinates

General cylindrical coordinate systems leave the  $z$ -coordinate unchanged ( $u_3 = z$ ). Thus, the metrical coefficients are

$$h_1 = h_1(u_1, u_2) \quad h_2 = h_2(u_1, u_2) \quad h_3 = 1. \quad (\text{D.29})$$

In conformity with the terminology of waveguide theory, we will refer to  $u_1$  and  $u_2$  as the *transverse coordinates* and introduce the transverse vector

$$\mathbf{F}_t = \sum_{n=1}^2 F_n \mathbf{i}_n.$$

Since there is no interference between the longitudinal ( $z$ ) and transverse coordinates (see Equation (D.29)), we can introduce transverse differential operators and thus keep the formulations coordinate-free with respect to the transverse coordinates. The corresponding formulas are

$$\nabla \phi = \nabla_t \phi + \frac{\partial \phi}{\partial z} \mathbf{i}_z$$

$$\begin{aligned}\nabla^2\phi &= \nabla_t^2\phi + \frac{\partial^2\phi}{\partial z^2} \\ \nabla \cdot \mathbf{F} &= \nabla_t \cdot \mathbf{F}_t + \frac{\partial F_z}{\partial z} \\ \nabla \times \mathbf{F} &= \nabla_t \times \mathbf{F}_t - \mathbf{i}_z \times \left( \nabla_t F_z - \frac{\partial \mathbf{F}_t}{\partial z} \right) \\ \Delta \mathbf{F} &= \Delta_t \mathbf{F}_t + \frac{\partial^2 \mathbf{F}}{\partial z^2} + \nabla_t^2 F_z \mathbf{i}_z.\end{aligned}\quad (\text{D.30})$$

We can reformulate Equations (D.5) for the transverse differential operators and obtain

$$\begin{aligned}\nabla_t \times (\nabla_t \phi) &= 0 \\ \nabla_t \times (\nabla_t \times \mathbf{F}_t) &= \nabla_t(\nabla_t \cdot \mathbf{F}_t) - \Delta_t \mathbf{F}_t \\ \nabla_t(\phi \cdot \mathbf{F}) &= \mathbf{F}_t \cdot \nabla_t \phi + \phi \nabla_t \cdot \mathbf{F}_t \\ \nabla_t \times (\phi \mathbf{F}) &= \nabla_t \phi \times \mathbf{F} + \phi \nabla_t \times \mathbf{F} \\ \nabla_t \cdot (\mathbf{F} \times \mathbf{G}) &= \mathbf{G} \cdot (\nabla_t \times \mathbf{F}) - \mathbf{F} \cdot (\nabla_t \times \mathbf{G}) \\ \nabla_t \times (\mathbf{F} \times \mathbf{G}) &= \mathbf{F}(\nabla_t \cdot \mathbf{G}) - \mathbf{G}(\nabla_t \cdot \mathbf{F}) + (\mathbf{G} \cdot \nabla_t)\mathbf{F} - (\mathbf{F} \cdot \nabla_t)\mathbf{G}.\end{aligned}\quad (\text{D.31})$$

If  $A_t$  is an area within the transverse plane (i.e., orthogonal to  $\mathbf{i}_z$ ) and  $\partial A_t$  its boundary, the two-dimensional form of the divergence theorem can be written as

$$\int_{A_t} da \nabla_t \cdot \mathbf{F}_t = \int_{\partial A_t} dl \mathbf{n} \cdot \mathbf{F}_t,\quad (\text{D.32})$$

Green's theorems are given by

$$\int_{A_t} da (\phi \nabla_t^2 \psi - \psi \nabla_t^2 \phi) = \int_{\partial A_t} dl \cdot (\phi \nabla_t \psi - \psi \nabla_t \phi)\quad (\text{D.33})$$

and

$$\begin{aligned}\int_A da [\mathbf{F}_t \cdot \nabla_t \times \phi(\nabla_t \times \mathbf{G}_t) - \mathbf{G}_t \cdot \nabla_t \times \phi(\nabla_t \times \mathbf{F}_t)] \\ = \int_{\partial A} dl \cdot [\mathbf{G}_t \times \phi(\nabla_t \times \mathbf{F}_t) - \mathbf{F}_t \times \phi(\nabla_t \times \mathbf{G}_t)].\end{aligned}\quad (\text{D.34})$$

For the transverse differential operators we obtain

$$\nabla_t \phi = \frac{1}{h_1} \frac{\partial \phi}{\partial u_1} \mathbf{i}_1 + \frac{1}{h_2} \frac{\partial \phi}{\partial u_2} \mathbf{i}_2,$$

$$\begin{aligned}\nabla_t^2 \phi &= \frac{1}{h_1^2} \frac{\partial^2 \phi}{\partial u_1^2} + \frac{1}{h_1 h_2} \frac{\partial(h_2/h_1)}{\partial u_1} \frac{\partial \phi}{\partial u_1} \\ &\quad + \frac{1}{h_2^2} \frac{\partial^2 \phi}{\partial u_2^2} + \frac{1}{h_1 h_2} \frac{\partial(h_1/h_2)}{\partial u_2} \frac{\partial \phi}{\partial u_2} \\ \nabla_t \cdot \mathbf{F}_t &= \frac{1}{h_1 h_2} \left[ \frac{\partial(h_2 F_1)}{\partial u_1} + \frac{\partial(h_1 F_2)}{\partial u_2} \right] + \frac{\partial F_z}{\partial z} \\ \nabla_t \times \mathbf{F}_t &= \frac{1}{h_1 h_2} \left[ \frac{\partial(h_2 F_2)}{\partial u_1} - \frac{\partial(h_1 F_1)}{\partial u_2} \right] \mathbf{i}_z.\end{aligned}\quad (\text{D.35})$$

For the transverse vectorial Laplacian we find

$$\begin{aligned}\Delta_t \mathbf{F}_t &= ((\nabla_t^2 + D)F_1 + \Delta_{ND}F_2) \mathbf{i}_1 \\ &\quad + ((\nabla_t^2 + D)F_2 - \Delta_{ND}F_1) \mathbf{i}_2\end{aligned}\quad (\text{D.36})$$

with

$$\begin{aligned}D &= \frac{1}{h_1} \frac{\partial(1/(h_1 h_2))}{\partial u_1} \frac{\partial h_2}{\partial u_1} + \frac{1}{h_1^2 h_2} \frac{\partial^2 h_2}{\partial u_1^2} \\ &\quad + \frac{1}{h_2} \frac{\partial[1/(h_1 h_2)]}{\partial u_2} \frac{\partial h_1}{\partial u_2} + \frac{1}{h_1 h_2^2} \frac{\partial^2 h_1}{\partial u_2^2}\end{aligned}\quad (\text{D.37})$$

and

$$\Delta_{ND}F_n = \frac{1}{h_1} \frac{\partial}{\partial u_1} \left[ \frac{1}{h_1 h_2} \frac{\partial(h_1 F_n)}{\partial u_2} \right] - \frac{1}{h_2} \frac{\partial}{\partial u_2} \left[ \frac{1}{h_1 h_2} \frac{\partial(h_2 F_n)}{\partial u_1} \right].\quad (\text{D.38})$$

The calculations of the expressions for line, surface and volume elements are straightforward. The following are explicit expressions for the most important coordinate systems:

- *Cartesian coordinates* ( $u_1 = x$ ,  $u_2 = y$ ,  $u_3 = z$ ) represent the most elementary coordinate system. Since all metrical coefficients are equal to unity  $h_1 = h_2 = h_3 = 1$ , we find

$$\begin{aligned}\nabla_t \phi &= \frac{\partial \phi}{\partial x} \mathbf{i}_x + \frac{\partial \phi}{\partial y} \mathbf{i}_y, \\ \nabla_t^2 \phi &= \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} \\ \nabla_t \cdot \mathbf{F}_t &= \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y}\end{aligned}\quad (\text{D.39})$$

$$\nabla_t \times \mathbf{F}_t = \left( \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right) \mathbf{i}_z.$$

$$\Delta_t \mathbf{F}_t = (\nabla_t^2 F_x) \mathbf{i}_x + (\nabla_t^2 F_y) \mathbf{i}_y$$

- Cylindrical coordinates ( $u_1 = r$ ,  $u_2 = \varphi$ ,  $u_3 = z$ ) constitute the appropriate coordinate system for circular waveguides. They are related to the Cartesian coordinates via

$$x = r \cos \varphi \quad y = r \sin \varphi$$

and the metrical coefficients are

$$h_1 = 1 \quad h_2 = r \quad h_3 = 1.$$

$$\begin{aligned} \nabla_t \phi &= \frac{\partial \phi}{\partial r} \mathbf{i}_r + \frac{1}{r} \frac{\partial \phi}{\partial \varphi} \mathbf{i}_\varphi, \\ \nabla_t^2 \phi &= \frac{\partial^2 \phi}{\partial r^2} + \frac{1}{r} \frac{\partial \phi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \phi}{\partial \varphi^2} \\ \nabla_t \cdot \mathbf{F}_t &= \frac{1}{r} \frac{\partial(rF_r)}{\partial r} + \frac{1}{r} \frac{\partial F_\varphi}{\partial \varphi} \\ \nabla_t \times \mathbf{F}_t &= \frac{1}{r} \left( \frac{\partial(rF_\varphi)}{\partial r} - \frac{\partial F_r}{\partial \varphi} \right) \mathbf{i}_z. \\ \Delta_t \mathbf{F}_t &= \left( \nabla_t^2 F_r - \frac{F_r}{r^2} - \frac{2}{r^2} \frac{\partial F_\varphi}{\partial \varphi} \right) \mathbf{i}_r + \left( \nabla_t^2 F_\varphi - \frac{F_\varphi}{r^2} - \frac{2}{r^2} \frac{\partial F_r}{\partial \varphi} \right) \mathbf{i}_\varphi. \end{aligned} \quad (D.40)$$

- Elliptical coordinates ( $u_1 = \theta$ ,  $u_2 = \varphi$ ,  $u_3 = z$ ) constitute the appropriate coordinate system for elliptical waveguides. They are related to the Cartesian coordinates via

$$x = c \cosh \theta \cos \varphi \quad y = c \sinh \theta \sin \varphi$$

and the metrical coefficients are

$$h_1 = h_2 = h = c \sqrt{\frac{\cosh(2\theta) - \cos(2\varphi)}{2}}$$

$$\nabla_t \phi = \frac{1}{h} \frac{\partial \phi}{\partial \theta} \mathbf{i}_\theta + \frac{1}{h} \frac{\partial \phi}{\partial \varphi} \mathbf{i}_\varphi$$

$$\begin{aligned} \nabla_t^2 \phi &= \frac{1}{h^2} \frac{\partial^2 \phi}{\partial \theta^2} + \frac{1}{h^2} \frac{\partial^2 \phi}{\partial \varphi^2} \\ \nabla_t \cdot \mathbf{F}_t &= \frac{1}{h^2} \left[ \frac{\partial(hF_\theta)}{\partial \theta} + \frac{\partial(hF_\varphi)}{\partial \varphi} \right] \\ \nabla_t \times \mathbf{F}_t &= \frac{1}{h^2} \left[ \frac{\partial(hF_\varphi)}{\partial \theta} - \frac{\partial(hF_\theta)}{\partial \varphi} \right] \mathbf{i}_z \\ D &= -\frac{c^4}{4h^6} [\sinh^2(2\theta) + \sin^2(2\varphi)] \\ \Delta_{ND} F_n &= -\frac{c^2}{h^4} \left[ \sin^2(2\varphi) \frac{\partial F_n}{\partial \theta} - \sinh^2(2\theta) \frac{\partial F_n}{\partial \varphi} \right]. \end{aligned} \quad (D.41)$$

## D.2 LINEAR OPERATORS

We will now briefly discuss the mathematical background of the eigenvalue problems treated in Chapter 3. They will turn out to be of the type

$$\mathcal{H} |m\rangle = \epsilon_m |m\rangle,$$

where  $\mathcal{H}$  is a linear operator,  $\epsilon_m$  is the eigenvalue and  $|m\rangle$  is an eigenmode of  $\epsilon_m$ . This eigenvalue problem is similar to that of the stationary Schrödinger equation, and it becomes obvious that we can use the formalism of quantum mechanics.

This section summarizes the most important results of Hilbert space theory without any proof.

### Dirac Vectors

The quantum mechanical notation  $|\varphi\rangle$  defines a Dirac vector. Such vectors satisfy the fundamental axioms of vector space, i.e., *associativity*

$$|\varphi\rangle + |\psi + \chi\rangle = |\varphi + \psi\rangle + |\chi\rangle,$$

*commutativity*

$$|\varphi\rangle + |\psi\rangle = |\psi\rangle + |\varphi\rangle$$

and *distributivity*

$$c |\varphi + \psi\rangle = c |\varphi\rangle + c |\psi\rangle$$

where  $c$  is a complex number. Here  $|\varphi|$  denotes the complementary vector which also satisfies all of the above axioms. It is usual to call  $|\varphi|$  the bra vector and  $|\varphi\rangle$  the ket vector.

### Scalar Product

The (complex-valued) scalar product (*bra-c-ket*) is defined by

$$\langle \varphi | \psi \rangle = \langle \psi | \varphi \rangle^*.$$

The scalar product of a vector with itself is always real and non-negative

$$\langle \varphi | \varphi \rangle \geq 0$$

and only the scalar product of the null-vector with itself vanishes. Two “non-null” vectors are called orthogonal, if the equation

$$\langle \varphi | \psi \rangle = 0$$

holds.

### Base Functions

An arbitrary Dirac vector  $|\varphi\rangle$  can be expanded into any complete set of base functions  $|\varphi_j\rangle$ , i.e.,

$$|\varphi\rangle = \sum_j \langle \varphi_j | \varphi \rangle |\varphi_j\rangle \quad (\text{D.42})$$

The symbol  $\Sigma$  denotes a hybrid between a sum for a discrete set and an integral for a continuous set of base functions.<sup>3</sup> The scalar product  $\langle \varphi_j | \varphi \rangle$  describes the projection of the vector  $|\varphi\rangle$  onto the base vector  $|\varphi_j\rangle$ . The base functions are orthonormal in most cases, i.e.,

$$\langle \varphi_i | \varphi_j \rangle = \delta(i, j).$$

The symbol  $\delta(i, j)$  describes either Kronecker's  $\delta_{ij}$

$$\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{else} \end{cases}$$

for a discrete set, or Dirac's  $\delta$ -function for a continuous set of base functions.

### Dirac's $\delta$ -function

Dirac's  $\delta$ -function over a definition subspace  $V$  is a distribution which is defined via

$$\int_V d\bar{r} f(\bar{r}) \delta(\mathbf{r} - \bar{r}) = \begin{cases} f(\mathbf{r}) & \text{for } \mathbf{r} \in V \\ \frac{1}{2}f(\mathbf{r}) & \text{for } \mathbf{r} \in \partial V \\ 0 & \text{else.} \end{cases} \quad (\text{D.43})$$

3. Dirac vectors are elements of the rigged Hilbert space  $H$  which contains both discrete and continuous base functions, i.e., distributions.

In physical terms, a distribution  $\delta(\mathbf{r})$  is not a function but a density. It can be regarded as the mathematical limit of a family of functions  $\delta_\epsilon(\mathbf{r})$  which become increasingly concentrated at the origin ( $\mathbf{r} = 0$ ) for  $\epsilon \rightarrow 0$ , i.e.,

$$\int_V d\mathbf{r} f(\mathbf{r}) \delta(\mathbf{r}) = \lim_{\epsilon \rightarrow 0} \int_V d\mathbf{r} f(\mathbf{r}) \delta_\epsilon(\mathbf{r}).$$

Examples of one-dimensional realizations of this family of functions are box functions,

$$\delta_\epsilon(x) = \begin{cases} \frac{1}{\epsilon} & \text{for } |x| < \frac{\epsilon}{2} \\ 0 & \text{else;} \end{cases}$$

Lorentz's functions,

$$\delta_\epsilon(x) = \frac{1}{\pi} \frac{\epsilon}{x^2 + \epsilon^2};$$

Gauss's functions,

$$\delta_\epsilon(x) = \frac{1}{\sqrt{2\pi\epsilon}} \exp\left(-\frac{x^2}{2\epsilon^2}\right)$$

and the sinc functions,

$$\delta_\epsilon(x) = \frac{1}{\pi\epsilon} \operatorname{sinc}^2\left(\frac{x}{\epsilon}\right).$$

We will briefly list some properties of Dirac's  $\delta$ -function. *Dirac's  $\delta$ -function of another function* is given by

$$\delta(g(x)) = \sum_n \frac{1}{|g'(x_n)|} \delta(x - x_n). \quad (\text{D.44})$$

where  $g(x)$  is a function having  $n$  zeros  $x_n$  with  $g'(x_n) \neq 0$ . The *derivatives of Dirac's  $\delta$ -function* are

$$\int_{-\infty}^{+\infty} d\bar{x} f(\bar{x}) \delta^{(n)}(x - \bar{x}) = (-1)^n f^{(n)}(x). \quad (\text{D.45})$$

Its *Fourier representation* is given by

$$\delta(\mathbf{r}) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dq e^{i\mathbf{q}\cdot\mathbf{r}}. \quad (\text{D.46})$$

*Cauchy's principal value of Dirac's  $\delta$ -function* is

$$\lim_{\epsilon \rightarrow 0} \int_{-\infty}^{+\infty} dx \frac{f(x)}{x \pm i\epsilon} = \mathcal{P} \int_{-\infty}^{+\infty} dx \frac{f(x)}{x} \mp i\pi f(0). \quad (\text{D.47})$$

This formula is used to derive the Kramers-Kronig relations.

## Linear Operators

An operator

$$\mathcal{L} : D \rightarrow R$$

acting on Dirac vectors maps the definition area  $D$  onto a range  $R$ . Both are part of the rigged Hilbert space  $H$ . The momentum operator

$$\mathcal{Q} = -i\nabla,$$

for example, can be applied only to differentiable functions. It is important to keep in mind that the definition area  $D = D_1 \dots D_n$  of a compound operator expression cannot be larger than that of the individual operators  $D_i$ . In general, it will be smaller. The compound operator  $\mathcal{Q}^n$ , for example, restricts the definition area to the  $n$ -times differentiable functions, whereas the individual elements  $\mathcal{Q}$  can be applied to the larger area of differentiable functions. Nevertheless, even "small" definition areas  $D$  are often dense in the original space, i.e., every function in  $H$  can be approximated by a sequence of functions of  $D$ .

An operator  $\mathcal{L}$  is called linear if it satisfies the following two conditions

$$\begin{aligned}\mathcal{L} |\varphi + \psi\rangle &= \mathcal{L} |\varphi\rangle + \mathcal{L} |\psi\rangle \\ \mathcal{L} |c\varphi\rangle &= c\mathcal{L} |\varphi\rangle.\end{aligned}$$

Two linear operators

$$\mathcal{L} + \mathcal{M} = M + \mathcal{L}$$

added to each other can be commuted. But the product of two operators does not in general satisfy the commutation rules. The commutator

$$[\mathcal{L}, \mathcal{M}] = \mathcal{L}\mathcal{M} - \mathcal{M}\mathcal{L}$$

describes the operator difference between both arrangements. Only the unity operator  $\mathcal{E}$  commutes with any other operator, i.e.,  $[\mathcal{L}, \mathcal{E}] = 0$ . The inverse operator is defined by the condition

$$\mathcal{L}\mathcal{L}^{-1} = \mathcal{L}^{-1}\mathcal{L} = \mathcal{E}$$

and an operator  $\mathcal{L}^\dagger$  is called adjoint to  $\mathcal{L}$ , if

$$\langle \mathcal{L}^\dagger \varphi | \psi \rangle = \langle \varphi | \mathcal{L} \psi \rangle$$

is satisfied.

The unity operator  $\mathcal{E}$  can be expanded into any complete set of base functions  $|\varphi_j\rangle$  by

$$\mathcal{E} = \sum_j |\varphi_j\rangle \langle \varphi_j|. \quad (\text{D.48})$$

Equation (D.48) represents a special formulation of Parseval's theorem. Putting such representations of unity at suitable locations into operator expressions represents an operator technique of great practical importance.

The eigenvalue problem of an operator  $\mathcal{H}$  is defined by

$$\mathcal{H} |\varphi_j\rangle = \epsilon_j |\varphi_j\rangle, \quad (\text{D.49})$$

where  $\epsilon_j$  is called the eigenvalue and  $|\varphi_j\rangle$ , the eigenvector or eigenmode of the operator  $\mathcal{H}$ . The spectrum of eigenvalues can be discrete and/or continuous. If  $n$  (linear independent) eigenvectors belong to one eigenvalue, this eigenvalue is designated as  $n$ -fold degenerated.

Obviously, any complete set of eigenvectors can be used as base functions.

The following small theorem is sometimes useful. *Two operators  $\mathcal{L}$  and  $\mathcal{M}$  which commute ( $[\mathcal{L}, \mathcal{M}] = 0$ ) have the same spectrum.*

## Hermitian and Essentially Self-Adjoint Operators

An operator  $\mathcal{L}$  is called Hermitian, if its definition area is dense in the rigged Hilbert space  $H$  and the condition

$$\langle \mathcal{L}\varphi | \psi \rangle = \langle \varphi | \mathcal{L}\psi \rangle$$

is satisfied, i.e., if an operator can be shifted through scalar products. If, in addition, the definition areas  $D(\mathcal{L}^\dagger) = D(\mathcal{L}^{\dagger\dagger})$  coincide, it is called essentially self-adjoint.

*Essentially self-adjoint operators provide a complete set of eigenvectors. The eigenvalues  $\epsilon_i$  are real, i.e.,*

$$\epsilon_i^* = \epsilon_i,$$

*and the eigenvectors corresponding to different eigenvalues are orthogonal. We can therefore always choose an orthonormal set of eigenvectors  $|\varphi_i\rangle$  such that*

$$\langle \varphi_i | \varphi_j \rangle = \delta(i, j).$$

*For any other operator  $\mathcal{L}$ , which provides a complete set of eigenvectors, we can construct an orthonormal set of base functions using the eigenvectors  $|\varphi_i\rangle$  of  $\mathcal{L}$  and the eigenvectors  $|\bar{\varphi}_i\rangle$  of the adjoint operator  $\mathcal{L}^\dagger$ , i.e.,*

$$\langle \bar{\varphi}_i | \varphi_j \rangle = \delta(i, j).$$

Examples are:

- *Position Operator  $\mathcal{R}$  and Momentum Operator  $\mathcal{Q}$ .* The corresponding eigenvalue problems are

$$\mathcal{R} |\mathbf{r}\rangle = \mathbf{r} |\mathbf{r}\rangle \quad (\text{D.50})$$

$$\mathcal{Q} |\mathbf{q}\rangle = \mathbf{q} |\mathbf{q}\rangle. \quad (\text{D.51})$$

The scalar products between eigenvectors of both operators in  $\mathbb{R}^n$  are plane waves, i.e.,

$$\langle \mathbf{r} | \mathbf{q} \rangle = (2\pi)^{-\frac{n}{2}} e^{i\mathbf{qr}}. \quad (\text{D.52})$$

We will now use the representation of the operators in the conventional space, i.e., with respect to the eigenvectors  $|r\rangle$ . Then, the position vector is simply  $\mathcal{R} = r$  and the momentum operator becomes

$$\mathcal{Q} = -i\nabla.$$

The scalar product becomes an integral in position space, i.e., the position operator  $\mathcal{R}$  can obviously be shifted through the scalar product, and the shift of the momentum operator  $\mathcal{Q}$  corresponds to a partial integration in which the factor  $-i$  keeps the operator Hermitian. The commutator between the position and the momentum operator is given by

$$[\mathcal{Q}, \mathcal{R}] = -i\mathcal{E}. \quad (\text{D.53})$$

- *Hamiltonians of the Form  $\mathcal{H} = \mathcal{Q}^2 - k_0^2\epsilon(\mathcal{R})$ .* These Hamiltonians  $\mathcal{H}$  are Hermitian for any Taylor expansion

$$\epsilon(\mathcal{R}) = \sum_n \epsilon_n \mathcal{R}^n$$

of the position operator with real coefficients  $\epsilon_n$ . The Schrödinger and Helmholtz operators describing the propagation of a TE mode through a lossless slab waveguide are Hamiltonians of this type.

### Unitary Operators

An operator  $\mathcal{U}$  defined on the complete rigged Hilbert space is called unitary if the condition

$$\mathcal{U}^{-1} = \mathcal{U}^\dagger$$

is satisfied. We can conclude that any product of unitary operators is unitary and that scalar products remain unchanged under unitary transformations, i.e.,

$$\langle \mathcal{U}\varphi | \mathcal{U}\psi \rangle = \langle \varphi | \psi \rangle.$$

As a further consequence, the eigenvalues  $\epsilon_j$  of an unitary operator  $\mathcal{U}$  are pure phase factors, i.e.,

$$|\epsilon_j| = 1.$$

As the analog transformation in a vector space, the unitary transformation in the rigged Hilbert space  $H$  conserves angles and lengths. Examples are:

- *Unitary Transformations of Hermitian Operators.* It can easily be proven that the operator

$$\mathcal{H}' = \mathcal{U}\mathcal{H}\mathcal{U}^\dagger$$

is Hermitian, if  $\mathcal{H}$  is Hermitian and  $\mathcal{U}$  is unitary. The eigenvalues  $\epsilon_j$  of the Hamiltonian  $\mathcal{H}$  remain unchanged under the unitary transformation, the eigenvectors  $|\varphi_j\rangle$  are transformed by the unitary transformation into a new set of orthogonal eigenvectors. The transformed eigenvalue problem is given by

$$\mathcal{H}' |\mathcal{U}\varphi_j\rangle = \epsilon_j |\mathcal{U}\varphi_j\rangle.$$

- *Unitary Group.* If  $\mathcal{H}$  is a self-adjoint operator and  $t$  is a real number, then the operators

$$\mathcal{U}(t) = e^{i\mathcal{H}t} \quad (\text{D.54})$$

form a continuous group.<sup>4</sup> The unitary operator  $\mathcal{U}$  is called a propagator, since it is a solution of the initial value problem

$$\mathcal{U} = i\mathcal{H}\mathcal{U} \quad (\text{D.55})$$

with the initial value  $\mathcal{U}(0) = \mathcal{E}$ . We saw in Chapter 4 of this book that the initial value problem for beam propagation is a more complicated variant of this problem.

- *Fourier Transformation.* The Fourier transformation

$$\mathcal{F}f = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} d\mathbf{r} e^{-i\mathbf{qr}} f(\mathbf{r}) \quad (\text{D.56})$$

is a unitary operator. Here  $\mathbf{r}$  and  $\mathbf{q}$  are elements of the vector space  $\mathbb{R}^n$ . Incidentally here are some more properties of the Fourier transformation.

$$(\mathcal{F}^{-1}f)(\mathbf{r}) = \mathcal{F}f(-\mathbf{r}) \quad (\text{D.57})$$

$$\mathcal{F}^4 = \mathcal{E} \quad (\text{D.58})$$

$$\mathcal{F}Qf = -\mathbf{q}\mathcal{F}f \quad (\text{D.59})$$

The convolution of two functions  $f(\mathbf{r})$  and  $g(\mathbf{r})$  is defined by

$$g \otimes f = (2\pi)^{-\frac{n}{2}} \int_{\mathbb{R}^n} d\bar{\mathbf{r}} g(\mathbf{r} - \bar{\mathbf{r}}) f(\bar{\mathbf{r}}). \quad (\text{D.60})$$

The Fourier transformation of a convolution is

$$\mathcal{F}(f \otimes g) = (\mathcal{F}f)(\mathcal{F}g). \quad (\text{D.61})$$

4. It can be proven that the operator  $\mathcal{U}$  is defined on the complete rigged Hilbert space even if  $\mathcal{H}$  is defined only on a part of it.

## REFERENCES

- [1] Courant, R., and Hilbert, D., *Methods of Mathematical Physics*, English ed., New York, Interscience Publishers, 1953.
- [2] Morse, P. M., and Feshbach, H., *Methods of Theoretical Physics*, New York, McGraw-Hill, 1953.
- [3] Gradshteyn, I. S., and Ryzhik, I. M., *Tables of Integrals, Series and Products*, English ed., New York, Academic Press, 1965.
- [4] Abramowitz, M., and Stegun, I. A., *Handbook of Mathematical Functions*, New York, Dover, 1965.

## About the Author

**Reinhard März** was born in Milan, Italy, in 1956. He received his diploma and Ph.D. in theoretical physics from the University of Frankfurt, in 1980 and 1983, respectively.

In 1984, he joined the Research Laboratories of Siemens AG, in Munich. Since that time he has been working on integrated optics within the InGaAsP/InP and glass (ion exchange and SiO<sub>2</sub>/Si) material systems. His activities have focussed on the design of many passive components, including Bragg gratings, various types of co- and contradirectional coupler, meander couplers, fiber-to-chip couplings, focusing reflection gratings and phased arrays. He has contributed to an enhanced beam propagation method which is based on a *z*-transient variational principle. He designed the first CAD system for integrated optics (SIGRAPH-OPTIK) and a simple model for the yield and cost analysis of integrated optical chips.

Dr. März is a member of the Deutsche Physikalische Gesellschaft (DPG) and of the advisory board of Fiber and Integrated Optics.

## Index

- Abbe number, 155
- ABCD* matrix, 148
- Aberrations
  - chromatic, 155–156
  - coma, 258
  - of spectrographs, 258
  - spherical, 152–154, 258
- Absorption, 11
- Airy's summation, 25, 120
- Angular spectrum, 103
- Bar state, 185, 189
- Beam parameter, 132
  - transformation rules, 143–147
- Beam propagation
  - adaptive finite elements, 119
  - benchmark results, 125–128
  - bidirectional, 120–123
  - boundary conditions, 95–97
  - classical algorithm, 98–109
  - classical propagator, 101–103, 109
  - eigenmode based, 112–113
  - finite differences (FD), 113–117
  - finite elements (FE), 117–120
  - free space, 98–101
  - Hermite-Gaussian mode, 134
  - mathematical assessment, 92–94
  - optimized paraxial approximations, 94
  - paraxial propagator, 99
  - prototype software, 297–298
  - reference dielectric constant, 91, 94
  - reflection operator, 122
  - semi-circle propagator, 99
  - software, 292–295
  - square-root propagator, 91
  - vectorial, 120
  - wide-angle approximations, 92
- z-transient variational principle, 117–120
- Beam radius, 135
- Beam waist, 133
- Blackman window, 197
- Blaze point, 264, 278
- Bohr-Sommerfeld quantization, 83
- Boundary conditions
  - absorbing, 96
  - dielectric, 18–21
  - Dirichlet, 95, 116
  - for beam propagation, 95–97
  - metallic, 95
  - Neumann, 95, 116
  - periodic, 96
  - transparent, 97
- BPM, *see* beam propagation
- Bragg condition, 30
- Bragg grating, 231–247
  - Brewster's angle, 238
  - compound, 242–244
  - conservation law, 232
  - curved, 244–247
  - filter characteristic, 233
  - Goos-Hänchen shift, 239
  - oblique incidence, 235–242
  - polarization conversion, 238
  - propagator, 243
  - quarter-wave shifted, 243
  - stopband, 233
  - TE-TM coupling, 238
- Brewster's angle, 24
- Bragg grating, 238
- Brillouin waves, 82
- Cauchy's principal value, 323
- Caustic, 36–37
- Characteristic admittance, 22, 28
- Characteristic equation, 75
- Characteristic impedance, 23

- Characteristic matrix, 27, 75  
 Charge density, 10  
 Chebyshev polynomials, 30, 210  
 Chirped coupling, 229  
 Commutation rule, 66  
 Computer aided design (CAD), 291  
 Confocal length, 135  
 Contradirectional coupling  
     chirped and tapered, 229  
     dispersion, 227–229  
     filter characteristic, 227–229  
     Goos-Hänchen shift, 239  
     polarization conversion, 238, 247  
     radiation, 224–226  
     TE-TM coupling, 238, 247  
 Convolution, 110  
 Coordinates, 315–321  
     cartesian, 319  
     cylindrical, 320  
     differential operator, 313, 318  
     elliptical, 320  
     general cylindrical, 317–321  
     metrical tensor, 316  
     transverse, 40, 90, 317  
     transverse differential operator, 318  
     unity vector, 313  
     vector operator, 313  
 Cost modeling, 301–311  
 Coupled mode theory  
     codirectional, 169–181  
     conservation law, 173, 222, 230  
     contradirectional, 220–231  
     coupling coefficients, 175–178, 193–195, 223  
     4-wave, 238, 247  
     mathematical assessment, 230–231  
     phenomenological, 173–175, 220–231  
     propagator, 174  
     tapered coupling, 178–181  
     transfer matrix, 175, 181, 183  
 Coupled waveguide array, 169, 220  
 Coupler  
     asymmetrical, 185, 191–197  
      $\Delta\beta$ -, 211–212, 298  
     contradirectional, 247–250  
      $\Delta\kappa$ -, 212–216  
     Mach-Zehnder, 207–209  
     symmetrical, 188–191  
     uniform, 181–188  
     with periodic overlay, 209–216  
 Cross state, 185, 189  
 Crossing, 199  
 Crosstalk, 283  
 Crosstalk attenuation, 267  
 Current density, 10  
 Cusp  
     caustic, 37, 154  
 Cutoff, 42  
 Cutoff condition, 84  
 Debye's expansion, 34  
 Detuning, 182, 191  
     effective, 192, 215  
 Device modeling, 290  
 Dielectric constant, 10  
     effective, 41  
     reference, 91, 94  
 Diffraction, 98–101  
     angle, 135  
     Fraunhofer integral, 110–111  
     Fresnel integral, 110–111  
 Dirac vector, 321  
     base functions, 322  
 Dirac's  $\delta$ -function, 322–323  
 Dispersion, 54–55  
     asymmetrical coupler, 192  
     contradirectional coupler, 249  
     contradirectional coupling, 227–229  
     intermodal, 54  
     material, 11, 54  
     waveguide, 54  
 Dispersion relation, 15  
 Divergence theorem, 314  
 Dynamic system, 173  
 Effective index method (EIM), 80–82  
 Eigenmode  
     closed solutions, 65–73  
     coupled waveguide array, 170–172  
     evanescent radiation, 42  
     guided, 42  
     Hermite-Gaussian, 69–70  
     Laguerre-Gaussian, 70–73  
     orthogonality of, 50–53  
     perturbation theory of, 58  
     propagating radiation, 42  
     propagation constant of, 41  
     transverse, 48  
     transverse electric (TE), 64, 65, 73, 75, 79,  
         80, 83  
     transverse magnetic (TM), 64, 75, 81, 83  
 Eigenvalue problem  
      $E_z/H_z$  formulation of, 47–48  
 mathematical background, 324  
 formulations, 45–50  
 Marcatili's, 79–80  
 separable, 77–82  
 vector E-field, 46–47  
 vector H-field, 46–47  
 Eikonal equation, 32–34  
 Electric energy, 11  
 Error function, 154  
 Euler-Lagrange equation, 35  
 Fermat's principle, 36, 256  
 Fiber  
     circular parabolic, 69–73  
     elliptical, 55  
     single-mode, 44  
 Fiber-to-chip coupling  
     lensed, 157–163  
 Filter characteristic  
      $\lambda/4$ -shifted Bragg grating, 244  
     asymmetrical coupler, 191–193  
     Bragg grating, 233  
     contradirectional coupling, 227–229  
     symmetrical coupler, 189  
 Finesse, 27  
 Focal length, 145, 148, 149  
 Forward Helmholtz equations, 89–92  
 Fourier transform, 327  
     Dirac's  $\delta$ -function, 323  
     fast, 107  
     Hermite-Gaussian mode, 134  
     Maxwell's equations, 10  
     numerical, 105–109, 297  
     sampling theorem, 107  
 Fresnel  
     diffraction integral, 92, 110–111  
     formulas, 23, 35  
 Full-width half-maximum (FWHM), 193  
 GaAlAs/GaAs, 4, 42, 285  
 Gaussian beam, 65, 267  
     coupling, 136–143  
     defocusing, 140  
     elliptical, 139  
     properties, *see* Hermite-Gaussian beam  
     tilted beam axes, 142  
     transverse misalignment, 140  
 Gaussian optics, 143–156  
     compound systems, 146  
     dielectric interface, 145  
     graded-index lens (GRIN), 146  
 spacer, 145  
 thin lens, 145  
 Gaussian units, 9  
 Geometrical optics, 32–37  
 Glass materials, 4, 42, 275  
 Goos-Hänchen shift, 24  
     Bragg grating, 239  
 Grating  
     reflection, 269–276  
     stigmatic, 270–274  
     theory, *see* spectrograph  
 Grating line, 252  
 Green's theorem, 314  
 Group delay, 54  
 Group velocity, 54  
 Hamilton  
     canonical equations, 35  
     principle, 35  
 Hamilton-Jacobi equation, 34  
 Hamiltonian, 89  
 Hamming window, 197  
 Helmholtz equations, 12–15  
     backward, 90  
     eigenvalue problems, 45–50  
     forward, 89–92  
     paraxial, 92  
     scalar, 60  
     vectorial, 120  
 Hermite polynomial, 67  
 Hermite-Gaussian beam, 131–136  
     ABCD matrix, 148  
     beam parameter, 132, 143–147  
     beam radius, 135  
     confocal length, 135  
     diffraction angle, 135  
     propagation, 134  
 Hermite-Gaussian mode, 69–70  
 Hilbert space, 321  
 Huygen's principle, 111, 252  
 InGaAsP/InP, 3, 42, 167, 273, 285, 302, 307  
 Insertion loss, 283  
 Interference filter, 27–32  
 Interferometer  
     Fabry-Perot, 25–27  
     Mach-Zehnder, 205–206  
 Kaiser window, 197  
 Kramers-Kronig relations, 11  
 Kronecker's delta, 313

- $\kappa$ -tapering, 197
- Laguerre polynominal, 72
- Laguerre-Gaussian mode, 70–73
- Laplacian, scalar and vectorial, 45, 316
- Leaky mode, 42
- Lens
- Abbe number, 155
  - ball, 158
  - fiber-to-chip coupling by, 157–163
  - focal length, 145, 149
  - graded-index (GRIN), 146, 160
  - plano-convex, 158
  - thick, 149
  - thin, 145
- Lens formula, 145, 146
- Light-path function, 252–255, 261, 278, 299
- $\text{LiNbO}_3$ , 5, 42
- Linear operator, 321–327
- commutator, 324
  - definition area, 323
  - eigenvalue problem, 324
  - Hermitian, 325–326
  - propagator, 326
  - scalar product, 321
  - self-adjoint, 325–326
  - unitary, 326–327
  - unitary group, 326
  - unity, 324
- Local normal mode theory, 163–166, 179
- Local oscillator, 285
- Lorentzian curve, 193, 216
- Magnetic energy, 11
- Magnetic monopole, 12
- Magnetization density, 9
- Magnification
- angular, 148, 150
  - spot, 148, 150, 157
- Marcatili's approach, 79–80
- Maxwell's equations, 9–12
- time-harmonic, 10
- Metrical tensor, 316
- Microlens, *see* lens
- Momentum operator, 325
- Mounting
- Rowland, 258, 273–274, 279–280
  - stigmatic, 261, 270–274, 280
- Notation
- Laplacian, 45
- quantum-mechanical, 45
- Optical backplane, 287
- Optical filter, 285
- Optical frequency division multiplex (OFDM), 285
- Optical interconnect, 287
- Optical sensor, 287–288
- Optical switching, 211
- Optical system
- $ABCD$  matrix, 148–152
  - chromatic aberrations, 155–156
  - focal length, 148
  - imaging, 148–152
  - magnification, 157
  - Newton's law, 151
  - one-lens, 162–163
  - principal planes, 149
  - spherical aberrations, 152–154
  - two-lens, 162–163
- Optical transmission systems, 284–287
- Orthogonality relations, 51
- Overlap integral, 176, 177, 185
- Parseval's theorem, 102, 175
- Passband, 30
- Perturbation theory, 56–58
- Phase front, 135, 294
- Phase thickness, 25, 28, 189
- Phase velocity, 10, 54
- Phased array, 277–280
- blazed, 278
  - light-path function, 278
  - Rowland mounting, 279–280
  - stigmatic mounting, 280
  - theory, *see* spectrograph
- Plane wave, 15
- Plasmon, 265
- Polarization
- senkrecht* (s), 22
  - circular, 17
  - elliptical, 17, 55
  - linear, 17
  - parallel (p), 22
  - transverse electric (TE), 13, 90, 95, 98, 265
  - transverse magnetic (TM), 13, 90, 95, 265
- Polarization density, 9
- Polarization scrambling, 55
- Polymers, 5
- Position operator, 325
- Power flux, 11
- tracking, 294
- Poynting's theorem, 11
- Poynting's vector, 11, 22, 53, 222
- generalized, 50
- Principal plane, 149
- Principle of causality, 11
- Process modeling, 290
- Profile interchange format (PIF), 292
- Propagation constant
- effective, 41
  - free space, 10
- Raised cosine window, 197
- Rate equation, 302
- Ray equation, 34
- Rayleigh's criterion, 267
- Rayleigh-Schrödinger theory, 56–58
- Reflection
- dielectric interface, 21–25
  - total, 24, 82
- Reflection grating
- Rowland, 273–274
  - stigmatic, 270–273
  - theory, *see* spectrograph
- Refraction, 21–25
- Refractive index, 11
- effective, 41
- Riccati equation, 196, 229
- Ritz's method, 55–56
- Rotating wave approximation (RWA), 214, 222
- Rowland circle, 258
- Rowland-type mountings, 258
- Sampling theorem, 107
- Scalar Helmholtz equation, 13
- beam propagation, 89–92, 97–120
  - eigenvalue problems, 58–64
- Scalar product, 51, 61, 321
- Schrödinger equation, 60, 92
- $\text{SiO}_2/\text{Si}$ , 4, 42, 275
- Snell's law, 22, 35
- Software
- $\Delta\beta$ -coupler, 298–299
  - beam propagation method, 292–295
  - BPM prototype, 297–298
  - computer aided design (CAD), 291
  - interchange format, 292
  - rapid prototyping, 295–300
- Spectrograph
- aberration theorem, 258
  - blazed, 264
- chirped, 254, 261
- coma, 258, 274–276
- crosstalk attenuation, 267
- defocusing, 274–276
- diffraction order, 268
- effective propagation constant, 252
- efficiency, 262–266
- flat-field, 274–276
- focal curve, 256
- free spectral range, 268–269
- grating equation, 256
- grating line, 252
- groove function, 254
- imaging, 256–261
- light-path function, 252–255, 261, 278, 299
- magnification, 257
- Rayleigh's criterion, 267
- Rowland circle, 258
- Rowland-type mounting, 258
- spectral resolution, 266–268
- spherical aberrations, 258, 273–276
- stigmatization, 261–262
- Spherical wave, 16, 135
- Spurious solutions, 12
- Staircase operator, 66
- Stigmatic point, 261–262
- Stokes's theorem, 314
- Stopband, 30, 224, 233
- Supermode, 169
- Susceptibility, 10
- Taper, 166–168
- Tapered coupling, 178–181
- asymmetrical coupler, 195–197
  - contradirectional couplers, 229
  - symmetrical coupler, 189
- Transfer matrix
- codirectional coupling, 175
  - crossing, 199
  - directional coupler, 186–188
  - Mach-Zehnder devices, 205–209
  - network, 202–205
  - periodic coupler, 209
  - phase shifter, 199
  - tapered coupling, 181
  - theory, 197–205
  - uniform directional coupler, 183
  - waveguide, 199
  - Y-branch, 199–202
- Transmission systems, 284–287
- Transmittance, 23

Transverse differential operator, 318

Unitary group, 205, 230, 327  
 $(N, M)$ -unitary, 230  
 conservation laws, 230

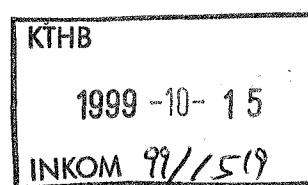
V parameter, 84, 172  
 Variational principle  
   Fermat's, 36, 256  
   Hamilton's, 35  
   Ritz's, 55–56  
   z-transient, 117–120  
 Vector correction formulas, 58, 61  
 Vector Helmholtz equation, 12  
   beam propagation, 89–92, 120  
   eigenvalue problems, 45–50

Wave  
   plane, 15  
   spherical, 16, 135  
 Waveguide  
   buried-rib, 44  
   fiber-matched, 44  
   general slab, 74–77  
   geometrical optics of the, 82–85  
   lossless, 50  
   nearly uniform, 59–62  
   parabolic slab, 65–69  
   power flux, 53–54  
   rib, 44  
   slab-like, 62–64  
   strip-loaded, 44  
   strongly guiding, 44  
   types of, 42–44  
   weakly guiding, 58–64

Wavelength  
   vacuum, 11  
 Wavelength division multiplex (WDM), 285  
 Wood's anomaly, 265

Y-branch, 199–202, 205, 206

Yield  
   array of identical devices, 307–310  
   continuous process, 303  
   critical process step, 304  
   extrapolation, 305–306  
   hidden defect, 304  
   modeling, 302–305  
   morphological defect, 304



## *The Artech House Optoelectronics Library*

Brian Culshaw, Alan Rogers, and Henry Taylor, *Series Editors*

- Acousto-Optic Signal Processing: Fundamentals and Applications*, Pankaj Das
- Amorphous and Microcrystalline Semiconductor Devices, Optoelectronic Devices*, Jerzy Kanicki, editor
- Bistabilities and Nonlinearities in Laser Diodes*, Hitoshi Kawaguchi
- Electro-Optical Systems Performance Modeling*, Gary Waldman and John Wootten
- The Fiber-Optic Gyroscope*, Hervé Lefèvre
- Field Theory of Acousto-Optic Signal Processing Devices*, Craig Scott
- Fundamentals of Multiaccess Optical Fiber Networks*, Denis J. G. Mestdagh
- Germanate Glasses: Structure, Spectroscopy, and Properties*, Alfred Margaryan and Michael A. Piliavin
- High-Power Optically Activated Solid-State Switches*, Arye Rosen and Fred Zutavern, editors
- Highly Coherent Semiconductor Lasers*, Motoichi Ohtsu
- Iddq Testing for CMOS VLSI*, Rochit Rajsuman
- Integrated Optics: Design and Modeling*, Reinhard März
- Introduction to Electro-Optical Imaging and Tracking Systems*, Khalil Seyrafi and S. A. Hovanessian
- Introduction to Glass Integrated Optics*, S. Iraj Najafi
- Introduction to Radiometry and Photometry*, William Ross McCluney
- Optical Control of Microwave Devices*, Rainee N. Simons
- Optical Document Security*, Rudolf L. van Renesse

*Optical Fiber Amplifiers: Design and System Applications*, Anders Bjarklev

*Optical Fiber Sensors, Volume I: Principles and Components*, John Dakin and Brian Culshaw, editors

*Optical Fiber Sensors, Volume II: Systems and Applications*, John Dakin and Brian Culshaw, editors

*Optical Interconnection: Foundations and Applications*, Christopher Tocci and H. John Caulfield

*Optical Network Theory*, Yitzhak Weissman

*Optical Transmission for the Subscriber Loop*, Norio Kashima

*Principles of Modern Optical Systems*, Volumes I and II, I. Andonovic and D. Uttamchandani, editors

*Reliability and Degradation of LEDs and Semiconductor Lasers*, Mitsuo Fukuda

*Semiconductor Raman Laser*, Ken Suto and Jun-ichi Nishizawa

*Semiconductors for Solar Cells*, Hans Joachim Möller

*Single-Mode Optical Fiber Measurements: Characterization and Sensing*, Giovanni Cancellieri

*For further information on these and other Artech House titles, contact:*

Artech House  
685 Canton Street  
Norwood, MA 02062  
617-769-9750  
Fax: 617-769-6334  
Telex: 951-659  
email: artech@world.std.com

Artech House  
Portland House, Stag Place  
London SW1E 5XA England  
+44 (0) 71-973-8077  
Fax: +44 (0) 71-630-0166  
Telex: 951-659  
email: bookco@artech.demon.co.uk