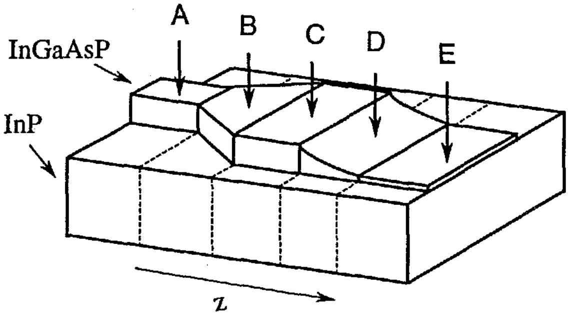
Stellingen behorende bij het proefschrift

Analysis, design and fabrication of tapered integrated optical structures

Implementation of the fully vectorial 3-D beam propagation method



Daoping Li

1. The propagating electric field in a longitudinally invariant, charge-free and dielectric waveguide structure has transverse components (E_x, E_y) that are independent of the longitudinal component (E_z); but the longitudinal component is dependent of the transverse components. Energy conservation is ensured through change of its magnetic field.

   This thesis, Eqs.(2.7-8).

2. Tensile strength of a material depends on the size and number of Griffith cracks on its surface. Without a coating layer and a buffer jacket, under the same environment conditions, the thinner an optical glass fiber is, the higher its tensile strength will be.

3. The discovery of quasi-crystals*, which have a fivefold rotation axis and thus contradict the traditional crystal theory**, shows that this theory has its preconditions.

   *D. Shechtman, I. Blech, D. Gratias and J.W. Cahn: *Phys. Rev. Lett.*, **53**, 1951-1953 (1984).
   **C. Kittel: *Introduction to Solid State Physics*, Sixth edition, John Wiley & Sons, Inc., 8 (1986).

4. Optical integrated circuits based upon InGaAsP/InP nowadays rely too much on the use of expensive techniques (MBE or MOVPE); with cheaper technology the development of such circuits would occur much faster.

5. Mistakes often occur in complicated formulas as appearing in scientific journals. Authors and referees should spend more time re-deriving in order to correct them all.

6. Correct use of words like *the* and *a* in English is so difficult for the majority of the world population that one has to doubt whether it is wise to introduce these two articles in a world language.

7. About one third of our whole life is used for sleep, which is an unbelievable waste of time. Investigations to decrease this would be important for the future development of mankind.

8. The names of Chinese dishes in the Netherlands are for a Chinese as unintelligible as for the Dutch.

9. The western standards of democracy and human rights can not be applied everywhere. Their acceptance depends at least on the local culture.

10. The recent decision of the French government to resume limited nuclear tests is being seriously criticized while previously the USA and the former USSR already did the same more than one thousand times. This implies that, if you want to do questionable things without being criticized, you should do it early.

11. The average length of the Dutch people keeps increasing, which is bad for the environment since tall people tend to use more O_2 and thus make more CO_2 which is the cause of the greenhouse effect.

Analysis, design and fabrication of tapered integrated optical structures

Implementation of the fully vectorial 3-D beam propagation method

The figure on the cover shows schematically the designed tapered waveguide structure. The whole structure is covered with InP or silicon nitride (not shown in the figure for convenience of view, see also Section 4.1)

Analysis, design and fabrication of tapered integrated optical structures

Implementation of the fully vectorial 3-D beam propagation method

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof.ir. K.F. Wakker,
in het openbaar te verdedigen ten overstaan van een commissie,
door het College van Dekanen aangewezen,
op dinsdag 19 september 1995 te 16:00 uur
door

Daoping LI

Master of Science,
geboren te Anhui, The People's Republic of China

Dit proefschrift is goedgekeurd door de promotor:
Prof.dr.ir. H.J. Frankena

Toegevoegd promotor:
Dr.ir. H. van Brug

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter
Prof.dr.ir. H.J. Frankena, TU Delft, promotor
Dr.ir. H. van Brug, TU Delft, toegevoegd promotor
Prof.dr.ir. H. Blok, TU Delft
Prof.dr.ir. J.J.M. Braat, TU Delft
Prof.dr. S.W. de Leeuw, TU Delft
Prof.dr. B.H. Verbeek, Philips Research
Prof.dr. W.Th. Wenckebach, TU Delft

Published and distributed by:

Delft University Press
Stevinweg 1
2628 CN Delft
The Netherlands
Telephone +31-15-783254
Fax +31-15-781661

CIP-DATA KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Li, Daoping

Analysis, design and fabrication of tapered integrated optical structures: implementation of the fully vectorial 3-D beam propagation method / Daoping Li. - Delft: Delft University Press. - Ill.
Thesis Delft University of Technology. - With index, ref. - With summary in Dutch.
ISBN 90-407-1154-2
NUGI 841
Subject headings: integrated optics / vectorial beam propagation method / fiber-chip coupling.

Printed in The Netherlands

To *Qiuxia*

Preface

This thesis is written to fulfill the requirements for the Ph.D. degree at Delft University of Technology, the Netherlands. It contains most of the results obtained within a four years (9/91 - 9/95) research period carried out in Optics Research Group, Dept. of Applied Physics under the guidance of Prof.dr. H.J. Frankena and Dr. H. van Brug. It reports the development of a fully vectorial finite difference beam propagation method for integrated optics and the analysis, design and fabrication of tapered integrated optical structures on InGaAsP/InP for fiber-chip coupling.

First, sincere thanks are given to Prof.dr. H.J. Frankena. As a busy professor, he has spent ample time to guide my research and revise this thesis. I am also deeply impressed by his character: responsible, direct, decisive and precise. My future career will certainly benefit from the serious training given by him. Eddy van Brug is gratefully acknowledged. He is more like a friend than a co-adviser, kind and patient. We had many fruitful discussions. Roland Horsten provided very good management of workstations and networks. He always solved problems immediately for me, even in weekends. Liang Shi (Dept. of Electrical Engineering) and Tom Scholtes took care of PECVD and proved to be reliable partners. Photospectrometer measurements were kindly performed by Ab Kuntze. Fokke Groen provided technological experience and useful advice. The technical support of Aad van der Lingen is also appreciated. Jos van der Tol and Jørgen Pedersen (PTT Research) provided new waveguide structures for simulation and co-authored two publications. The assistance of photolithography and atomic beam etching by Koos van Uffelen (Dept. of Electrical Engineering) is gratefully acknowledged. The kind help of Willem Vermin and Bert van Corler (SARA, Amsterdam) for using the super computer is appreciated. Also, thanks to Jin Song Gu (Institute of Quantum Electronics, ETH, Zürich, Switzerland) for the use of his FEM program. All other staff and students in the Optics Research Group are acknowledged for their cooperation and help. Finally, I am most grateful to my wife Qiuxia for her support, understanding and numerous lonely weekends and evenings.

Contents

Chapter 1

Introduction

1.1 Optical integrated circuits

An *Optical Integrated Circuit* (OIC) is based primarily on the fact that light waves can propagate through and be contained inside very thin layers (films) of transparent materials. Such a circuit is designed to fulfill certain functions and can, for example, consist of a laser diode as source, functional components such as switches/modulators, interconnection waveguides and photodiodes as detectors. Through integration, a more compact, stable, and functional optical system can be expected. If all components are integrated on a single substrate, it is called a *monolithic* OIC, otherwise, we speak of a *hybrid* OIC. Realization of a monolithic OIC is now possible in III-V semiconductor materials, for instance, the quaternary semiconductor InGaAsP epitaxially grown on an InP-substrate, which allows for the generation and detection of light at wavelengths between 1.3 and 1.55 μm, where glass fiber properties are optimal for long-distance communication [1]. On a hybrid OIC, as an example, the source, the waveguide and the detector can be made of compound semiconductors, dielectric materials and silicon, respectively.

Although the monolithic type OIC is the better solution as seen from several points of view, its implementation is still difficult at present. The separate parts of the hybrid type are relatively easy to fabricate, but there exist problems with assembling the components. Waveguide devices for constructing an OIC can further be classified into two categories: *passive devices*, which exhibit fixed characteristics for optical waves, and *active devices*, in which guided waves are controlled via electrooptic, acoustooptic, magnetooptic, thermooptic, or nonlinear optic effects.

Because the propagation velocity of an optical signal is much larger than that of an electric signal, it is often simply concluded that signal processing can operate much faster in OICs than that in their electric counterparts. Actually, however, the propagation velocity of an optical pulse in dielectric waveguides is only a few times larger than that of an electric pulse in an electric IC [2]. Therefore, the use of OICs is not only owing to its fast signal processing. Instead, its use mainly comes from utilizing the specific wave properties of optical signals, e.g. the much larger bandwidth. This fact is very important to understand the operation of OICs clearly.

Despite the fact that fundamental work has been performed even previously, the year 1969 is generally taken as the birth year of *integrated optics*, because then the term was firstly proposed by S.E. Miller of Bell Laboratories [3]. Since the very beginning, integrated optics has been motivated by its important role in broad band optical communication systems, where OICs can serve as high speed multiswitches, high speed modulators, wavelength (de)multiplexers, and time (de)multiplexers. Nowadays, 40 millions of kilometers of optical fibers are installed worldwide in the long and medium distance networks and the optical fiber is gradually penetrating the local networks [4]. The Japanese telecommunications company NTT envisages that by the year 2015 all houses will be connected by optical fiber [5]. In addition, OICs have other wide applications. For example, in signal processing they can serve as radio frequency spectrum analyzers, correlators, and convolvers or in sensor technology, as temperature sensors, position sensors, and optical disk signal pickup.

1.2 Fiber-chip coupling

Although many sophisticated electrooptical components for advanced optical communication systems have been demonstrated in recent years, comparatively little attention has been paid to the implementation of effective fiber-chip coupling techniques. The efficient, low cost and reliable coupling of guided waves from an optoelectronic chip to single mode fibers is an essential requirement for the application of these devices in future communication systems. To achieve this, a mode size transformation from the 1-3 μm diameter of the (usually elliptical) guided mode field in the chip to the 8-10 μm range for the (circular) guided mode field in the fiber is needed.

1.2.1 Coupling methods

There exist many kinds of coupling methods. We restrict our discussions to coupling between a single-mode fiber and an InGaAsP/InP chip. The following methods are generally used:

Butt-coupling: Butt-coupling is the most simple coupling method. The endfaces of the waveguide and the fiber are directly put into contact. However, an abruptly terminated (flat-end) single-mode fiber has a mode width of about 8-10 μm, whereas a normal integrated waveguide has a mode width of about 1-3 μm. This serious mismatch will lead to a low coupling efficiency, which is the biggest disadvantage of butt-coupling.

Fiber taper and microlens: A tapered fiber end combined with a microlens formed on its endface can reduce the mode size. Therefore, the coupling efficiency is increased.

The beamwaist of a Gaussian beam is that position where the beam has its smallest cross-section. For the abruptly terminated fiber, the beamwaist occurs exactly on the endfaces, but for the fiber with a microlens at the tip, there is a distance between the beamwaist and the endface. The coupling between a waveguide and such a fiber needs a separation for their beamwaists to occur in the same cross-section. This kind of coupling is called end-fire coupling. To compare with a simple butt-coupling configuration using a flat-end fiber, higher coupling efficiency can be expected since there is a better mode match, but a much more stringent alignment both in transverse and in longitudinal directions is needed. Transverse or lateral displacements in the order of 1 μm may cause around 1 dB additional loss.

The fabrication is normally carried out in two steps. First, a fiber taper is made by pulling it in an arc discharge [6], by etching in hydrofluoric acid [7] or by dipping the fiber end into negative photoresist while the fiber core carries the light beam from a HeNe laser [8]. Then the tip of the taper is rounded by melting, usually in an electric spark gap [9]. Surface tension, acting over the small dimensions involved, generates surfaces which are approximately spherical. However, it can also be fabricated in one step by laser micromachining [10].

***Tapered waveguide*:** Tapered waveguides on an InGaAsP/InP chip, serving as mode size transformers, have attracted increasing interest in recent years. Both laterally and vertically tapered waveguides have been demonstrated [11-18] and suggested [19-21]. Since the field of a waveguide mode on the chip can be transformed to match that of the fiber and the propagation loss through the tapered structure can be lower than 0.5 dB, the total coupling efficiency can be more than 80%. On the other hand, the alignment requirements are tolerant and can be technologically realized.

The fabrication of laterally tapered waveguide structures on InP/InGaAsP can be realized through a conventional lithography process. The production of vertically tapered structures, however, is relatively difficult. Selective wet chemical etching has been used to fabricate such structures. One approach involves the controlled dipping of the sample into an etchant, which is a fairly delicate process [22]. Another approach proceeds by consecutively removing individual layers of a diluted multiple-quantum-well (MQW) structure [11,17]. In addition, a combination of metal organic vapor phase epitaxy (MOVPE) and ion milling has been used [12]. Furthermore, a combination of MOVPE and wet chemical etching has been presented [23]. Beside this, the use of photo-resist techniques, oxide shadow and direct shadow etching are reported [24]. Also, a combination of MOVPE and a mask technique has been developed [25]. All these reported methods, however, are expensive.

There are several methods to realize a accurate alignment which is important for the coupling efficiency. A mechanical manipulator is a basic device to accomplish this which, however, has a relatively big size. A V-groove on the motherboard is another choice. One suggestion involves milling a groove in a glass substrate by a microdrill with a fine diamond paste as the polishing agent [26]. Another suggestion is to fabricate the V-groove on a silicon substrate by wet chemical etching [27]. The horizontal position of the V-groove with respect to the waveguide is defined by photolithography. The vertical position of the fiber core with respect to the waveguide is adjusted by the width of the V-groove. The precise alignment of the vertical position is obtained by etching the V-groove slightly wider than necessary, and covering it with an elastic layer of photoresist.

Taking into account both coupling efficiency and alignment tolerance, a tapered waveguide structure is the most promising candidate, but fabrication costs should be further decreased. Furthermore, for practical uses, the Si motherboard with V-grooves is recommended for the alignment, which has a simple configuration and a small size.

1.2.2 Coupling efficiency

Based upon the orthogonality of modes, the coupling efficiency between fields, propagating along the z- direction, of a single mode fiber and a waveguide can be derived as [28]:

$$\eta = \frac{\left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbf{E}_f \times \mathbf{H}_g^*\right)\cdot \mathbf{k}\,\mathrm{dx}\,\mathrm{dy} \right|^2}{\left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbf{E}_f \times \mathbf{H}_f^*\right)\cdot \mathbf{k}\,\mathrm{dx}\,\mathrm{dy} \right| \left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbf{E}_g \times \mathbf{H}_g^*\right)\cdot \mathbf{k}\,\mathrm{dx}\,\mathrm{dy} \right|}, \tag{1.1}$$

where **E**, **H** denote vectorial electric and magnetic fields, respectively, and **k** denotes the unit vector in z- direction, while the subscripts *f*, *g* denote quantities in the fiber and the waveguide, respectively. In fact, Eq.(1.1) is difficult to evaluate. Usually, a good approximation only containing electric fields, i.e.

$$\eta \cong \frac{\left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbf{E}_f \cdot \mathbf{E}_g^*\right)\mathrm{dx}\,\mathrm{dy} \right|^2}{\left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbf{E}_f \cdot \mathbf{E}_f^*\right)\mathrm{dx}\,\mathrm{dy} \right| \left| \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbf{E}_g \cdot \mathbf{E}_g^*\right)\mathrm{dx}\,\mathrm{dy} \right|}, \tag{1.2}$$

is widely used, which is known as the overlap integral [28]. To improve the coupling efficiency, the cross-section of the waveguide mode should match the fiber mode field as good as possible. Our comparison of coupling methods presented in literature with our own results will be based upon these expressions.

1.3 Simulation

As discussed in Section 1.1, the field of integrated optics concerns a wide range of phenomena involving light being guided along and controlled by thin dielectric films or strips. The wavelengths of interest lie mostly between 0.5 and 2 μm. Since the lateral waveguide sizes are usually in the order of the wavelength, a ray-optical treatment can not be fully adequate. Therefore, a rigorous electromagnetic theory is required, which usually corresponds to complicated mathematics.

The availability of advanced and accurate simulation techniques, which can provide a fast and reliable prediction of the field distributions or propagating behavior for designers, is crucial to any further progress. Among many simulation techniques for integrated optics, the beam propagation method (BPM) is a powerful tool to simulate wave propagation in waveguide structures. In principle, propagating fields in a waveguide structure can be calculated at any cross-section by the BPM, thus enabling scientists to evaluate and design new waveguide structures with the help of computers. From the very beginning when the initial BPM scheme was developed, the BPM has been widely accepted as a powerful tool. Numerous structures have been designed, simulated and subsequently fabricated, e.g. couplers, splitters and converters.

The BPM scheme was firstly developed by Fleck *et al* in 1976 [29]. In the original BPM, scalar wave equations are solved as an initial value problem, propagating step by step along the longitudinal direction by use of the fast Fourier transform (FFT) [29-32]. Its theoretical background is summarized in Appendix A. In addition, scalar versions of the BPM based upon the finite difference scheme (FD-BPM) have also been presented by other investigators [33-40]. Furthermore, the iterative Lanczos reduction method [41-42] has been developed to solve the scalar beam propagation problems based upon dominant eigenvalues and eigenvectors of the discretized lateral Laplace operator in the Helmholtz equation. However, scalar versions of the BPM in which the vectorial nature of electromagnetic waves is ignored are not fully adequate for polarization sensitive guided-wave devices. On the other hand, a semi-vectorial FD-BPM [43] based upon a series expansion of the Helmholtz propagation operator has been developed. Recently, fully vectorial versions of the BPM have also been presented [44-50]. There, the field in some straight waveguide structures with constant cross sections could be simulated numerically, under the assumption that the transverse field components are independent of the longitudinal coordinate.

1.4 Scope of this research

The final purpose of this research is the optimization of the fiber-chip coupling, where the fiber is a single mode fiber (SMF) and the chip is a III-V semiconductor InP/InGaAsP chip. This includes extensive calculations at possible structures, complicated processes to produce them and extensive measurements to assess the optimal coupling.

Although fiber microlenses generate certain effects that improve the coupling efficiency, they also lead to stringent alignment requirements. Tapered waveguides on chips serving as mode size transformers, however, can be directly butt-coupled to a flat-end fiber. In such configurations, a high coupling efficiency of more than 80% may be realized. Therefore, they are promising choices for an efficient fiber-chip coupling between a single mode fiber and an InGaAsP/InP chip, although they are still difficult and expensive to realize.

To design and fabricate tapered waveguide structures, wave propagation in the structures must be accurately simulated. Therefore, we first have to develop a fully vectorial FD-BPM for three-dimensional waveguide structures. The theoretical derivation is presented in Chapter 2. Then in Chapter 3, we discuss the program, show benchmark test results, and investigate influences of parameters on simulation results with our BPM by simulating the wave propagation through various waveguide structures. Subsequently we use it to design 3D tapered waveguide structures on InGaAsP/InP chips for obtaining high coupling efficiency with a flat-end single mode fiber in Chapter 4. Then, we develop a practical fabrication process to realize the designed structures using a combination of atomic beam etching, photolithography and plasma enhanced chemical vapor deposition (PECVD) in Chapter 5.

1.5 References

[1] J.M. Senior: Optical fiber communications: principles and practice, Prentice Hall, second edition, 84-96 (1992).

[2] H. Nishihara, M. Haruna and T. Suhara: *Optical Integrated Circuit*, McGraw-Hill, 3 (1987).

[3] S.E. Miller: "Integrated optics: an introduction," *Bell Syst. Tech. J.*, **48** (7), 2059-2068 (1969).

[4] M.R. Amersfoort: *Phased-array wavelength demultiplexers and their integration with photodetectors*, Delft University Press, 7 (1994).

[5] T. Ikegami: "Towards 'photonics through the network' in Japan," in: *Proc. ECOC'92*, 661-662 (1992).

[6] G.D. Khoe, J. Poulissen and H.M. de Vrieze: "Efficient coupling of laser diodes to

tapered monomode fibers with high-index end," *Electron. Lett.*, **19**, 205-207 (1983).

[7] H. Kuwahara, M. Sasaki and N. Tokoyo: "Efficient coupling from semiconductor lasers into single-mode fibers with tapered hemispherical ends," *Appl. Opt.*, **19**, 2578-2583 (1980).

[8] P.D. Bear: "Microlenses for coupling single-mode fibers to single-mode thin-film waveguides," in: *Proc. SPIE*, **239**, 98-103 (1980).

[9] G. Eisenstein and D. Vitello: "Chemically etched conical microlenses for coupling single-mode lasers into single-mode fibers," *Appl. Opt.*, 3470-3474 (1982).

[10] H.M. Presby, A.F. Benner and C.A. Edwards: "Laser micromachining of efficient fiber microlenses," *Appl. Opt.*, **29**, 2692-2695 (1990).

[11] T.L. Koch, et al: "Tapered waveguide InGaAs/InGaAsP multiple-quantum-well lasers," *IEEE Photon. Technol. Lett.*, **2**, 88-90 (1990).

[12] G. Müller, B. Stegmüller, H. Westermeier and G. Wenger: "Tapered InP/InGaAsP waveguide structure for efficient fiber-chip coupling," *Electron. Lett.*, **27**, 1836-1838 (1991).

[13] G. Wenger, G. Müller, B. Sauer, D. Seeberger and M. Honsberg: "Highly efficient multi-fiber chip coupling with large alignment tolerances by integrated InGaAsP/InP spot size transformer," in: *Proc. ECOC'92*, 927-930 (1992).

[14] R. Zengerle et al: "Low loss fiber-chip-fiber butt coupling by buried nonlinearly tapered InP/InGaAsP waveguide," in: *Proc. ECOC'93*, **2**, 249-252 (1993).

[15] T. Brenner and H. Melchior: "Highly efficient fiber-waveguide coupling achieved by InGaAsP/InP integrated optical modeshape adapters," in: *Proc. ECOC'93*, **2**, 329-332 (1993).

[16] K. Kasaya, O. Mitomi, M. Naganuma, Y. Kondo and Y. Noguchi: "A simple laterally tapered waveguide for low-loss coupling to single-mode fibers," *IEEE Photon. Technol. Lett.*, **5**, 345-347 (1993).

[17] N. Yoshimoto, K. Kawano, H. Takeuchi, S. Kondo and Y. Naguchi: "Spot size convertors using InP/InAlAs multiquantum well waveguides for low loss single-mode fiber coupling," *Electron. Lett.*, **28**, 1610-1611 (1992).

[18] T. Schwander et al: "Simple and low-loss fiber-to-chip coupling by integrated field-matching waveguide in InP," *Electron. Lett.*, **29**, 326-328 (1993).

[19] E.A.J. Marcatili: "Dielectric tapers with curved axes and no loss," *IEEE J. Quantum Electron.*, **QE-21**, 307-314 (1985).

[20] J. Buus, W.J. Stewart, J. Haes, J. Willems and R.G. Baets: "Spot size expansion for laser-to-fiber coupling using an integrated multimode coupler," *J. Lightwave Technol.*, **11**, 582-588 (1993).

[21] Z. Weissman and A. Hardy: "2-D mode tapering via tapered channel waveguide segmentation," *Electron. Lett.*, **28**, 1514-1516 (1992).

[22] T. Brenner, W. Hunziker, M. Smit, M. Bachmann, G. Guekos and H. Melchior: "Vertical InP/InGaAsP tapers for low loss optical fiber waveguide coupling," *Electron. Lett.*, **28**, 2040-2041 (1992).

[23] T. Brenner and H. Melchior: "Local etch-rate control of masked InP/InGaAsP by diffusion-limited etching," *J. Electrochem. Soc.*, **141**, 1954-1958 (1994).

[24] G. Müller, G. Wenger, L. Stoll, H. Westermeier and D. Seeberger: "Fabrication techniques for vertically tapered InP/InGaAsP spot-size transformers with very low loss," in: *Proc. ECIO'93*, 1414-1415 (1993).

[25] I. Moerman, M. D'Hondt, W. Vanderbauwhede, G. Coudenys, J. Haes, P. De Dobbelaere, R. Baets, P. Van Daele and P. Demeester: "Monolithic integration of a spot size transformer with a planar buried heterostructure InGaAsP/InP-laser using shadow mask growth," *IEEE Photon. Technol. Lett.*, **6**, 888-890 (1994).

[26] P.S. Chung, and M.J. Millington: "Milled-grove method for fiber-to-waveguide couplers," *J. Lightwave Technol.*, **5**, 1721-1726 (1987).

[27] C.A. Armiento, M. Tabasky, C. Jagannath, T.W. Fitzgerald, C.L. Shieh, V. Barry, M. Rothman, A. Negri, P.O. Haugsjaa and H.F. Lockwood: "Passive coupling of InGaAsP/InP laser array and singlemode fibers using silicon waverboard," *Electron. Lett.*, **27**, 1109-1111 (1991).

[28] R. Syms and J. Cozens: *Optical guided waves and devices*, McGraw-Hill Book Company, 134-144 (1992).

[29] J.A. Fleck, Jr., J.R. Morris and M.D. Feit: "Time-dependent propagation of high-energy laser beams through the atmosphere," *Appl. Phys.*, **10**, 129-160 (1976).

[30] M.D. Feit and J.A. Fleck, Jr.: "Light propagation in graded-index optical fibers," *Appl. Opt.*, **17**, 3990-3998 (1978).

[31] C.Yeh, L. Casperson and B. Szein: "Propagation of truncated Gaussian beams in multimode fiber guides," *J. Opt. Soc. Am.*, **68**, 989-993 (1978).

[32] C.Yeh, W.P. Brown and B. Szein: "Multimode inhomogeneous fiber couplers," *Appl. Opt.*, **18**, 489-495 (1979).

[33] T.B. Koch, J.B. Davies, and D. Wickramasinghe, "Finite element/finite difference propagation algorithm for integrated optical devices," *Electron. Lett.*, **25**, 514-516 (1989).

[34] Y. Chuang and N. Dagli, "An assessment of finite difference beam propagation method," *IEEE J. Quantum Electron.*, **26**, 1335-1338 (1990).

[35] Y. Chuang and Dagli, "Analysis of z-invariant and z-variant semiconductor rib waveguides by explicit finite difference beam propagation with nonuniform mesh configuration," *IEEE J. Quantum Electron.*, **27**, 2296-2305 (1991).

[36] C.M. Kim and R.V. Ramaswamy, "Modeling of graded-index channel waveguides using nonuniform finite difference method," *J. Lightwave Technol.*, **7**, 1581-1589 (1989).

[37] B. Hermansson, D. Yevick, W. Bardyszewski, and M. Glasner, " The unitary of split-operator finite difference and finite-element methods: applications to longitudinally varying semiconductor rib waveguides," *J. Lightwave Technol.*, **8**, 1866-1873 (1990).

[38] Y. Chuang and N. Dagli, "Explicit finite difference beam propagation method: application to semiconductor rib waveguide Y-junction analysis," *Electron. Lett.*, **26**, 711-713 (1990).

[39] R. Scarmozzino and R.M. Osgood, Jr., "Comparison of finite-difference and Fourier-transform solutions of the parabolic wave equation with emphasis on integrated-optics applications," *J. Opt. Soc. Am. A*, **8**, 724-731 (1991).

[40] D. Yevick and B. Hermansson, "Efficient beam propagation techniques," *IEEE Quantum Electron.*, **26**, 109-112 (1990).

[41] R.P. Ratowsky and J.A. Fleck, Jr., "Accurate numerical solution of the Helmholtz equation by iterative Lanczos reduction," *Optics Lett.*, **16**, 787-789 (1991).

[42] R.P. Ratowsky, J.A. Fleck, Jr., and M.D. Feit, "Helmholtz beam propagation in rib waveguides and couplers by iterative Lanczos reduction," *J. Opt. Soc. Am. A*, **9**,

265-273 (1992).

[43] P.C. Lee, D. Schulz, and E. Voges, "Three-dimensional finite difference beam propagation algorithms for photonic devices," *J. Lightwave Technol.*, **10**, 1832-1838 (1992).

[44] R. Clauberg, and P. Von Allmen: "Vectorial beam-propagation method for integrated optics," *Electron. Lett.*, **27**, 654-655 (1991).

[45] Y. Chuang, N. Dagli, and L. Thylen: "Explicit finite difference vectorial beam propagation method," *Electron. Lett.*, **27**, 2119-2121 (1991).

[46] J.M. Liu and L. Gomelsky, "Vectorial beam propagation method," *J. Opt. Soc. Am. A*, **9**, 1574-1585 (1992).

[47] W.P. Huang, C.L. Xu, S.T. Chu, and S.K. Chaudhuri, "The finite-difference vector beam propagation method: analysis and assessment," *J. Lightwave Technol.*, **10**, 295-304 (1992).

[48] W.P. Huang, C.L. Xu, and S.K. Chaudhuri, "A finite-difference vector beam propagation method for three-dimensional waveguide structures," *IEEE Photon. Technol. Lett.*, **4**, 148-151 (1992).

[49] W.P. Huang, C.L. Xu, S.T. Chu, and S.K. Chaudhuri, "A vector beam propagation method for guided-wave optics," *IEEE Photon. Technol. Lett.*, **3**, 910-913 (1991).

[50] W.P. Huang, C.L. Xu, and S.K. Chaudhuri, "Application of the finite-difference vector beam propagation method to directional coupler devices," *IEEE J. Quantum Electron.*, **28**, 1527-1532 (1992).

Chapter 2

The fully vectorial beam propagation method

In this chapter, we present a fully vectorial BPM for three-dimensional waveguide structures including longitudinally varying structures. The vectorial formulation is presented in Section 2.1. Subsequently, the finite difference scheme is shown in Section 2.2. In Section 2.3, we solve the discretized differential equations by using the relaxation method and analyze the stability and dissipation, simultaneously, discussing possibilities to speed up the numerical simulation.

2.1 Vectorial formulation

The scalar FFT-BPM has been discussed as a simple introduction in Appendix A. If possible, expressions in this section will be linked to their counterpart there.

Starting point is the vectorial wave equation for the time-harmonic electric field in an inhomogeneous dielectric medium which is source free and isotropic:

$$\nabla \times \nabla \times \mathbf{E}(x,y,z) = k_0^2 n^2(x,y,z)\mathbf{E}(x,y,z), \tag{2.1}$$

where k_0 is the wave number in vacuum and n(x,y,z) is the local refractive index. Here, we introduced a Cartesian coordinate system x,y,z with its z-axis along the propagation direction. Therefore:

$$\nabla^2\mathbf{E}(x,y,z)+k_0^2n^2(x,y,z)\mathbf{E}(x,y,z)=\nabla\left(\nabla\cdot\mathbf{E}(x,y,z)\right). \tag{2.2}$$

In components, we obtain from this:

$$\begin{aligned}\nabla^2E_x(x,y,z)+k_0^2n^2(x,y,z)E_x(x,y,z)=-\frac{\partial}{\partial x}\Bigg(&E_x(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial x}\\&+E_y(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial y}+E_z(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial z}\Bigg),\end{aligned} \tag{2.3a}$$

$$\begin{aligned}\nabla^2E_y(x,y,z)+k_0^2n^2(x,y,z)E_y(x,y,z)=-\frac{\partial}{\partial y}\Bigg(&E_x(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial x}\\&+E_y(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial y}+E_z(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial z}\Bigg),\end{aligned} \tag{2.3b}$$

$$\begin{aligned}\nabla^2E_z(x,y,z)+k_0^2n^2(x,y,z)E_z(x,y,z)=-\frac{\partial}{\partial z}\Bigg(&E_x(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial x}\\&+E_y(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial y}+E_z(x,y,z)\frac{\partial\ln n^2(x,y,z)}{\partial z}\Bigg),\end{aligned} \tag{2.3c}$$

where $E_x(x,y,z)$, $E_y(x,y,z)$ and $E_z(x,y,z)$ are the three Cartesian components of the electric field strength. In the scalar FFT-BPM, there exists only one component and $\nabla(\nabla\cdot\mathbf{E})$ has been discarded, as shown in Eq.(A.2). Furthermore, we write

$$E_x(x,y,z)=F_x(x,y,z)\exp(i\beta z), \tag{2.4a}$$

$$E_y(x,y,z)=F_y(x,y,z)\exp(i\beta z), \tag{2.4b}$$

$$E_z(x,y,z)=F_z(x,y,z)\exp(i\beta z), \tag{2.4c}$$

where β is the reference propagation constant, while F_x, F_y and F_z are the three components of the complex wave amplitude. Eqs.(2.4) are analogous to (A.3). β must be chosen such that a slowly varying envelope approximation of the field is allowed, implying that

$$\left|\frac{\partial^2 F_p(x,y,z)}{\partial z^2}\right| \quad << \quad 2\beta\left|\frac{\partial F_p(x,y,z)}{\partial z}\right|, \tag{2.5}$$

which is similar to (A.5) and here F_p stands for each of the three components. Then, the following set of equations is obtained:

$$\begin{aligned} i\frac{\partial F_x(x,y,z)}{\partial z} = -\frac{1}{2\beta}\Bigg[& \frac{\partial^2 F_x(x,y,z)}{\partial x^2} + \frac{\partial^2 F_x(x,y,z)}{\partial y^2} + \left(k_0^2 n^2(x,y,z)\right. \\ & \left. -\beta^2\right) F_x(x,y,z) + \frac{\partial}{\partial x}\left(F_x(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial x}\right. \\ & \left. + F_y(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial y} + F_z(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial z}\right)\Bigg], \end{aligned} \tag{2.6a}$$

$$\begin{aligned} i\frac{\partial F_y(x,y,z)}{\partial z} = -\frac{1}{2\beta}\Bigg[& \frac{\partial^2 F_y(x,y,z)}{\partial x^2} + \frac{\partial^2 F_y(x,y,z)}{\partial y^2} + \left(k_0^2 n^2(x,y,z)\right. \\ & \left. -\beta^2\right) F_y(x,y,z) + \frac{\partial}{\partial y}\left(F_x(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial x}\right. \\ & \left. + F_y(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial y} + F_z(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial z}\right)\Bigg], \end{aligned} \tag{2.6b}$$

$$\begin{aligned} i\frac{\partial F_z(x,y,z)}{\partial z} = -\frac{1}{2\beta}\Bigg[& \frac{\partial^2 F_z(x,y,z)}{\partial x^2} + \frac{\partial^2 F_z(x,y,z)}{\partial y^2} + \left(k_0^2 n^2(x,y,z)\right. \\ & \left. -\beta^2\right) F_z(x,y,z) + \frac{\partial}{\partial z}\left(F_x(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial x}\right. \\ & \left. + F_y(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial y} + F_z(x,y,z)\frac{\partial \ln n^2(x,y,z)}{\partial z}\right)\Bigg]. \end{aligned} \tag{2.6c}$$

Eqs.(2.6) are equivalent to (A.6) or (A.7) and can be rewritten as:

$$\frac{\partial F_x(x,y,z)}{\partial z} = H_{xx}F_x(x,y,z)+H_{xy}F_y(x,y,z)+H_{xz}F_z(x,y,z), \tag{2.7a}$$

$$\frac{\partial F_y(x,y,z)}{\partial z} = H_{yx}F_x(x,y,z)+H_{yy}F_y(x,y,z)+H_{yz}F_z(x,y,z), \tag{2.7b}$$

$$\frac{\partial F_z(x,y,z)}{\partial z} = H_{zx}F_x(x,y,z)+H_{zy}F_y(x,y,z)+H_{zz}F_z(x,y,z), \tag{2.7c}$$

where the operators H_{xx}, ⋯, H_{zz} are defined by

$$H_{qr}F_r = \frac{i}{2\beta}\left[\frac{\partial}{\partial q}\left(\frac{2}{n}\frac{\partial n}{\partial r}F_r\right)\right], \tag{2.8a}$$

$$H_{pp}F_p = \frac{i}{2\beta}\left[\frac{\partial^2 F_p}{\partial x^2}+\frac{\partial^2 F_p}{\partial y^2}+\left(k_0^2 n^2-\beta^2\right)F_p+\frac{\partial}{\partial p}\left(\frac{2}{n}\frac{\partial n}{\partial p}F_p\right)\right], \tag{2.8b}$$

with p, q, r = x, y, z and q ≠ r. For a z-invariant waveguide, $H_{xz}F_z$ and $H_{yz}F_z$ vanish, whereas for a straight slab waveguide, $H_{xy}F_y$, $H_{xz}F_z$, $H_{yx}F_x$, $H_{yz}F_z$ and $H_{zy}F_y$ are zero. Contrary to the original scalar FFT-BPM, we'll apply a FD scheme to solve Eqs.(2.7).

2.2 Numerical scheme

2.2.1 Formulation

The waveguide structure is discretized by a grid division with mesh sizes Δx, Δy and Δz, sampling the function values at the intersection points. Let *H*, *J* and *L* are total numbers of nodes in x-, y- and z- directions, respectively, then x = (*h-1*)Δx, y = (*j-1*)Δy and z = (*l-1*)Δz, where *h* = 2, 3, 4, ⋯, *H-1*; *j* = 2, 3, 4, ⋯, *J-1*; *l* = 1, 2, 3, ⋯, *L*. A new matrix equation

$$\mathbf{M}\cdot\mathbf{X}=\mathbf{b}, \tag{2.9}$$

can be obtained from Eqs.(2.7) and (2.8) for one propagation step by use of the standard Crank-Nicholson finite difference scheme (see Appendix B), where **M** denotes a square matrix of known coefficients, **X** is the solution column vector (F_x, F_y and F_z) and **b** is a column vector containing known elements only. Eq.(2.9) can be solved following the simultaneous relaxation algorithm [1].

Compared with the known scalar methods, the vectorial method has numerical complications originating mainly from the coupling among field components. Based upon the standard Crank-Nicholson finite difference scheme, the matrix **M** in Eq.(2.9) for a three-dimensional waveguide structure will have the dimensions $[(H\text{-}2)\cdot(J\text{-}2)]$ by $[(H\text{-}2)\cdot(J\text{-}2)]$ in a scalar method, while it has the size of $[2\cdot(H\text{-}2)\cdot(J\text{-}2)]$ by $[2\cdot(H\text{-}2)\cdot(J\text{-}2)]$ in the vectorial method for a structure with a constant cross section. Since the simulation speed of the relaxation method is inversely proportional to H^3 (if $H = J$), a speed difference of a factor of 8 between a scalar method and the vectorial method for that simple case can be expected. On similar grounds, for a tapered three-dimensional waveguide structure the speeds differ by a factor of 27. It should be pointed out that there exist techniques to speed up the scalar BPM, such as the split-step procedure [2], leading to an even bigger speed difference between a scalar method and the vectorial method. Therefore, special steps for speeding up the vectorial BPM are required to arrive at a useful method. The most important step is carried out by introducing the *Modified Crank-Nicholson scheme* [3].

An *explicit* finite difference scheme means that field components in step $l+1$ can be calculated explicitly from the known quantities in the previous step l, while an *implicit* scheme means that implicit equations containing field components in step $l+1$ have to be solved. An explicit scheme is unstable, in the sense that it yields exponentially growing results after many propagation steps, while an implicit scheme is stable but leads to a dissipation in the calculated wave propagation resulting from the numerical method [1]. A standard Crank-Nicholson scheme, which is stable and is the least dissipative, uses the average of the unknowns from the explicit and implicit schemes. However, we apply a modified Crank-Nicholson scheme to Eqs.(2.7), using only the explicit scheme to the coupling items, while ensuring the stability by introducing a finite difference scheme parameter α ($\alpha > 0.5$). Thus we arrive at

$$\begin{aligned}\frac{F_x^{l+1}(h,j)-F_x^{l}(h,j)}{\Delta z} &= \alpha\, \boldsymbol{FD}\left[H_{xx}F_x^{l+1}(h,j)\right]+(1-\alpha)\boldsymbol{FD}\left[H_{xx}F_x^{l}(h,j)\right] \\ &\quad +\boldsymbol{FD}\left[H_{xy}F_y^{l}(h,j)\right]+\boldsymbol{FD}\left[H_{xz}F_z^{l}(h,j)\right],\end{aligned} \tag{2.10a}$$

$$\begin{aligned}\frac{F_y^{l+1}(h,j)-F_y^{l}(h,j)}{\Delta z} &= \alpha\,\boldsymbol{FD}\big[H_{yy}F_y^{l+1}(h,j)\big]+(1-\alpha)\boldsymbol{FD}\big[H_{yy}F_y^{l}(h,j)\big]\\ &\quad+\boldsymbol{FD}\big[H_{yx}F_x^{l}(h,j)\big]+\boldsymbol{FD}\big[H_{yz}F_z^{l}(h,j)\big],\end{aligned} \tag{2.10b}$$

$$\begin{aligned}\frac{F_z^{l+1}(h,j)-F_z^{l}(h,j)}{\Delta z} &= \alpha\,\boldsymbol{FD}\big[H_{zz}F_z^{l+1}(h,j)\big]+(1-\alpha)\boldsymbol{FD}\big[H_{zz}F_z^{l}(h,j)\big]\\ &\quad+\boldsymbol{FD}\Big[H_{zx}F_x^{l+\frac{1}{2}}(h,j)\Big]+\boldsymbol{FD}\Big[H_{zy}F_y^{l+\frac{1}{2}}(h,j)\Big],\end{aligned} \tag{2.10c}$$

where $h = 2, 3, 4, \cdots, H\text{-}1$; $j = 2, 3, 4, \cdots, J\text{-}1$; $l = 1, 2, 3, \cdots, L$. ***FD***[PP] means that PP has to be represented by an expression obtained through the finite difference operation, which will be shown in the following. Eq.(2.10c) contains the superscript $l+\frac{1}{2}$, which indicates the FD operation for half the step since the z- derivative exists for those quantities. As a result, each of the above three equations can be rewritten as $\mathbf{M}\cdot\mathbf{X} = \mathbf{b}$ where the coefficient matrices are sparse having a tridiagonal form with fringes. Thus, the vectorial BPM is converted to three mutually related scalar BPM schemes, resulting in a higher simulation speed.

We next show the results of the finite difference operation on Eq.(2.10a) at discretized nodes *h,j,l+1* and *h,j,l*. Based upon Eqs.(2.7a) and (2.8), we first have

$$\begin{aligned}\boldsymbol{FD}\big[H_{xx}F_x^{l+1}(h,j)\big]=\boldsymbol{FD}\Bigg[\frac{i}{2\beta}\Bigg\{&\frac{\partial^2 F_x}{\partial x^2}+\frac{\partial^2 F_x}{\partial y^2}+\left(k_0^2n^2-\beta^2\right)F_x\\ &+\frac{\partial}{\partial x}\left(\frac{2}{n}\frac{\partial n}{\partial x}F_x\right)\Bigg\}\Bigg]_{h,j}^{l+1},\end{aligned} \tag{2.11}$$

with

$$\boldsymbol{FD}\left[\frac{\partial^2 F_x}{\partial x^2}\right]_{h,j}^{l+1}=\frac{F_x^{l+1}(h+1,j)-2F_x^{l+1}(h,j)+F_x^{l+1}(h-1,j)}{\Delta x^2}, \tag{2.12a}$$

$$\boldsymbol{FD}\left[\frac{\partial^2 F_x}{\partial y^2}\right]_{h,j}^{l+1}=\frac{F_x^{l+1}(h,j+1)-2F_x^{l+1}(h,j)+F_x^{l+1}(h,j-1)}{\Delta y^2}, \tag{2.12b}$$

$$\boldsymbol{FD}\left[(k_0^2 n^2-\beta^2)F_x\right]_{h,j}^{l+1}=\left(k_0^2 n^2(h,j,l+1)-\beta^2\right)F_x^{l+1}(h,j), \tag{2.12c}$$

$$\begin{aligned}\boldsymbol{FD}\left[\frac{\partial}{\partial x}\left(\frac{2}{n}\frac{\partial n}{\partial x}F_x\right)\right]_{h,j}^{l+1}&=\frac{2[n(h+1,j,l+1)-n(h,j,l+1)]}{[n(h+1,j,l+1)+n(h,j,l+1)]\Delta x^2}\left[F_x^{l+1}(h+1,j)\right.\\&\left.+F_x^{l+1}(h,j)\right]+\frac{2[n(h-1,j,l+1)-n(h,j,l+1)]}{[n(h-1,j,l+1)+n(h,j,l+1)]\Delta x^2}\left[F_x^{l+1}(h-1,j)+F_x^{l+1}(h,j)\right].\end{aligned} \tag{2.12d}$$

For the expression

$$\begin{aligned}\boldsymbol{FD}\left[H_{xx}F_x^l(h,j)\right]=\boldsymbol{FD}\left[\frac{i}{2\beta}\left\{\frac{\partial^2F_x}{\partial x^2}+\frac{\partial^2F_x}{\partial y^2}+\left(k_0^2n^2-\beta^2\right)F_x\right.\right.\\\left.\left.+\frac{\partial}{\partial x}\left(\frac{2}{n}\frac{\partial n}{\partial x}F_x\right)\right\}\right]_{h,j}^{l},\end{aligned} \tag{2.13}$$

the same procedures as those mentioned above can be used, and therefore are omitted here. On the other hand, we have

$$\begin{aligned}\boldsymbol{FD}\left[H_{xy}F_y^l(h,j)\right]&=\boldsymbol{FD}\left[\frac{i}{2\beta}\frac{\partial}{\partial x}\left(\frac{2}{n}\frac{\partial n}{\partial y}F_y\right)\right]_{h,j}^{l}\\&=R_1\left[F_y^l(h+1,j)+F_y^l(h,j)\right]-R_2\left[F_y^l(h-1,j)+F_y^l(h,j)\right],\end{aligned} \tag{2.14}$$

with

$$R_1=\frac{i[n(h+1,j+1,l)+n(h,j+1,l)-n(h+1,j-1,l)-n(h,j-1,l)]}{4\beta\,\Delta x\,\Delta y[n(h+1,j,l)+n(h,j,l)]}, \tag{2.15a}$$

$$R_2=\frac{i[n(h-1,j+1,l)+n(h,j+1,l)-n(h-1,j-1,l)-n(h,j-1,l)]}{4\beta\,\Delta x\,\Delta y[n(h-1,j,l)+n(h,j,l)]}. \tag{2.15b}$$

In addition, we obtain

$$
\begin{aligned}
\boldsymbol{FD}\left[\mathrm{H}_{xz}\mathrm{F}_z^l(h,j)\right] &= \boldsymbol{FD}\left[\frac{i}{2\beta}\frac{\partial}{\partial \mathrm{x}}\left(\frac{2}{\mathrm{n}}\frac{\partial \mathrm{n}}{\partial \mathrm{z}}\mathrm{F}_z\right)\right]_{h,j}^{l} \\
&= \mathrm{S}_1\left[\mathrm{F}_z^l(h+1,j)+\mathrm{F}_z^l(h,j)\right]-\mathrm{S}_2\left[\mathrm{F}_z^l(h-1,j)+\mathrm{F}_z^l(h,j)\right],
\end{aligned}
\tag{2.16}
$$

with

$$
\mathrm{S}_1 = \frac{i\left[\mathrm{n}(h+1,j,l+1)+\mathrm{n}(h,j,l+1)-\mathrm{n}(h+1,j,l)-\mathrm{n}(h,j,l)\right]}{2\beta\,\Delta\mathrm{x}\,\Delta\mathrm{z}\left[\mathrm{n}(h+1,j,l)+\mathrm{n}(h,j,l)\right]},
\tag{2.17a}
$$

$$
\mathrm{S}_2 = \frac{i\left[\mathrm{n}(h-1,j,l+1)+\mathrm{n}(h,j,l+1)-\mathrm{n}(h,j,l)-\mathrm{n}(h-1,j,l)\right]}{2\beta\,\Delta\mathrm{x}\,\Delta\mathrm{z}\left[\mathrm{n}(h-1,j,l)+\mathrm{n}(h,j,l)\right]}.
\tag{2.17b}
$$

The above calculation is related to Eq.(2.10a), but Eqs.(2.10b-c) can be treated similarly. As a result, by substituting Eqs.(2.11-17) into Eqs.(2.10), we obtain

$$
\begin{aligned}
&\mathrm{A}_1^l(h,j)\mathrm{F}_x^{l+1}(h+1,j)+\mathrm{A}_2^l(h,j)\mathrm{F}_x^{l+1}(h-1,j)+\mathrm{A}_3^l(h,j)\mathrm{F}_x^{l+1}(h,j+1) \\
&+\mathrm{A}_4^l(h,j)\mathrm{F}_x^{l+1}(h,j-1)+\mathrm{A}_5^l(h,j)\mathrm{F}_x^{l+1}(h,j)= \\
&\qquad \mathrm{B}_1^l(h,j)\mathrm{F}_x^l(h+1,j)+\mathrm{B}_2^l(h,j)\mathrm{F}_x^l(h-1,j)+\mathrm{B}_3^l(h,j)\mathrm{F}_x^l(h,j+1) \\
&\qquad +\mathrm{B}_4^l(h,j)\mathrm{F}_x^l(h,j-1)+\mathrm{B}_5^l(h,j)\mathrm{F}_x^l(h,j)+\mathrm{B}_6^l(h,j)\mathrm{F}_y^l(h+1,j) \\
&\qquad +\mathrm{B}_7^l(h,j)\mathrm{F}_y^l(h-1,j)+\mathrm{B}_8^l(h,j)\mathrm{F}_y^l(h,j)+\mathrm{B}_9^l(h,j)\mathrm{F}_z^l(h+1,j) \\
&\qquad +\mathrm{B}_{10}^l(h,j)\mathrm{F}_z^l(h-1,j)+\mathrm{B}_{11}^l(h,j)\mathrm{F}_z^l(h,j),
\end{aligned}
\tag{2.18a}
$$

$$
\begin{aligned}
&\mathrm{C}_1^l(h,j)\mathrm{F}_y^{l+1}(h+1,j)+\mathrm{C}_2^l(h,j)\mathrm{F}_y^{l+1}(h-1,j)+\mathrm{C}_3^l(h,j)\mathrm{F}_y^{l+1}(h,j+1) \\
&+\mathrm{C}_4^l(h,j)\mathrm{F}_y^{l+1}(h,j-1)+\mathrm{C}_5^l(h,j)\mathrm{F}_y^{l+1}(h,j)= \\
&\qquad \mathrm{D}_1^l(h,j)\mathrm{F}_y^l(h+1,j)+\mathrm{D}_2^l(h,j)\mathrm{F}_y^l(h-1,j)+\mathrm{D}_3^l(h,j)\mathrm{F}_y^l(h,j+1) \\
&\qquad +\mathrm{D}_4^l(h,j)\mathrm{F}_x^l(h,j-1)+\mathrm{D}_5^l(h,j)\mathrm{F}_y^l(h,j)+\mathrm{D}_6^l(h,j)\mathrm{F}_x^l(h,j+1) \\
&\qquad +\mathrm{D}_7^l(h,j)\mathrm{F}_x^l(h,j-1)+\mathrm{D}_8^l(h,j)\mathrm{F}_x^l(h,j)+\mathrm{D}_9^l(h,j)\mathrm{F}_z^l(h,j+1) \\
&\qquad +\mathrm{D}_{10}^l(h,j)\mathrm{F}_z^l(h,j-1),+\mathrm{D}_{11}^l(h,j)\mathrm{F}_z^l(h,j),
\end{aligned}
\tag{2.18b}
$$

$$\begin{aligned}
&G_1^l(h,j)F_z^{l+1}(h+1,j)+G_2^l(h,j)F_z^{l+1}(h-1,j)+G_3^l(h,j)F_z^{l+1}(h,j+1)\\
&+G_4^l(h,j)F_z^{l+1}(h,j-1)+G_5^l(h,j)F_z^{l+1}(h,j)=\\
&\qquad H_1^l(h,j)F_z^l(h+1,j)+H_2^l(h,j)F_z^l(h-1,j)+H_3^l(h,j)F_z^l(h,j+1)\\
&\qquad +H_4^l(h,j)F_z^l(h,j-1)+H_5^l(h,j)F_z^l(h,j)+H_6^l(h,j)F_x^{l+1}(h,j)\\
&\qquad +H_7^l(h,j)F_x^l(h,j)+H_8^l(h,j)F_y^{l+1}(h,j)+H_9^l(h,j)F_y^l(h,j),
\end{aligned} \tag{2.18c}$$

where $A_1, \cdots, A_5$; $B_1, \cdots, B_{11}$; $C_1, \cdots, C_5$; $D_1, \cdots, D_{11}$; $G_1, \cdots, G_5$, and $H_1, \cdots, H_9$ are all complicated expressions, which have been arranged in Appendix C.

2.2.2 Stability analysis

The stability of a finite difference scheme is critical. Despite its lack of rigor, the von Neumann method for stability analysis [1] generally gives valid answers and is much easier to apply than more exact methods [4].

The von Neumann method acts locally: we suppose that the coefficients of a difference equation vary slowly such that they can be considered constant over the whole space. In that case, the independent solutions of the difference equation are all of the form

$$F^l(h,j)=\zeta^l e^{ik_x h\Delta x} e^{ik_y j\Delta y}, \tag{2.19}$$

where k_x and k_y are real spatial wave vector components while ζ is a complex number called the amplification factor.

It can be shown that there exists a critical value α_c ($\alpha_c > 0.5$), such that if $\alpha < \alpha_c$, the scheme is unstable ($|\zeta| > 1$), leading to an unrealistic increase of the field power propagating through subsequent cross-sections. If $\alpha > \alpha_c$, however, there is a numerically induced power dissipation ($|\zeta| < 1$). The optimal choice is $\alpha = \alpha_c$ ($|\zeta| = 1$) which corresponds to a stable propagation without numerical loss . Actually, for a longitudinally varying waveguide structure, the exact value of α_c is difficult to calculate and could slightly change along the propagation direction. We usually look for the best value obtained from numerical simulation. To ensure the stability, a slightly bigger value should be used, leading to an inevitable but small spurious power dissipation, which will be discussed in Section 3.3.1 where we consider waveguide structures and simulation results.

2.2.3 Influence of finite difference operation

Most of the InP/InGaAsP waveguide structures have stepped index changes, which means that the index distribution is discontinuous on some interfaces. Finite difference operation in the neighborhood of waveguide corners should be considered carefully, especially when the step sizes of the discretization are small and the index differences large. We studied the possible consequence of the finite difference operation on different but equivalent expressions which are functions of the local refractive index, especially the non-physical noise produced in the neighborhood of waveguide corners where the index is discontinuous.

In the vectorial formulation of Section 2.1, the first- and second-order derivatives of the refractive index with respect to x an y are non-vanishing. For example, the following expressions occur in Eqs.(2.3)

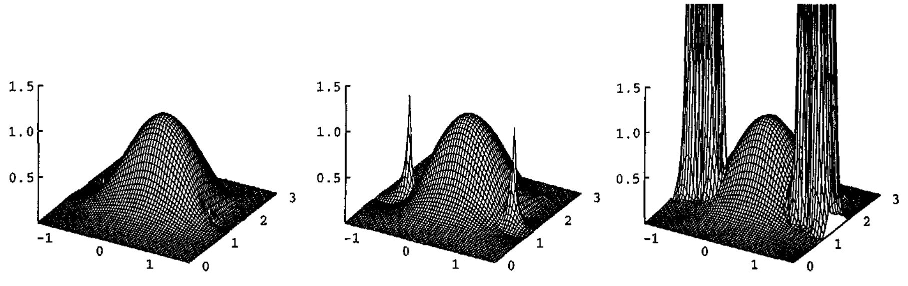
$$\frac{\partial}{\partial \mathrm{q}}\left(\mathrm{F}_{\mathrm{r}}^{l}(h,j) \frac{\partial \ln \mathrm{n}^2(h,j,l)}{\partial \mathrm{r}} \right) \qquad (\mathrm{q},\mathrm{r}=\mathrm{x},\mathrm{y}). \tag{2.20}$$

The finite difference operation in the node *h,j,l* will lead to the introduction of the index values at the following five nodes, i.e. (*h+1,j,l*), (*h-1,j,l*), (*h,j+1,l*), (*h,j-1,l*) and (*h,j,l*). We know that the refractive index distribution is discontinuous at waveguide interfaces and especially in corners. Therefore, the finite difference operation on the following equivalent expressions

$$\frac{2}{\mathrm{n}(h,j,l)}\frac{\partial \mathrm{n}(h,j,l)}{\partial \mathrm{x}} \equiv \frac{\partial \ln \mathrm{n}^2(h,j,l)}{\partial \mathrm{x}} \equiv \frac{1}{\mathrm{n}^2(h,j,l)}\frac{\partial \mathrm{n}^2(h,j,l)}{\partial \mathrm{x}}. \tag{2.21}$$

give different results.

We have investigated the influence of the three types of expressions in Eq.(2.21) on propagating fields. The HE_{00}-mode field was launched into the rib waveguide [5]. Figure 2.1 shows the E_y components after propagation over 2 μm. If the latter two expressions are used for the finite difference operation, both lead to the introduction of nonphysical noise in the neighborhood of the corners and this noise becomes bigger as the propagation proceeds. However, the first expression generates that problem on a much lower scale. On account of this analysis we suggest that, prior to the finite difference operation, the expressions containing the derivatives of the refractive index should be converted to their most simple form, i.e. no square, no logarithm.



(a) (b) (c)

Figure 2.1 Amplitude distribution of E_y for a HE_{00}-mode coupled into the rib waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ µm; $\alpha = 0.65$, $n_0 = 3.39$. The propagation distance is 2 µm.

2.3 Solving the linear algebraic equations

2.3.1 Boundary condition

A boundary as discussed here is the boundary of a computational window, not a boundary of the waveguide structure. There are many kinds of boundary conditions, among which the Dirichlet and the Neumann boundary conditions are widely used [1]. If the field strengths at the boundary are not small enough, a noticeable reflection of fields by the boundary will occur, which is non-physical and causes spurious interference.

To prevent boundary reflection, the most common way is the insertion of artificial absorption regions adjacent to the pertinent boundaries such that field strengths will be negligible at the boundary [6]. The thickness of these regions, their maximum absorption coefficient, and the spatial dependence of the absorption must all be carefully chosen for the method to work properly. In addition, extra absorption regions result in computational penalties concerning runtime and storage space.

A new transparent boundary condition [7-8] has been developed, which allows the radiation to escape through the boundaries without noticeable reflection, such that a radiative flux back into the computational window is prevented.

Since the treatment of the four boundaries of a computational window is essentially similar [7-8], we only focus on the right-hand boundary (see Figure 2.2). By use of the simultaneous relaxation algorithm, we obtain the field components of the propagation step $l+1$ in all nodes excluding those on the four boundaries from the field components of the propagation step l. The following implementation is used:

$$\frac{F_x^{l+1}(h, J)}{F_x^{l+1}(h, J-1)} = \frac{F_x^{l+1}(h, J-1)}{F_x^{l+1}(h, J-2)} = \exp(ik_r \Delta y), \tag{2.22}$$

where k_r is a complex propagation constant and $h = 2, 3, 4, \cdots, H-1$; $l = 1, 2, 3, \cdots$. It can be proven [7] that if the real part of k_r is positive, radiative energy can only flow out through the right-hand boundary. If the real part of k_r is negative, we use instead of (2.22):

$$F_x^{l+1}(h, J) = F_x^{l+1}(h, J-1), \tag{2.23}$$

thus neglecting a back flow of energy into the computational window. For field components in the y- and z- directions, a similar implementation is used. Such a treatment will ensure that radiated energy can only flow out of the considered boundary.

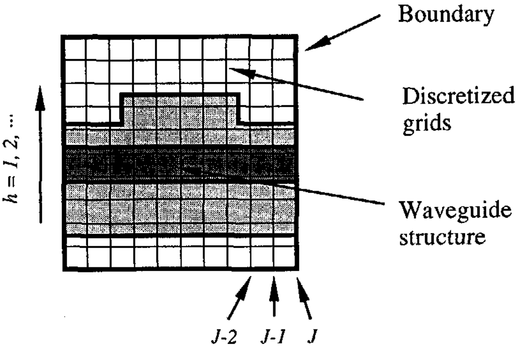


Figure 2.2 Schematic view of a discretized rib waveguide

2.3.2 Relaxation method

In (2.18), each equation has the following form with $h = 2, 3, 4, \cdots, H\text{-}1$ and $j = 2, 3, 4, \cdots, J\text{-}1$:

$$\begin{aligned}&\mathrm{a}(h,j)\mathrm{U}(h+1,j)+\mathrm{b}(h,j)\mathrm{U}(h-1,j)+\mathrm{c}(h,j)\mathrm{U}(h,j+1)\\&\quad+\mathrm{d}(h,j)\mathrm{U}(h,j-1)+\mathrm{e}(h,j)\mathrm{U}(h,j)=\mathrm{f}(h,j),\end{aligned} \tag{2.24}$$

where a(h, j), b(h, j), c(h, j), d(h, j), e(h, j) and f(h, j) are known coefficients. An iterative solution procedure is set up by solving (2.24) for U(h, j), which is first calculated from the initially estimated values $\mathrm{U}^{\mathrm{initial}}(h, j)$ via:

$$\begin{aligned}\mathrm{U}(h,j)=\frac{1}{\mathrm{e}(h,j)}\Big(&\mathrm{f}(h,j)-\mathrm{a}(h,j)\mathrm{U}^{\mathrm{initial}}(h+1,j)-\mathrm{b}(h,j)\mathrm{U}^{\mathrm{initial}}(h-1,j)\\&-\mathrm{c}(h,j)\mathrm{U}^{\mathrm{initial}}(h,j+1)-\mathrm{d}(h,j)\mathrm{U}^{\mathrm{initial}}(h,j-1)\Big).\end{aligned} \tag{2.25}$$

A new value $\mathrm{U}^{\mathrm{new}}(h, j)$ can be obtained as the weighted average of the calculated and initial values

$$\mathrm{U}^{\mathrm{new}}(h,j)=\omega\mathrm{U}(h,j)+(1-\omega)\mathrm{U}^{\mathrm{initial}}(h,j), \tag{2.26}$$

where ω is called the relaxation parameter. Thus

$$\mathrm{U}^{\mathrm{new}}(h,j)=\mathrm{U}^{\mathrm{initial}}(h,j)-\frac{\omega\xi(h,j)}{\mathrm{e}(h,j)}, \tag{2.27}$$

where in the first step the residual $\xi(h, j)$ is defined as

$$\begin{aligned}\xi(h,j)=\;&\mathrm{a}(h,j)\mathrm{U}^{\mathrm{initial}}(h+1,j)+\mathrm{b}(h,j)\mathrm{U}^{\mathrm{initial}}(h-1,j)\\&+\mathrm{c}(h,j)\mathrm{U}^{\mathrm{initial}}(h,j+1)+\mathrm{d}(h,j)\mathrm{U}^{\mathrm{initial}}(h-1,j)\\&+\mathrm{e}(h,j)\mathrm{U}^{\mathrm{initial}}(h,j+1)-\mathrm{f}(h,j).\end{aligned} \tag{2.28}$$

For each propagation step, the above procedure from Eq.(2.25) to Eq.(2.27) is repeated with $\mathrm{U}^{\mathrm{initial}}(h, j)$ replaced by $\mathrm{U}^{\mathrm{new}}(h, j)$ until a criterion for terminating the iteration is reached. The stopping criterion is related to the residual and will be discussed below. This way, the correct field components in all nodes of one propagation step can be obtained.

The choice of ω is very important. Only $0 < \omega < 2$ leads to a converging process [1]. The optimal value of ω depends on the individual situation and a proper choice will lead to a faster convergence. Besides, the necessary number of iterations for a $H \times J$ grid division depends on the initial guess, the number of grid nodes, step sizes and the reference propagation constants. On the other hand, for a converging process, the residual becomes smaller and smaller as the iteration carries on. To obtain accurate results, a high precision for both the real and imaginary parts of the quantities used in the iteration must be ensured during this process. Therefore, double precision should be used in the iteration subroutine of the computer program.

For convenience, the above discussion is based upon one equation, i.e. concerning a single set of h, j, l values. If we discuss all equations simultaneously, a matrix equation as in (2.9) has to be used. Then $\mathbf{U}(h, j)$ becomes the solution vector.

An iterative method will produce a sequence $\{\mathbf{U}^i(h,j)\}$ of vectors converging to a vector $\mathbf{U}(h,j)$ satisfying the matrix equation. To be effective, the process must decide when to stop. A good stopping criterion should

(1) identify if the error $\mathbf{e}^i = \mathbf{U}^i(h,j) - \mathbf{U}(h,j)$ is small enough to stop,
(2) stop if the error is no longer decreasing or is decreasing too slowly,
(3) limit the maximum amount of time spent iterating.

There are many kinds of stopping criteria. The most widely used one is to terminate the iteration when the error defined as $\sum_h \sum_j \left|\xi_{h,j}\right|$ reduced by a factor of 10^{-P} with respect to its initial value, where P is a prescribed integer. This criterion is simple and easy to program. However, it has the disadvantage of strongly depending on the initial error. Therefore, the choice of P must be treated carefully. For our BPM, we used this type of criterion. Afterwards, however, we always verified the absolute accuracy.

2.3.3 Acceleration of the iterations

We have accelerated the iterations by use of the following [9]:

Decreased initial error: An initial estimate of the solution vector for each propagation step is necessary for the iteration in the relaxation method, which in literature is usually chosen as the zero vector. An error defined as $\sum_h \sum_j \left|\xi_{h,j}\right|$ will be introduced by this initial guess. A bigger initial error will correspond to a bigger number of iteration steps to obtain the same absolute accuracy, leading to a lower speed. Therefore, an appropriate initial guess is important to speed up the simulation. Based upon the local continuity of the optical field, we expect that the calculated field in the previous propagation step is a suitable initial estimate for the field in the next step.

As an example, we simulated a rib waveguide structure as described in [5] with a 61×61 grid division ($\Delta x = \Delta y = \Delta z = 50$ nm) at a wavelength of 1500 nm. We chose $\beta = 14.205$ μm^{-1}. The reason is that there exists an optimal reference propagation constant for each structure, which corresponds to the negligible spurious numerical loss, as discussed in Section 3.3.1. The value of the scheme parameter α is chosen as 0.7 because it ensures the stability of the numerical scheme (also see Section 3.3.1). The HE_{00} guided mode field is launched as the input at $z = 0$ μm. To reach a given accuracy, the zero initial guess of the propagating field requires 90 iterations for one propagation step, while using the calculated field of the previous step as an initial guess only 58 iterations were required. Thus, a 55% speed increase is obtained.

Reduced iteration parameter: Usually, the iteration parameter P is chosen as $P \geq 5$. A bigger P will lead to a higher accuracy, but the simulation speed will decrease simultaneously. Therefore, P should be chosen as small as possible within the accuracy limits required for the results.

Since the field in the previous propagation step has been used as the initial estimate for the field in the following propagation step and the propagation step size is usually small (≤ 0.1 μm), the initial error is quite small. Our results show that $P = 3$ is enough for an accurate simulation, which means that the relative error between the calculated results (field amplitudes) with $P = 3$ and $P = 5$ is less than 10^{-4}. Still using the same waveguide structure mentioned above, 99 iterations are needed for $P = 5$, while only 58 iterations are needed for $P = 3$, leading to a 70% speed increase.

Transparent boundary condition: By choosing the transparent boundary condition, since the field remains undisturbed in the neighborhood of the boundaries, a smaller widow size can be used. Because the simulation speed is inversely proportional to J^3 for a $J \times J$ grid division, the speed has been effectively increased. As an example, a 61 × 61 grid division of the rib waveguide structure corresponds to a simulation time of 3.3 seconds for a single propagation step on a workstation (SUN Sparc 10/40), while 6.6 seconds for a 81 × 81 grid division. Since a small computational window can be used and thus a high simulation speed can be obtained, this transparent boundary condition is preferred in our computations.

2.4 Summary

A fully vectorial finite difference beam propagation method for three-dimensional waveguide structures, including z-variant structures has been developed. We solve the discretized differential equations by using the relaxation method. The stability, numerical dissipation and boundary condition are discussed. Several ways to speed up the simulation have been given, including: applying a modified Crank-Nicholson scheme, introducing a decreased initial error, using a reduced iteration parameter and applying a transparent boundary condition.

2.5 References

[1] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling: *Numerical Recipes*, Cambridge University Press (1986).

[2] D. Yevick and B. Hermansson: "A split-step finite difference analysis of rib waveguides," *Electron. Lett.*, **25**, 461-462 (1989).

[3] W.P. Huang, C.L. Xu, S.T. Chu, and S.K. Chaudhuri: "The finite-difference vector beam propagation method: analysis and assessment," *J. Lightwave Technol.*, **10**, 295-304 (1992).

[4] R.D. Richtmyer and K.W. Morton: *Difference methods for initial value problems, 2nd ed.*, Wiley-interscience (1967).

[5] D. Li, H. van Brug, and H.J. Frankena: "Vectorial FD-BPM for 3-D waveguide structures," in *Linear and Nonlinear Integrated Optics*, Giancarlo C. Righini, David Yevick, Editors, Proc. SPIE **2212**, 36-46 (1994).

[6] D. Yevick and B. Hermansson: "New formulations of the matrix beam propagation method: Application to rib waveguides," *IEEE J. Quantum Electron.*, **25**, 221-229, (1989).

[7] G. Ronald Hadley: "Transparent boundary condition for the beam propagation method," *IEEE J. Quantum Electron.*, **28**, 363-370 (1992).

[8] G. Ronald Hadley: "Transparent boundary condition for beam propagation," *Opt. Lett.*, **16**, 624-626 (1991).

[9] D. Li, H. van Brug, and H.J. Frankena: "Investigation on acceleration of vectorial beam propagation method," *Pure Appl. Opt.* (being published, 1995).

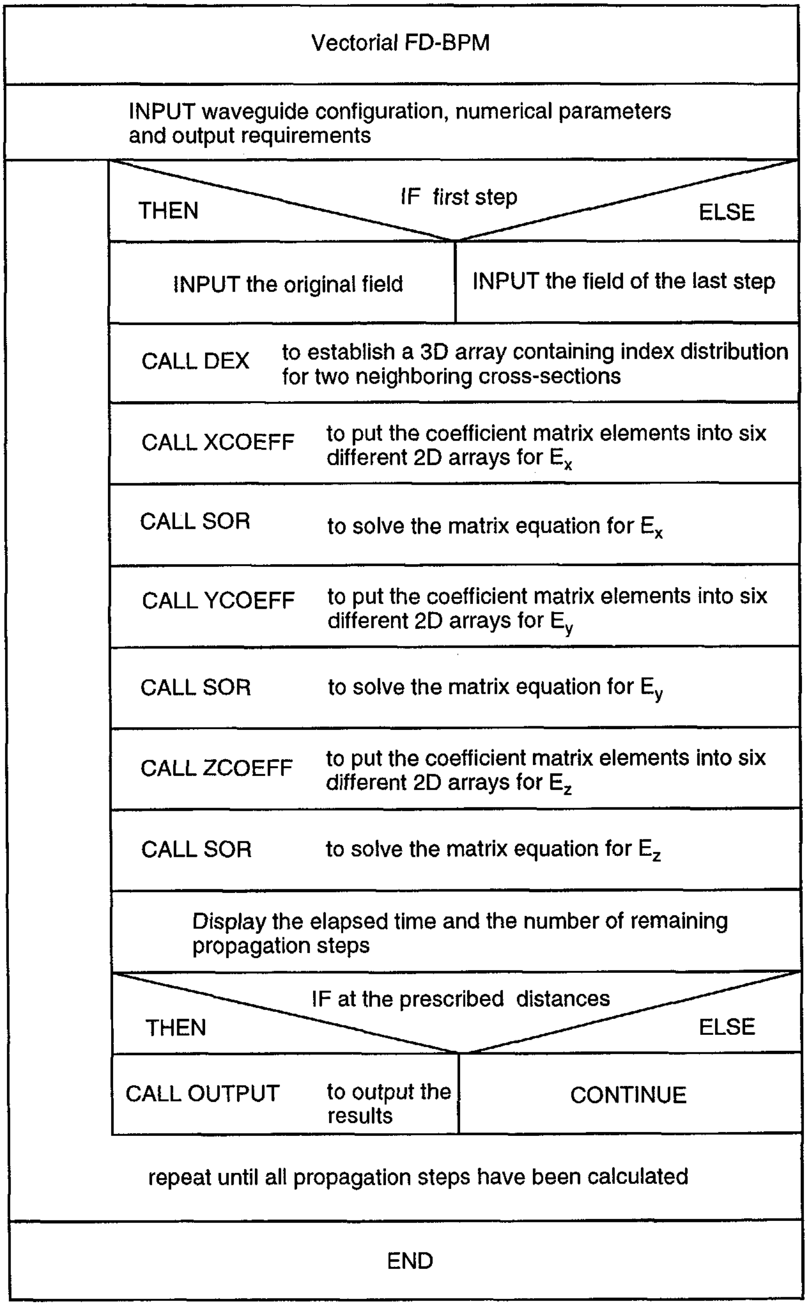
Chapter 3

Programming and simulation

In Section 3.1, we describe the programming and discuss some related procedures we have used. Results of a benchmark test including three types of computers are presented in Section 3.2. The influence of the finite difference scheme parameter α, the reference propagation constant β, the step sizes and the boundary condition on the simulation results are shown in Section 3.3. To this aim, the wave propagation through the following structures has been simulated: a straight rib waveguide, a longitudinally varying tapered waveguide, a polarization splitter, and a polarization converter.

3.1 Programming

Programming the fully vectorial FD-BPM includes the design of the software algorithms, choice of the numerical scheme, writing of the source code, test of the validity of the results, acceleration of the program and overall optimization. The source code has been written in FORTRAN language, and comprises various subroutines. Most computing is carried out on the SUN 10/40 workstation. A benchmark test also includes other two types of workstations: the HP 735 and the Cray C98/4256, as discussed in Section 3.2. We used SUN FORTRAN 1.4 as the programming language, which is an enhanced FORTRAN 77 development system and conforms to ISO 1539-1980, corresponding to the ANSI X3.9-1978 FORTRAN standard. It provides an IEEE standard 754-1985 floating-point package. The debugging tool *dbx* was used to debug our program.



```
Vectorial FD-BPM
INPUT waveguide configuration, numerical parameters
and output requirements
    IF first step
    THEN
        INPUT the original field
    ELSE
        INPUT the field of the last step
    CALL DEX        to establish a 3D array containing index distribution
                    for two neighboring cross-sections
    CALL XCOEFF     to put the coefficient matrix elements into six
                    different 2D arrays for $E_x$
    CALL SOR        to solve the matrix equation for $E_x$
    CALL YCOEFF     to put the coefficient matrix elements into six
                    different 2D arrays for $E_y$
    CALL SOR        to solve the matrix equation for $E_y$
    CALL ZCOEFF     to put the coefficient matrix elements into six
                    different 2D arrays for $E_z$
    CALL SOR        to solve the matrix equation for $E_z$
    Display the elapsed time and the number of remaining
    propagation steps
    IF at the prescribed distances
    THEN
        CALL OUTPUT     to output the
                        results
    ELSE
        CONTINUE
repeat until all propagation steps have been calculated
END
```

Figure 3.1 Flow chart of our computational setup

3.1.1 Program structure

The main program has been divided into three interactive sections, for the parameter values input, the simulation and the output, respectively. The parameter input section guides the user to provide various waveguide parameters, iteration parameters, input fields, and requirements for the output. In the simulation section, for each propagation step, the elements of the three coefficient matrices occurring in Eqs.(2.18) are calculated, and subsequently iterations are carried out to obtain the field distributions. This process is executed repeatedly. In the output section, various results such as field strengths, field amplitudes, field intensities, and power attenuation are stored according to the choice of user. The formatted output data can be directly used by *Gnuplot* software on a Unix workstation and by *Mathematica* software on a PC or Macintosh for plotting. The flow chart is shown in Figure 3.1.

3.1.2 Arrays

Coefficient matrix arrays: Each coefficient matrix occurring in Eqs.(2.18) is dimensioned as $[(H\text{-}2)\cdot(J\text{-}2)]$ by $[(H\text{-}2)\cdot(J\text{-}2)]$, but is sparse having a tridiagonal form with fringes. For a grid division $H = J = 102$, there will be 10^8 elements in a single matrix. Since one IEEE double precision value occupies 8 bytes and a double complex one 16 bytes, 10^8 elements occupy more than 800 Mb. It would prove disastrous trying to store all of them, since most computer systems can not handle that. To solve this problem, we have introduced six arrays dimensioned as $(H\text{-}2)$ by $(J\text{-}2)$ to store only the six groups of the non-zero elements, as indicated in Eqs.(18). Then, the necessary memory space is reduced to less than 1 Mb for one matrix.

Refractive index array: For a tapered waveguide structure, the cross-sectional distribution of refractive indices changes as the propagation continues. The storage of index values on all discretization nodes will take a huge memory and disk space. We have adopted two methods to overcome this problem. The first one is to introduce a function subroutine $R(h,j,l)$, which has a unique value for one set of h,j,l values. Such a treatment requires a small memory size. However, in subroutines XCOEFF, YCOEFF and ZCOEFF, frequent calls of the function subroutine $R(h,j,l)$ take a considerable CPU time. The second method is to use a three-dimensional array with dimensions $(H,J,2)$, where H and J are the numbers of nodes in x- and y- directions, respectively. This means that only the index values of two neighboring cross-sections are stored for each propagation step, thus requiring both low memory size and CPU time.

3.1.3 Numerical considerations

We have applied the standard Crank-Nicholson FD scheme to Eqs.(2.7) in the beginning of the investigation. As shown in Section 2.2.1, we then experienced unacceptably low simulation speeds. To enable any investigation, we only could allow a very small number of grid nodes (H, $J < 31$) at that time. This painful experience pushed us to introduce a modified Crank-Nicholson scheme and other acceleration procedures. In Sections 2.2.1 and 2.3.3, these effective measures have been presented and analyzed.

After using the modified Crank-Nicholson scheme, Eqs.(2.7) are converted to five-points differential equation problems. A relaxation algorithm through iterations has been used to obtain the propagating fields step by step. The choice of the parameter ω is critical for the convergence of the iterations (see Section 2.3.2). According to the literature [1], only the values $1 < \omega < 2$ can lead to a fast convergence, if Dirichlet and the Neumann boundary conditions are used. We have used the recently developed transparent boundary conditions (see Section 2.3.1). Our results show, however, that the fastest convergence is obtained if ω is about 0.9, which we have used as the preset parameter value in the simulations.

3.1.4 Double precision

According to the IEEE standards [3], a single precision value occupies 32 bits. Bits 0:22 store the 23-bit mantissa, bits 23:30 store the 8-bit exponent, and bit 31 stores the sign (positive or negative). The 23-bit mantissa combined with the implicit leading bit provides 24 bits of precision in normal numbers, which is always more accurate than 6 decimal digits but always less precise than 9 decimal digits [3]. Similarly, an IEEE double precision value is always more precise than 18 decimal digits and less accurate than 21 decimal digits.

The stopping criterion for an iteration process has been discussed in Section 2.3.2. For example, we can decide that the iteration process should not stop until an absolute value 10^{-10} of the error, defined as $\sum_h \sum_j |\xi(h, j)|$ based upon Eq.(2.28), is reached. Therefore, all variables in the expression for $\xi(h,j)$ must have an absolute precision higher than 10^{-10}, otherwise the iteration stopping criterion can never be reached. In this case, a single precision will not be enough which can only ensure a precision up to 6 decimal digits. Consequently, in the iteration subroutine, a higher precision is extremely important to determine when the iteration stops and thus influences the accuracy of the calculated results. We have used double precision for all constants and variables (real or complex) in the pertinent subroutine.

3.1.5 Debugging

For a big program whose source code contains more than one thousand lines, finding all errors is complicated. After removing all syntactical errors during compilation, an executable file is produced. However, a compilable program doesn't mean that results will be correct. For example, there might be programming mistakes or typing errors in formulas, which can not be recognized by the compiler. To find such bugs, we perform simulations on simple and well-known structures. The results can be compared with those obtained by other methods. The debuggers *dbx* and *dbxtool* on the SUN workstation have proven to be useful for finding these bugs. With *dbx*, the programmer can display and modify variables, set breakpoints, trace variables and invoke procedures in the program being debugged without having to recompile. *dbxtool* allows a more effective use of *dbx* by replacing the original, terminal-oriented interface with a window- and mouse-based interface.

To use *dbx* or *dbxtool*, one must load and compile the program with the **-g** flag, i.e. use the command

```
f77 -o test.exe -g test.f,
```

thus compiling **test.f** and producing an executable file ***test.exe***. Then, by typing ***dbx test.exe***, the debugging is invoked. After the *(dbx)* prompt appears, **run** must be typed to start the execution. The detailed description of the debugging can be found from the literature [3].

3.1.6 Optimization

Nowadays almost all compilers offer various optimization choices, which can greatly improve the calculation speed (see Section 3.2). However, every machine has its limited abilities. The manual optimization of a source code is always recommendable. In literature, [4] proved to be a useful reference book.

For example, nested *DO* loops (≥ 3) should be avoided if possible, since they usually take more CPU time. Special optimization efforts are needed for codes inside the innermost loop. In fact, the subroutine SOR in our program contains essentially three nested loops. This subroutine takes 95% of the total CPU time needed. Therefore, optimization efforts are mainly directed to improve this subroutine. In addition, we experienced that different computers have different requirements for an optimal structure of the source code file (see Section 3.2.3).

3.2 Benchmark test

A benchmark test including three computers (SUN 10/40, HP 735 and Cray C98/4256) has been carried out. We simulated the field propagation through a rib waveguide structure [1] using a 61 × 61 grid division over the propagation distance of 1 μm, with the steps $\Delta x = \Delta y = \Delta z = 0.05$ μm. Exactly the same FORTRAN source files have been compiled at all three computers and then executed, with and without the optimization choices which will be discussed later.

3.2.1 SUN 10/40

For the SUN 10/40 workstation, the basic compiling command has been used without optimization choices, using the command:

f77 -o *sun1.exe* test.f.

Then, the execution of ***sun1.exe*** took 227.3 seconds. However, using the following optimization choice,

f77 -fast -O3 -o *sun2.exe* test.f,

where **-fast** automatically selects the combination of compilation options that optimizes for speed of execution without excessive compilation time and **-O3** sets the optimization to the level 3*, the execution of ***sun2.exe*** required 74.9 seconds, a speed increase of 203%. According to the reference book [3], this should work closely to the maximum performance for most realistic applications.

3.2.2 HP 735

We also made computations on a HP 735 workstation. Without extra optimization choices, i.e. using the command

f77 -o *hp1.exe* test.f,

the execution of ***hp1.exe*** took 60.1 seconds. We further used the optimization choice, i.e.:

f77 +O4 -o *hp2.exe* test.f,

where **+O4** means that the optimization option is set to the highest level 4. This includes store/copy optimization, unused definition elimination, loop invariant code motion, common sub-expression elimination and register allocation etc. Then, the execution of ***hp2.exe*** required 29.7 seconds, meaning a speed increase of 102%.

3.2.3 Cray C98/4256

The Cray C98/4256 is a super computer, which contains four parallel processors. We only used one of them to test our program. Without extra optimization choices, i.e. using the command

cf77 -o *cray1.exe* test.f,

the execution of ***cray1.exe*** took 58.6 seconds. On the other hand, we used an optimization choice by calling:

f77 -Zv -o *cray2.exe* test.f,

where **-Zv** means that the compiler uses the highest vectorization option (level 3). Then

---

* The highest level actually is 4, but that proved to be worse than the level 3 when combining with the -fast flag in our program.

the execution time of ***cray2.exe*** required 17.8 seconds, a speed increase of 229%.

For a super computer, the above times are unacceptably long. An inspection on the source code file was carried out by specialized personnel*. They suggested that a minor change in one time-consuming subroutine should be introduced, as indicated in Appendix D. The purpose is to vectorize *DO* loops in the routine. Without an appropriate source code change, the innerest *DO* loop can not be fully vectorized, which has a critical influence on the speed.

Then, without extra optimization choices, i.e. calling:

```
cf77  -o cray3.exe  new.f,
```

the execution of the vectorized program ***cray3.exe*** took 3.8 seconds only, while with the optimization choice, i.e.:

```
f77  -Zv  -o cray4.exe new.f,
```

the execution of ***cray4.exe*** required 2.2 seconds, showing a huge increase of the computational speed as compared to the results before the source code change.

However, the above source code change has no influence on the calculation speed of the SUN 10/40, and even has a negative influence on that of the HP 735, which means that the CPU time increases. Therefore, the structure of a source code file has later been optimized according to the used computers.

We thus arrive at the following conclusions:

(1) without source code change and without extra optimizations, the calculation time ratios are

SUN 10/40 : HP 735 : Cray C98/4256 = 227.3 : 60.1 : 58.6

(2) without source code change but with extra optimizations, the ratios become

---

* The kind assitance of Willem Vermin and Bert van Corler of SARA, Amsterdam is gratefully acknowledged.

SUN 10/40 : HP 735 : Cray C98/4256 = 74.9 : 29.7 : 17.8

(3) with the source code change but without extra optimizations, the calculation time ratios are

SUN 10/40 : HP 735 : Cray C98/4256 = 226.2 : 60.1 : 3.8

(4) with the source code change and with extra optimizations, calculation times are

SUN 10/40 : HP 735 : Cray C98/4256 = 74.1 : 29.7 : 2.2

Theoretically, using one of its processors only, the Cray C98/4256 could be about 40 times faster than the SUN 10/40, under the assumption that the source code can be fully vectorized. Based upon the above source code change, with the source code optimization, we have reached a factor of 35. If all four parallel processors are used simultaneously, the calculation speed on the Cray may be further improved another 4 times, thanks to the possibility of multitasking operation [2].

3.3 Simulation of waveguide structures

For a BPM with the slowly varying envelope approximation, the choice of the reference propagation constant is important. There exists an optimal reference propagation constant corresponding to the lowest spurious numerical loss for each structure. This value is approximately equal to that of the guided mode of the local structure, as discussed in Section 3.3.1. In practice, we first calculate the propagation constants of the guided modes by use of the FEM analysis. Based upon the calculated values, the optimal reference propagation constants are searched. In addition, since the modified Crank-Nicholson FD scheme has been used, the introduced scheme parameter α ($\alpha > 0.5$) is found to be critical. There exists a critical value of α_c for each structure. If $\alpha < \alpha_c$, the numerical scheme is unstable, leading to an exponentially increasing field as the propagation proceeds. If $\alpha > \alpha_c$, the scheme is stable, but there exists a spurious numerical loss. The value of α will be chosen bigger than α_c but as small as possible to decrease the spurious numerical loss, as discussed in Section 2.2.2.

3.3.1 Rib waveguide

We first simulated the fields propagating through the same straight rib waveguide structure that has been simulated by several other researchers [5-6]. Such structures are widely used in OICs.

The structure is shown in Figure 3.2. The height and the width of the rib are 1.1 μm and 2.0 μm, respectively, the thickness of the side slab is 0.2 μm; the refractive indices of the cover, core and substrate are 1.00, 3.44 and 3.34, respectively. Here, the vacuum wavelength λ_0 is 1.5 μm.

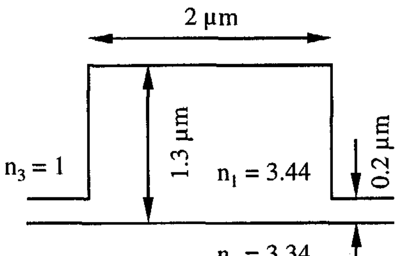


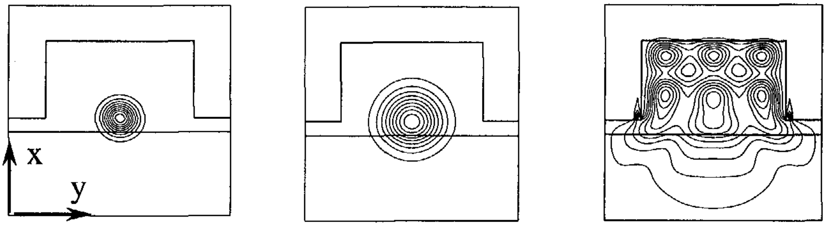
Figure 3.2 Rib waveguide structure.

Simulation results

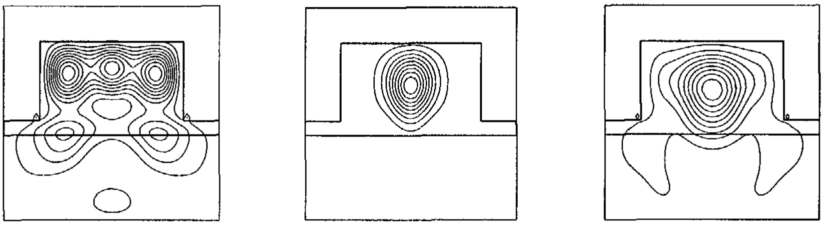
As an example, we show the simulation results for which the parameters are chosen as follows: the computational window of 3.0 × 3.0 μm is uniformly discretized with a 61 × 61 grid division; $\Delta x = \Delta y = 0.05$ μm and the propagation step Δz is also 0.05 μm. The reference propagation constant is taken as $\beta = k_0 n_0$ ($k_0 = 2\pi/\lambda_0$, $n_0 = 3.39$.) and the scheme parameter $\alpha = 0.7$ has been used. Both horizontally and vertically polarized Gaussian beams with a 0.3 μm radius (beam width) of the 1/e amplitude are coupled into the rib waveguide in the plane of $z = 0$ μm. Figures 3.3 and 3.4 show the evolution of isointensity contour patterns in the rib waveguide for horizontally and vertically polarized inputs, respectively.

To allow a comparison with results of other researchers who used a semi-vectorial method [5] or a scalar method [6], we show the dominant component of the electric field strength.

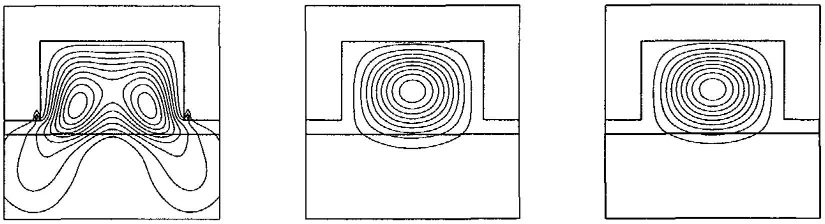
In both cases, the beams first expand by diffraction until they reach the dielectric-air interface. Subsequent reflections cause radiation modes propagating into the substrate. After propagation over 100 µm, the fields become stable, showing the field distribution of the guided HE_{00}- and EH_{00}-modes, respectively. There are evident differences between field profiles of Figures 3.3 and 3.4, where the polarization directions are different. Therefore, a scalar description will be not fully adequate for such a kind of structure.



(a). z = 0 µm (b). z = 1 µm (c). z = 3 µm

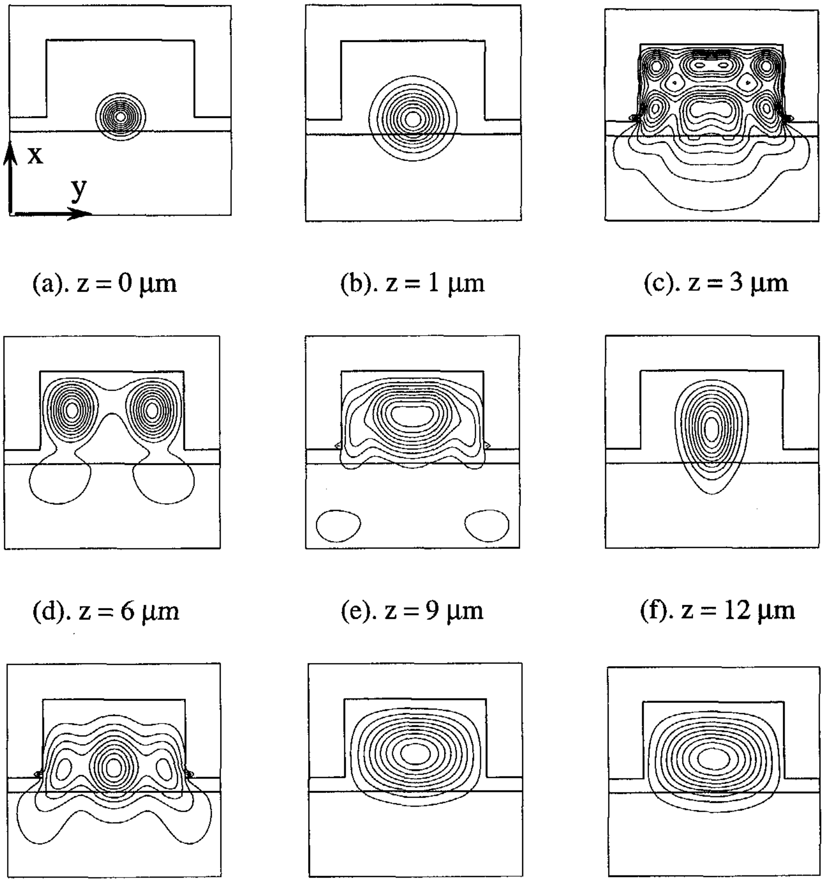


(d). z = 6 µm (e). z = 9 µm (f). z = 12 µm



(g). z = 15 µm (h). z = 50 µm (i). z = 100 µm

Figure 3.3 Evolution of iso-intensity contours of E_y for a horizontally polarized Gaussian beam coupled into the rib waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ µm; $\alpha = 0.7$, $n_0 = 3.39$.



(a). z = 0 μm (b). z = 1 μm (c). z = 3 μm

(d). z = 6 μm (e). z = 9 μm (f). z = 12 μm

(g). z = 15 μm (h). z = 50 μm (i). z = 100 μm

Figure 3.4 Evolution of iso-intensity contours of E_y for a vertically polarized Gaussian beam coupled into the rib waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ μm; $\alpha = 0.7$, $n_0 = 3.39$.

Schulz *et al* [5] have simulated the same structure up to z = 15 μm with a semi-vectorial FD-BPM based upon a series expansion of the Helmholtz propagation operator. They took into account one polarization direction and expressed fields in one component, thus neglecting the couplings between the three electric field components. Both so-called Quasi-TE and Quasi-TM cases have been simulated, respectively. Their results are in perfect agreement with ours, but their method can not be applied when an arbitrarily polarized field is launched. Ratowsky *et al* [6] have also simulated this structure up to $z = 7\ \mu m$ with the iterative Lanczos reduction method. They launched a Gaussian beam of the 0.3 μm 1/e intensity radius (0.4242 μm 1/e amplitude radius) at 1.55 μm wavelength into the waveguide, using a scalar method. Our simulation shows that their results are only approximately identical to our results with a horizontally polarized Gaussian beam input.

Launching the guided mode field calculated using the finite element method* (FEM) and the Gaussian beam with a 0.3 μm beam width, respectively, we next consider the power attenuation which is defined as $-10\times\log_{10}(P/P_0)$, where P_0 is the power at $z = 0$ and P is the power after propagation over a given distance.

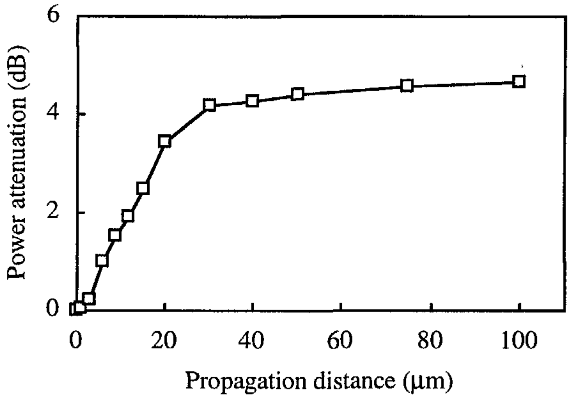


Figure 3.5 Relationship between the power attenuation and the propagation distance for a horizontally polarized Gaussian beam coupled into the rib waveguide. $\Delta x = \Delta y = \Delta z = 0.05\ \mu m$; $\alpha = 0.7$, $n_0 = 3.39$.

Figure 3.5 shows that the power attenuation is considerable for the horizontally polarized Gaussian beam input because this input is different from the HE_{00}-mode field and thus significant power will radiate into the substrate. The power attenuation finally stabilize when the propagated field becomes identical to the guided HE_{00}-mode field. The total power attenuation after propagation over 100 μm is about 4.65 dB. For the guided HE_{00}-

---

* Kindly provided by the Institute of Quantum Electronics, ETH Zürich, Switzerland.

mode field input, the power attenuation due to the numerical scheme appears to be extremely small. Then, the total power attenuation is less than 0.05 dB after propagation over 100 μm. The simulation time for this rib waveguide structure with the propagation distance up to 100 μm took about 1.8 hours on the SUN Sparc 10/40 workstation.

Influence of the reference propagation constant

Since the slowly varying envelope approximation has been used, the influence of the reference propagation constants on the simulation results is inevitable. Figure 3.6 shows the relationship between the reference refractive index and the total power attenuation for the structure with a horizontally polarized Gaussian beam as the input after the stable guided HE_{00}-mode is formed. There is a small range of optimal reference index values (between 3.38 and 3.39) corresponding to the small power loss. The FEM analysis shows that the HE_{00}-mode has a propagation constant of $\beta = k_0 n_g$ with $n_g = 3.391$. Thus, the optimal values are approximately the same as that for this mode. If the optimal reference index is unknown, it can be chosen the value of the guided mode calculated by the FEM analysis.

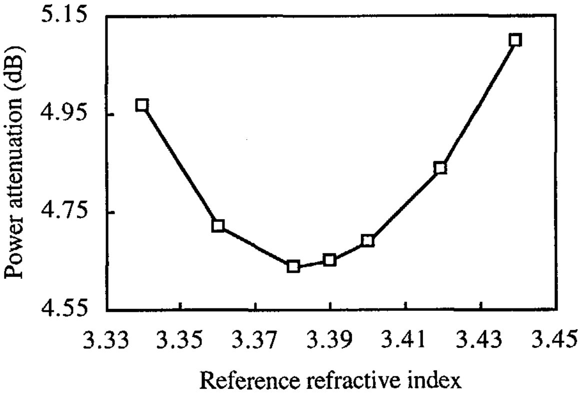


Figure 3.6 Relationship between the power attenuation and the reference refractive index for a horizontally polarized Gaussian beam coupled into the rib waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ μm; $\alpha = 0.7$. The propagation distance is 100 μm. The calculated effective index of the HE_{00}-mode is 3.391.

Influence of the scheme parameter

We have used the modified Crank-Nicholson finite difference scheme (see Section 2.2.1). The von Neuman analysis shows that, for a three-dimensional waveguide structure, there exists a critical scheme parameter $0.5 < \alpha_c < 1$ which depends on the structure parameters and on the reference propagation constant. If $\alpha < \alpha_c$, the numerical simulation will be unstable, leading to an unrealistic power increase of the propagating field. If $\alpha > \alpha_c$, the numerical simulation will be stable, however, an extra numerical dissipation occurs. The optimal choice is $\alpha = \alpha_c$.

Huang *et al* [7] have described a detailed analysis for the slab waveguide structure, where $\alpha \geq 0.5$ can be used. Our results show that, for the rib waveguide structure, if $\alpha < 0.6$, the numerical simulation will be unstable. But if $\alpha \geq 0.6$, stability is ensured.

The influence of α on the power attenuation depends on the choice of the reference propagation constant β. If a non-optimal reference propagation constant has been chosen, the influence of α on the spurious power attenuation will be stronger. A bigger value of α will lead to a bigger power attenuation. Therefore, the optimal reference refractive index should be used to minimize this influence of α. In addition, under the condition of ensured stability, a smaller α value is recommended in order to obtain the smallest power dissipation introduced by the numerical procedure.

Influence of step sizes

Bigger transverse step sizes Δx and Δy can lead to a bigger power attenuation, since then the structure can not be discretized precisely. Figure 3.7 shows the relationship between the power attenuation and the step sizes of Δx and Δy for the rib structure with a horizontally polarized Gaussian beam as the input. If $\Delta x = \Delta y \leq 0.1$ μm, the relative total spurious power attenuation is less than 2×10^{-3} and thus can be neglected.

On the other hand, a bigger propagation step size Δz can also lead to a bigger power attenuation (see Figure 3.8), where the spurious power loss can be also neglected if $\Delta y \leq 0.1$ μm. However, small step sizes of Δx, Δy and Δz correspond to more grid points, leading to a lower simulation speed. As a compromise, Δx, Δy and Δz between 0.05 and 0.1 μm are satisfactory choices for the particular structure under consideration by accounting for both the speed and the accuracy. For a structure with small waveguide dimensions, this choice is no longer correct. A fine grid division has to be used to ensure precision, thus a lower simulation speed must be expected.

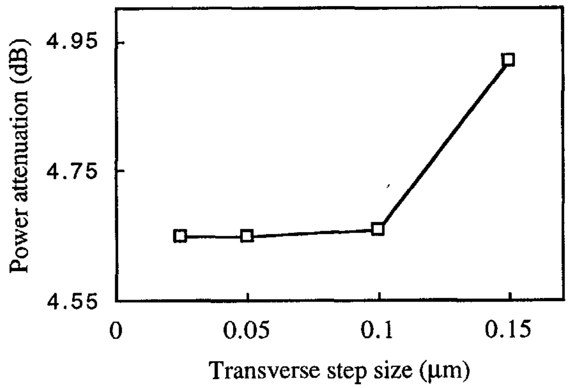


Figure 3.7 Relationship between the power attenuation and the transverse step size for a horizontally polarized Gaussian beam coupled into the rib waveguide. $\Delta z = 0.05$ μm; $\alpha = 0.7$, $n_0 = 3.39$. The propagation distance is 100 μm.

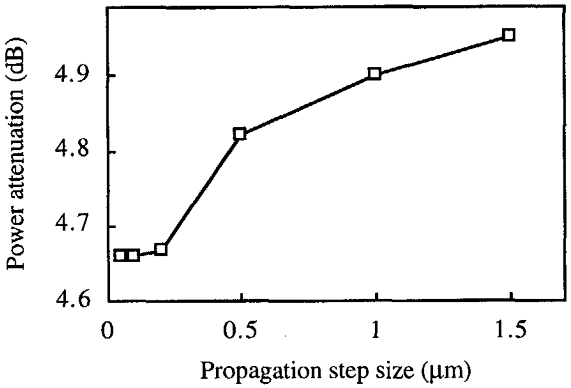


Figure 3.8 Relationship between the power attenuation and the propagation step size for a horizontally polarized Gaussian beam coupled into the rib waveguide. $\Delta x = \Delta y = 0.1$ μm; $\alpha = 0.7$, $n_0 = 3.39$. The propagation distance is 100 μm.

Extraction of propagation constants

Propagation constants can be extracted from the overlap between the input field and the propagated field [8]. The theoretical background is as follows:

An electric field can be expressed as the superposition of orthogonal modes [8]

$$\mathbf{E}(x,y,z)=\sum_{n,j}\left(A_{n,j}\mathbf{u}_{n,j}(x,y)\exp(i\beta_n z)\right), \tag{3.1}$$

where in the case of degeneration the summation includes all modes with the same β_n and where j is used to distinguish different modes within a degenerate set having the same propagation constant. Then

$$\begin{aligned}P(z)&=\iint\left[\mathbf{E}^*(x,y,0)\cdot\mathbf{E}(x,y,z)\right]dxdy\\&=\sum_{n,j}\left[\left|A_{n,j}\right|^2\exp(i\beta_n z)\right].\end{aligned} \tag{3.2}$$

Let $z=z_0+l\Delta z$ ($l = 0, 1, 2, \cdots, N-1$). Then, the discrete Fourier transform of Eq.(3.2) is

$$F(m)=\frac{1}{N}\sum_{l=0}^{N-1}\left[P(l)w(l)\exp\left(-i2\pi ml/N\right)\right], \tag{3.3}$$

where $m = 0, 1, 2, \cdots, N-1$ and $w(l) = 1-\cos(2\pi l/(N-1))$, which is the Hanning window function that had to be introduced since N is finite [8]. For a thorough discussion of the use of windows in numerical Fourier analysis, the reader is referred to [9]. Notice that $P(l)$ and $F(m)$ stand for $P(z_0+l\Delta z)$ and $F(m\Delta t)$, respectively, where $\Delta t = 1/(N\Delta z)$. Eq.(4.3) can be rewritten as

$$F(m)=\sum_{n,j}\left[\left|A_{n,j}\right|^2 L(\beta_n-\beta)\right], \tag{3.4}$$

where $\beta = 2\pi m/(N\Delta z)$ and $L(\beta_n-\beta)$ is

$$L(\beta_n-\beta)=\frac{1}{N}\sum_{l=0}^{N-1}\left[w(l)\exp\left(i(\beta_n-\beta)l\Delta z\right)\right]. \tag{3.5}$$

Therefore, calculated spectra of F(m) will display a series of resonance peaks with maxima at $\beta = \beta_n$. But in practice, only a finite set of values of P(z) is available. The sampled values of β do not exactly coincide with that of the guided modes. As a result, an inaccuracy will occur. The maximal uncertainty in β_n is $\pi/(N\Delta z)$ since $\beta = 2\pi m/(N\Delta z)$. However, through a line fitting method [8], the uncertainty can be effectively decreased, as indicated as follows. If a local maximum of F(m) exists at $m = m'$, we can obtain [8]

$$\beta_n = \beta_{m'} - \delta \frac{2\pi}{(N-1)\Delta z}, \tag{3.6}$$

where $-0.5 \le \delta \le 0.5$ and

$$\delta = \begin{cases} \dfrac{3r + \sqrt{9r^2 - 8}}{2} & (r < 0), \\ \dfrac{3r - \sqrt{9r^2 - 8}}{2} & (r > 0), \end{cases} \tag{3.7}$$

with

$$r = \frac{F(m'-1) + F(m'+1)}{F(m'-1) - F(m'+1)}.$$

As an example, we extract the propagation constant of the HE_{00} guided mode for the rib waveguide structure, through which only the zero order modes HE_{00} and EH_{00} can propagate. We launch a horizontally polarized Gaussian beam with a 1/e amplitude radius (beam width) of 0.4 μm at 1.5 μm wavelength and propagate over 1 mm.

Figure 3.9 shows the relationship F(β), the peak corresponds to the HE_{00} mode. The calculated value of the propagation constant from our BPM is 14.193 μm^{-1}, which is in excellent agreement with the calculated value 14.205 μm^{-1} by the FEM based program.

If a symmetrical input field is launched into a multimodal structure, only even modes will be excited. Therefore, an asymmetrical input field is necessary if all modes need to be extracted.

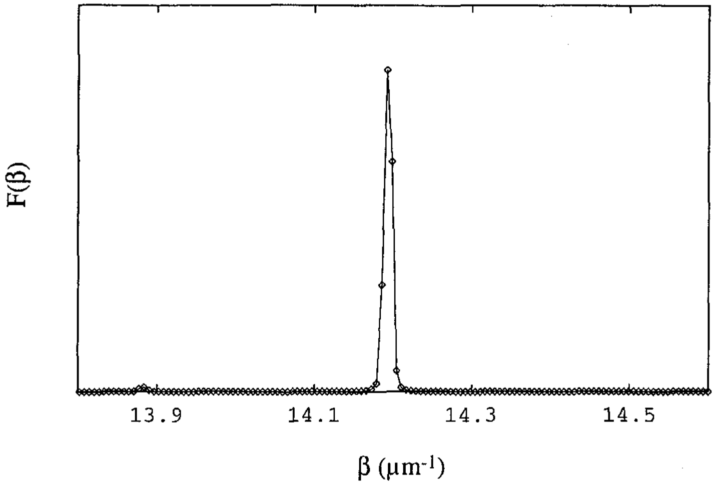


Figure 3.9 Relationship F(β) for a horizontally polarized Gaussian beam coupled into the rib waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ μm; $\alpha = 0.7$, $n_0 = 3.39$.

3.3.2 Tapered buried waveguide

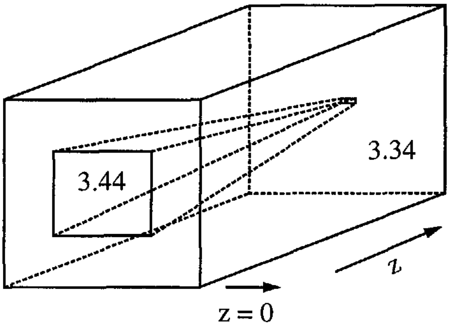
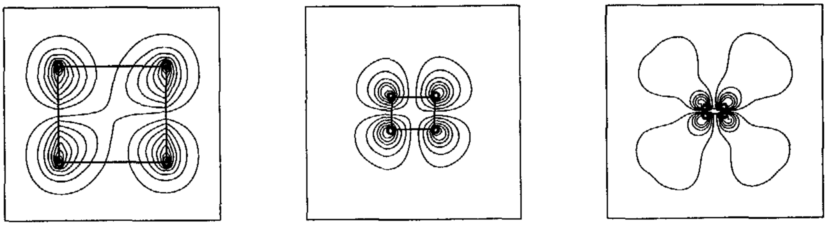


Figure 3.10 Schematic view of a tapered structure.

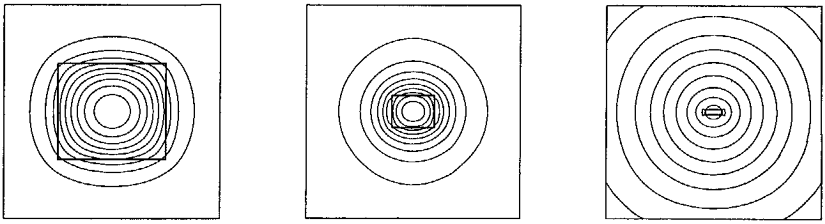
Our second structure for test computations is a tapered ridge waveguide structure (see Figure 3.10). The refractive indices of the core and the background are 3.44 and 3.34, respectively. The waveguide sizes taper both laterally and vertically along the propagation direction from 2.0 × 1.8 μm to 0.3 × 0.1 μm over a length of 53.125 μm. The transverse computational window of 4.0 × 4.0 μm is uniformly discretized with 81 × 81 grid points, thus $\Delta x = \Delta y = 0.05$ μm; the propagation step Δz is also chosen as 0.05 μm. The zero-order eigenmode field HE_{00} calculated from the FEM based program, is launched as the input in the plane $z = 0$, having a vacuum wavelength of 1.55 μm.

The three propagating electric field components E_x, E_y and E_z are shown in Figures 3.11, 3.12 and 3.13, respectively, where the reference propagation constant $\beta = k_0 n_0$ ($n_0 = 3.35$) and the scheme parameter value $\alpha = 0.7$ have been used. Nine iso-amplitude contours are drawn. where (a), (b) and (c) correspond to the propagating distances $z = 0$, 37.5 and 53.125 μm, corresponding to waveguide cross-sections of 2.0 × 1.8, 0.8 × 0.6 and 0.3 × 0.1 μm. The dominant component is E_y depicted in Figure 3.12, while the other two components are relatively small. We found that the field profile of E_y first contracts as the waveguide size decreases until it reaches a minimum in the case of Figure 3.12(b). Then the field gradually expands if the waveguide size decreases further to zero. The reason is that the zero-order mode size of a waveguide structure is closely linked to its waveguide size. Both a large and a small waveguide size correspond to large mode sizes, as discussed in [10]. This provides the principle of spot-size transformers with tapered waveguide structures. The simulation time for this structure took about 2.9 hours on the SUN Sparc 10/40 workstation.



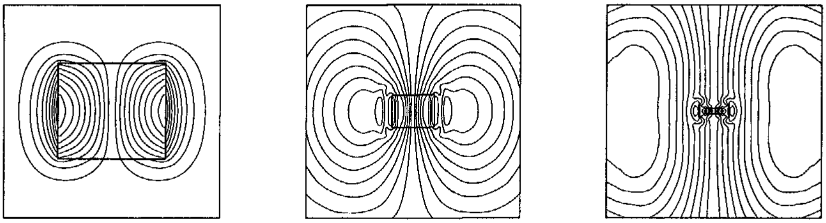
(a). z = 0 μm (b). z = 37.5 μm (c). z = 53.125 μm

Figure 3.11 Evolution of iso-amplitude contours of E_x for a HE_{00}-mode coupled into the tapered buried ridge waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ μm; $\alpha = 0.7$, $n_0 = 3.35$. The inner rectangles indicate the local waveguide cross-sections.



(a). z = 0 μm (b). z = 37.5 μm (c). z = 53.125 μm

Figure 3.12 Evolution of iso-amplitude contours of E_y for a HE_{00}-mode coupled into the tapered buried ridge waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ μm; $\alpha = 0.7$, $n_0 = 3.35$. The inner rectangles indicate the local waveguide cross-sections.



(a). z = 0 μm (b). z = 37.5 μm (c). z = 53.125 μm

Figure 3.13 Evolution of iso-amplitude contours of E_z for a HE_{00}-mode coupled into the tapered buried ridge waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ μm; $\alpha = 0.7$, $n_0 = 3.35$. The inner rectangles indicate the local waveguide cross-sections.

Since a small computational window is used, fields will extend to its boundaries. However, the fields remain undisturbed in the neighborhood of the boundaries, which indicates that the transparent boundary condition works appropriately.

For a longitudinally tapered structure, the propagation constant of a guided mode varies along the z-direction. An FEM analysis has shown that the effective refractive index of the HE_{00} guided mode changes gradually from 3.412 at the input to 3.34 at the output. Figure 3.14 gives the relationship between the power attenuation and the propagation distance for a HE_{00} guided mode field as the starting field. Our results show that the smallest power attenuation is obtained by using a variable reference refractive index n_0 which is equal to that of the guided mode for the corresponding local structure. If $n_0 = 3.35$, the power attenuation is relatively big in the beginning but increases only slowly later on since this value of n_0 means a considerable deviation from the effective index of the guided mode in the beginning. If $n_0 = 3.43$, the power attenuation is small in the beginning but increases rapidly since n_0 gradually deviates considerably from the effective index of the guided mode as the propagation continues. There are rapid increases of the power attenuation in Figure 3.14 after propagation for 45 μm. This occurs because we have chosen a small computational window of 4 × 4 μm. A certain amount of power has radiated through the boundary.

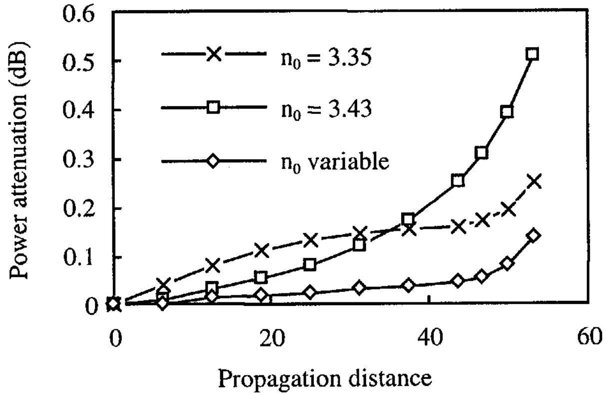


Figure 3.14 Relationship between the power attenuation and the propagation distance for a HE_{00}-mode field coupled into the buried ridge waveguide. $\Delta x = \Delta y = \Delta z = 0.05$ mm; $\alpha = 0.7$.

3.3.3 Polarization splitter

Integrated optical polarization splitters are important components for many applications, e.g. in coherent optical detection schemes, which use the concept of polarization diversity. Various concepts for polarization splitters have been presented in the literature [11-14]. Van der Tol *et al* [13] presented a new mode evolution polarization splitter in InGaAsP/InP based upon the waveguide birefringence which can be realized in a single etching step and which does not require metallization or other overlays. As a practical example, we next simulate this device and compare our results with the realized performance.

Structure of the splitter

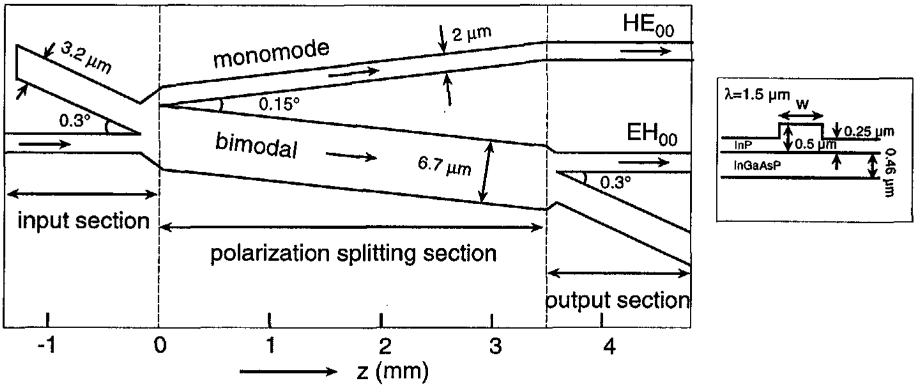


Figure 3.15 Schematic top view of the polarization splitter.

The polarization splitter consists of three sections: an input section, a splitting section, and an output section (see Figure 3.15). The input section with two asymmetrical monomode branches is used to inject the first order modes into the splitting section, thus functioning as a mode converter. The splitting section is constructed by combining a monomode waveguide and a bimodal waveguide to form an asymmetrical splitter. The monomode waveguide has a higher propagation constant for its HE$_{00}$-mode than that for the HE$_{01}$-mode of the bimodal waveguide, while at the same time it has a lower propagation constant for its EH$_{00}$-mode than that for the EH$_{01}$-mode of the bimodal waveguide. An injected mode will propagate preferentially as the mode in the Y-splitter which most closely matches its propagation constant [13]. Therefore, an HE$_{01}$-mode input will mainly lead to a HE$_{00}$-mode propagating in the monomode waveguide, while an EH$_{01}$-mode input will mainly lead to an EH$_{01}$-mode traveling through the bimodal waveguide. The output section is used to couple the EH$_{01}$-mode from the bimodal waveguide of the splitting section to the EH$_{00}$-mode of the monomode output waveguide, again having two asymmetrical monomode branches connected to the bimodal waveguide.

Simulation results

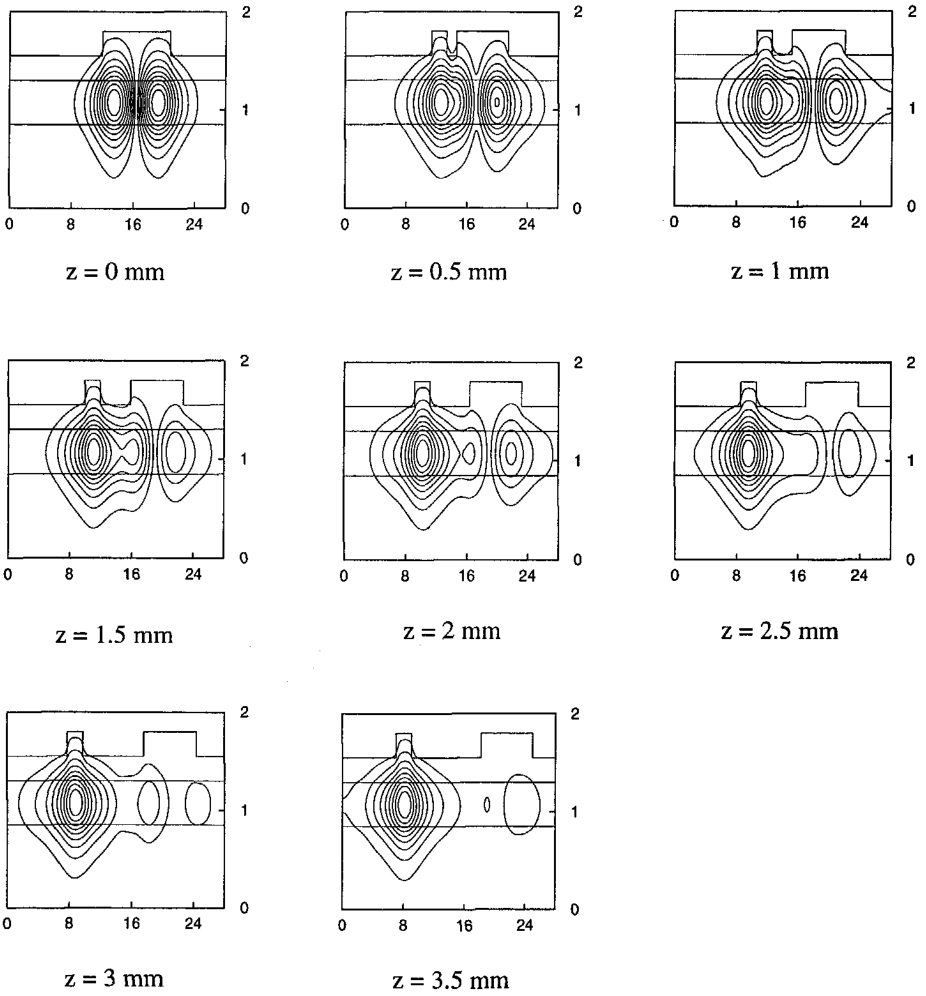
The first order modes calculated through the FEM analysis are the incident fields for the splitting section at z = 0 mm. $\Delta x = 0.05$ µm, $\Delta y = 0.2$ µm, $\Delta z = 0.1$ µm. The finite difference scheme parameter is chosen as 0.67, while the reference propagation constants are chosen as 13.78 and 13.74 μm^{-1} for the HE_{01}-mode and the EH_{01}-mode input, respectively. The wavelength equals 1.5 µm [14]. Figure 3.16 shows the field evolution of the dominant E_y-component, with the HE_{01}-mode as the input, while Figure 3.17 shows the field evolution of the dominant E_x-component, starting from the EH_{01}-mode. The polarization splitting effects can clearly be observed. The cross-talk values in dB, defined as the power ratio of the dominant (wanted) component in one branch and the non-dominant (unwanted) one in the other branch, are -13.8 and -11.1 dB for the HE_{01}-mode and the EH_{01}-mode inputs, respectively. The simulation time for such a splitter over a propagation distance of 3.5 mm took about 35 hours on the SUN Sparc 10/40 workstation.

Due to technical difficulties, the quoted experimental results [13] of the cross-talk values showed errors of ± 3 dB, which makes that both calculated results are in perfect agreement with the realized best performance of -12.0 and -13.1 dB [13], respectively.

The cross-talk depends on the branching angle and on the propagation constant difference of the modes. Actually, FEM analysis shows that the propagation constant difference between the HE_{00}-mode of the monomode waveguide and the HE_{01}-mode of the bimodal waveguide is only 1.47×10^{-4} μm^{-1}, and between the EH_{01}-mode of the bimodal waveguide and the EH_{00}-mode of the monomode waveguide is only 5.84×10^{-4} μm^{-1}. Such small differences correspond to a rather small branching angle in the order of 0.1°.

We have simulated another splitting section with the same parameter values as in Figure 3.15 except that the branching angle is increased to 0.3°. Then, the resulting cross-talk for the HE_{01}-mode input is increased -7.0 dB, demonstrating that a big angle can lead to the increase of cross-talk values. By application of a double masking technique, very small branching angles can be realized [13].

Figure 3.18 shows the relationship between the power attenuation of the propagating fields and the propagation distance in the splitting section. We notice that the EH_{01}-mode input experiences a lower power attenuation, however, its calculated cross-talk value is bigger than that of the HE_{01}-mode input. For the EH_{01}-mode input, the 3.5 mm section length can be decreased to about 2.5 mm without much change in performance (see Figure 3.17). However, 3.5 mm length is necessary for the HE_{01}-mode input (see Figure 3.16).



z = 0 mm

z = 0.5 mm

z = 1 mm

z = 1.5 mm

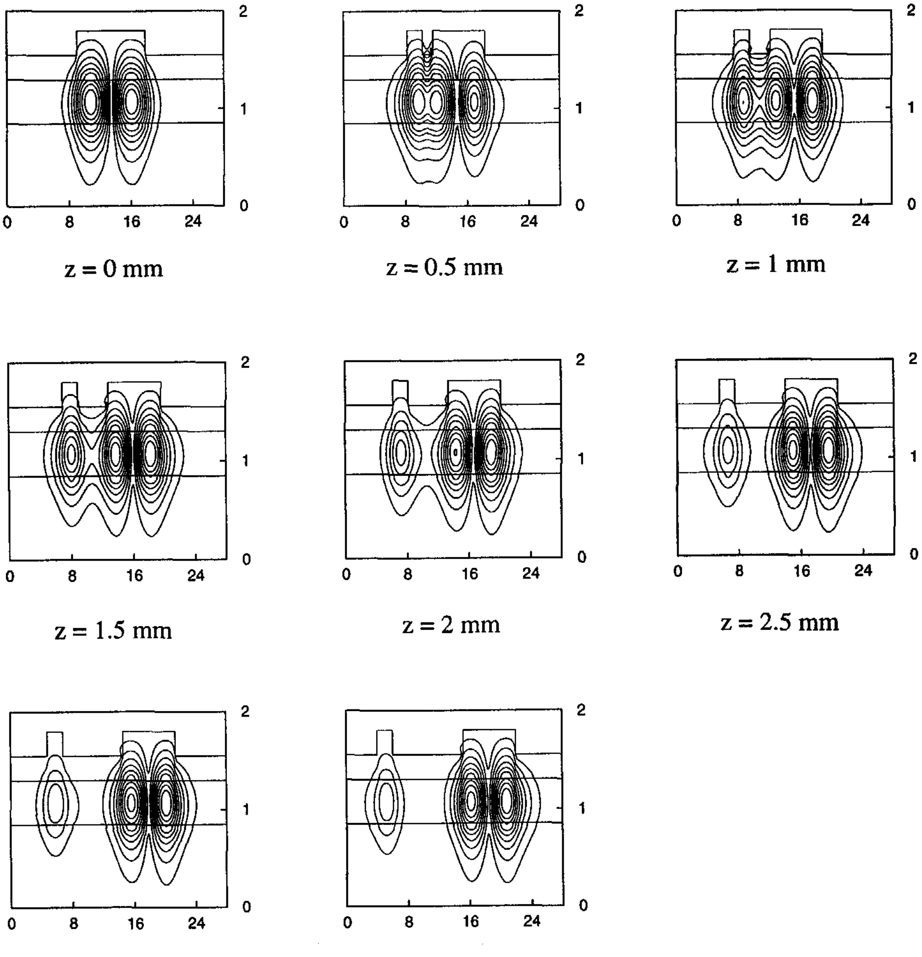
z = 2 mm

z = 2.5 mm

z = 3 mm

z = 3.5 mm

Figure 3.16 Evolution of the dominant component E_y in the splitting section with the HE_{01}-mode as input at z = 0 mm. The waveguide cross-sections are also indicated.



z = 0 mm

z = 0.5 mm

z = 1 mm

z = 1.5 mm

z = 2 mm

z = 2.5 mm

z = 3 mm

z = 3.5 mm

Figure 3.17 Evolution of the dominant component E_x in the splitting section with the EH_{01}-mode as input at z = 0 mm. The waveguide cross-sections are also indicated.

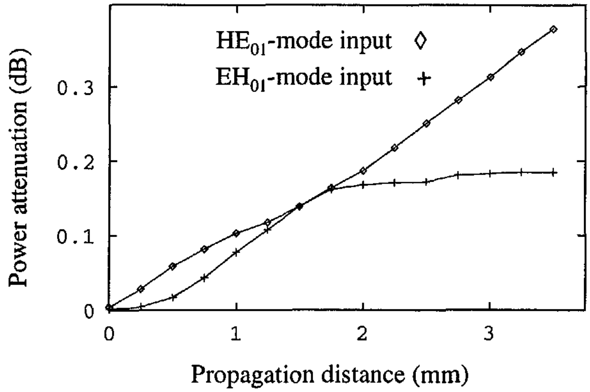


Figure 3.18 Relationship between the power attenuation of the propagating fields and the propagation distance in the splitting section with the injection of the first order modes at z = 0 mm.

Only the dominant field components have been shown in Figures 3.16 and 3.17. Our results show that the other components are relatively small. Therefore, scalar BPM results could also be used as an approximation for both cases of an HE_{01}-mode or an EH_{01}-mode input. However, the vectorial BPM is required if an arbitrarily polarized field is injected into the structure.

3.3.4 Polarization converter

Polarization converters act as the "retardation plates" of integrated optics. They are essential for a polarization diversity heterodyne receiver if it is integrated monolithically. They also can be used in a number of other applications, e.g. in polarization switching or scrambling. On InP, polarization converters using the electrooptic effect have been reported [15]. Alternatively, passive polarization converters made with periodic structures have also been mentioned in literature [16-17]. These components require less processing, since only waveguide sections have to be made. However, they usually are several mm long and have losses of a few dB. Recently, a new passive polarization converter on InP has been proposed [18]. It consists of a sequence of asymmetrical non-rectangular ridge waveguide segments with one angled sidewall. Adjacent segments have angled walls at opposite sides; at their junctions a coupling between the HE_{00} mode and the EH_{00} mode occurs. Choosing the correct segment length and waveguide parameters, a complete polarization conversion from the HE_{00} mode to the EH_{00} mode will occur after a given number of segments. The total length of such a device is about 1 mm and its total access loss is less than 0.5 dB.

Principle of the polarization converter

The proposed polarization converter is shown in Figure 3.19, it contains ridges having a straight and an angled side. The straight sidewalls are made by reactive ion etching (RIE). The angled sidewalls are produced through two selective wet chemical etching steps [19]: first on InP, then on InGaAsP. Figure 3.20 shows the waveguide's cross-sectional geometry in two adjacent sections. A lateral offset Δ between two adjacent ridges will be introduced, which can decrease the loss in the propagating field effectively.

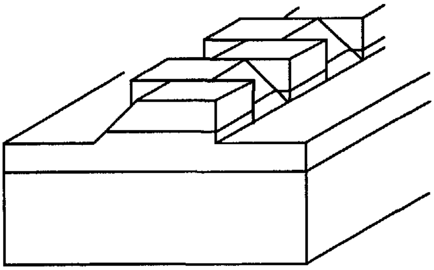


Figure 3.19 Periodic polarization converter with waveguide segments containing angled sides.

The angled sides cause a local tilt in the fields, which leads to bigger non-dominant field components. Figure 3.21 shows the non-dominant E_x-components of the HE_{00}-mode for the rectangular and non-rectangular structures calculated by use of the FEM-based program. The non-dominant E_x-component of the HE_{00}-mode for the rectangular structure is very small but symmetrical (Figure 3.21a), while in an angled side it is locally bigger and asymmetrical (Figure 3.21b).

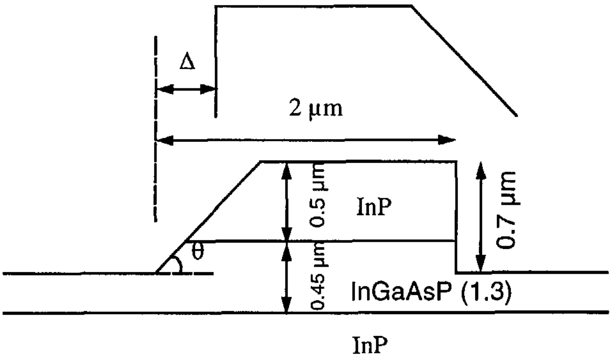


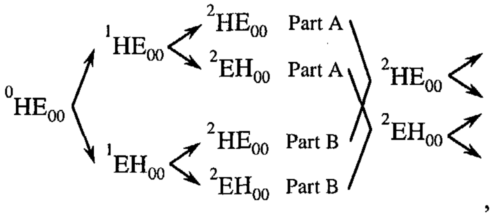
Figure 3.20 Waveguide cross-section of the polarization converter on InP.

(a)

(b)

Figure 3.21 Non-dominant E_x-components of the HE_{00} mode for (a) a rectangular structure and (b) a non-rectangular structure, both calculated by the FEM-based program. The facet angle θ is 45°.

We then can expect that at their junctions one propagating zero-order mode in the preceding section will excite two mutually perpendicular zero-order modes in the next section, as shown in the following:



,

where the number on the up left corners denotes the input field (if it is 0) and exited modes (1,2···). Part A and B denote excited modes from EH_{00}- and HE_{00}-modes, respectively. If the segments have a length that matches half the beatlength of the two mutually perpendicular zero-order modes, the two Parts A and B of $^{2}HE_{00}$-modes will be out of phase, and the two Parts A and B of $^{2}EH_{00}$-modes will be in phase. This phenomenon occurs at each later junctions. Consequently, a gradual polarization conversion will occur. After propagation through a certain number of segments, an incoming HE_{00}-mode will be converted completely to the EH_{00}-mode.

Simulation results

The considered wavelength is 1.5 μm. The facet angle is 45°. The cross-section of the polarization converter is discretized into grids with $\Delta x = \Delta y = 0.05$ μm. The propagation step size is also taken as 0.05 μm. The segment length of 90 μm is obtained by calculating $\pi/\Delta\beta$, where $\Delta\beta$ is the difference in the propagation constants of the HE_{00}- to EH_{00}-modes, calculated by the FEM-based program. The reference propagation constant is chosen as 13.72 μm^{-1} and the finite difference scheme parameter is chosen as 0.65 [20].

An HE_{00} modal field with a dominant E_y component is launched as the starting field. The field evolution of the E_x- and E_y-components is shown in Figure 3.22, where an offset of 0.2 μm has been used. The polarization conversion can be observed clearly from the evolution of field components. As the propagation continues, the E_y-component gradually decreases, while the E_x-component increases. After about 12 segments, a complete polarization conversion from HE_{00} to EH_{00} has been realized. If the field propagates further, the E_y-component will gradually increase while E_x decreases, repeating the process periodically after about 24 segments. The relative power of the HE_{00}- and EH_{00}-modes as functions of the propagation distance (measured in numbers of segments) is shown in Figure 3.23. The power attenuation after realizing a complete polarization conversion is about 0.5 dB.

The simulation time for such a converter to realize a complete polarization conversion took about 14 hours on the SUN Sparc 10/40 workstation.

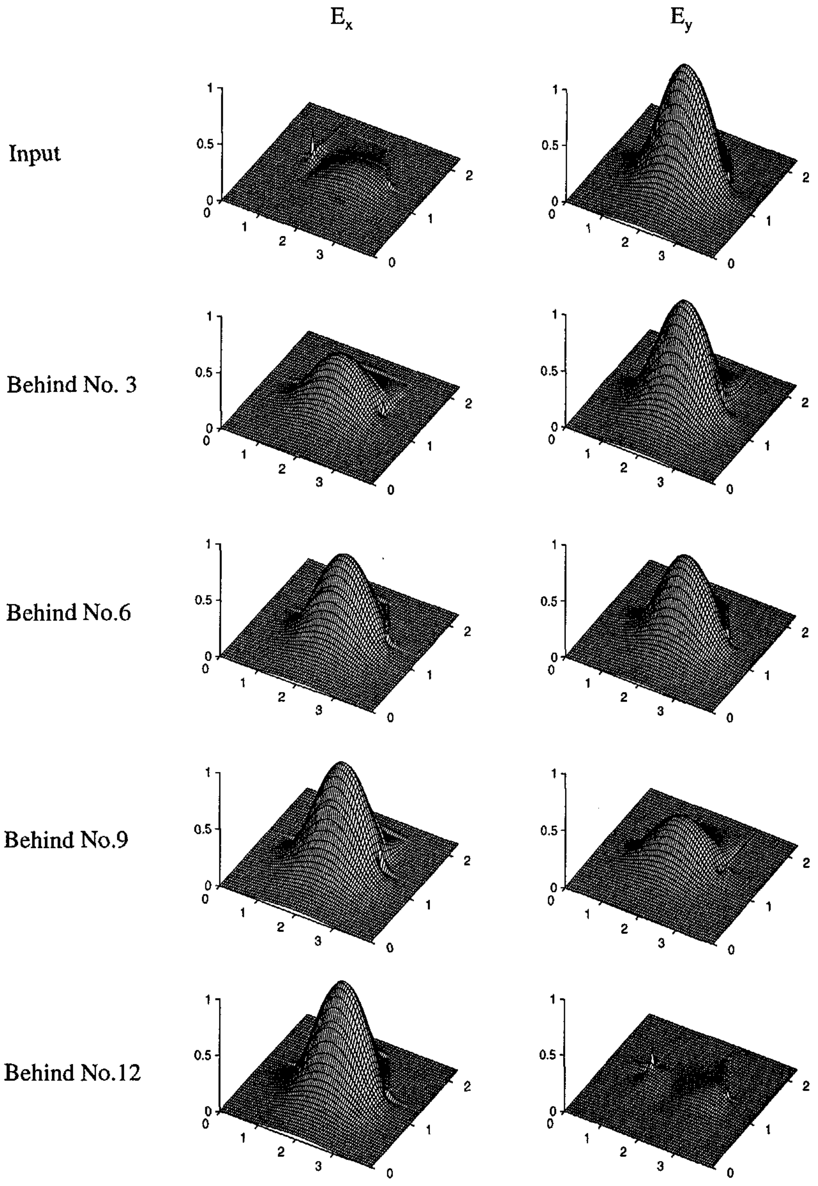


Figure 3.22 Evolution of the field propagating through a converter consisting of twelve segments.

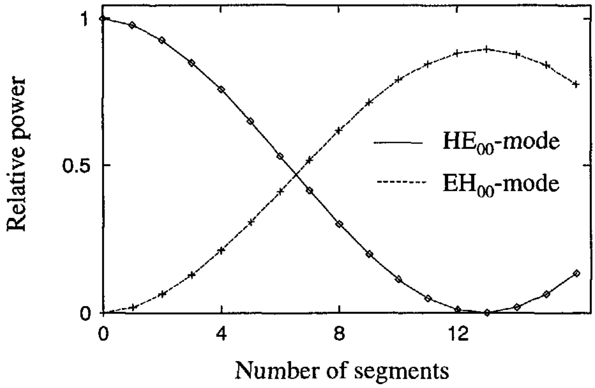


Figure 3.23 Relative powers of the HE$_{00}$- and EH$_{00}$-modes for the converter for a segment length of 90 μm and a lateral offset of 0.2 μm.

Lateral offset

The total access loss for a complete polarization conversion has been decreased by introducing a lateral offset between the two adjacent segments.

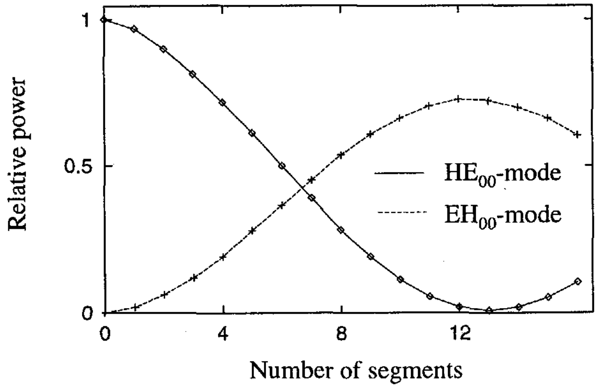


Figure 3.24 Relative powers of the HE$_{00}$- and EH$_{00}$-modes for the converter for a segment length of 90 μm but without a lateral offset.

Without this lateral shift, the total access loss for a complete conversion is obtained from Figure 3.24, being about 1.5 dB. With a 0.2 μm shift, however, this loss reduces to 0.5 dB (see Figure 3.23). Therefore, an appropriate lateral shift is important. We have simulated the polarization conversion for the structure with various offset values ranging from -0.5 to 1 μm. Our results show that the optimal lateral shift is about 0.2 μm.

Segment length

Another critical parameter is the segment length L. Since it is calculated from $\pi/\Delta\beta$, where $\Delta\beta$ is the difference in the propagation constants of the HE_{00}- and EH_{00}-modes in the angled structure, the value of L will depend on the calculation precision of the β of the guided modes. We experienced differences in $\Delta\beta$ if calculated by three different methods: one a FEM-based program, one a program based upon the "film mode matching" technique [21] and the third our vectorial FD-BPM to obtain $\Delta\beta$ by use of an extraction technique [8]. This technique can be used for a straight waveguide structure with a constant cross-section. The calculated $\Delta\beta$ is in the order of 0.035 μm^{-1}. Due to the inverse dependence of L on $\Delta\beta$ ($L = \pi/\Delta\beta$), a minor difference in $\Delta\beta$ values obtained from different methods will lead to a significant difference in L. The difference in L is in the order of 2 μm.

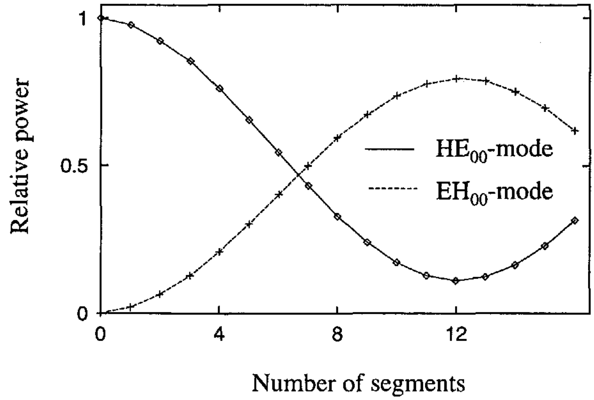


Figure 3.25 Relative powers of the HE_{00}- and the EH_{00}-modes for a segment length L = 87.5 μm, a deviation of 2.5 μm from the correct length, and with a lateral offset of 0.2 μm.

Our vectorial FD-BPM calculated the segment length as 90 μm. We then have investigated the influence of a segment length variation on the polarization conversion. The results show that the polarization conversion situation will deteriorate quickly if the length deviates from 90 μm. Figure 3.25 shows the situation with a segment length of 87.5 μm, a deviation of 2.5 μm. In that case a complete polarization conversion can never be realized. However, we found that if the deviation is less than 0.5 μm, the non-converted power is less than 1% after twelve segments. A precision of better than 0.1 μm can technologically be realized, if an electron beam pattern generation (EBPG) instrument and a double mask technique are used.

Method of overlap integrals

We now use another method, based upon overlap integrals [21], to analyze the fields in the above mentioned structure and we compare its results with that of our vectorial FD-BPM. The amplitude coupling coefficient between two complex guided modes is defined as:

$$C_{u,v} = \frac{I_{u,v}}{\sqrt{I_{u,u}\, I_{v,v}}}, \tag{3.8}$$

with

$$I_{u,v} = \left| \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(E_x^u H_y^v - E_y^u H_x^v \right) dx\, dy \right|, \tag{3.9}$$

where u and v denote two modes; E_x, E_y and H_x, H_y are the transverse components of their electric and magnetic fields, respectively. First, the guided mode fields are calculated by the FEM-based program. Then, the coupling efficiencies between modes of two adjacent segments are obtained from the overlap integrals. We again introduce a lateral offset of 0.2 μm. The calculated field coupling coefficients at one junction are 0.15 between the HE_{00}- and EH_{00}-modes of the adjacent segments, while 0.98 between the two HE_{00}-modes or two EH_{00}-modes of the neighboring segments. Subsequently, assuming that each segment has a correct length, the gradual change of power percentages of two zero-order modes as a function of the number of the segments can be calculated, using those field coupling efficiencies. The results are shown in Figure 3.26.

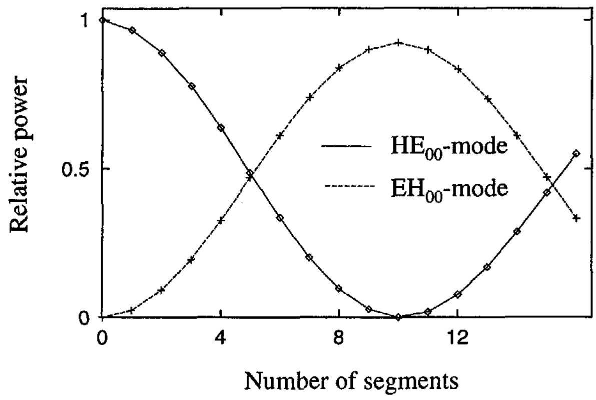


Figure 3.26 Relative powers of the HE$_{00}$- and the EH$_{00}$-modes for the converter calculated by the method of overlap integrals, with a lateral offset of 0.2 μm.

From Figure 3.26, a complete polarization conversion occurs for 10 segments instead of 12 segments from the FD-BPM analysis, mainly due to the fact that the method of overlap integrals ignores the influence of radiation, thus omitting extra phase shifts introduced by this radiation. The vectorial FD-BPM, which takes this radiation into account, therefore should lead to more exact results.

The polarization converter has been technologically realized [19], where a 90% conversion was obtained. The necessary number of segments for a maximal polarization conversion is between 10-13. The variation is due to technical difficulties. The total loss less than 1 dB has been obtained. These results have further proved the validity of our vectorial FD-BPM.

3.4 Summary

The considerations and procedures such as choices of the arrays, numerical treatment, double precision, debugging and manual source code optimization, have been discussed. A benchmark test including three computers was carried out. Our SUN workstation has the lowest speed. A 35 times speed increase has been obtained on the Cray C98/4256 using one of the four processors only. Optimization choices of compilers have been used, which speed up the simulation considerably.

Both exact guided mode fields calculated by the finite element method (FEM) and Gaussian beam fields have been launched as the input. The power attenuation coefficients and the field profiles have been calculated. The choices of the reference propagation constant, the finite difference scheme parameter and the step sizes are made on the basis of numerical analysis. An optimal reference propagation constant exists for each waveguide structure. For a tapered structure, this optimal value should be variable as the propagation continues. There also exists a critical value α_c of the finite difference scheme parameter. If $\alpha < \alpha_c$, the scheme is unstable. The validity and accuracy of the FD-BPM have been demonstrated by simulation of several particular waveguide structures, including a rib waveguide, a tapered ridge waveguide, a polarization splitter and a polarization converter. The simulation time for a single propagation step (0.05 μm) of a rib waveguide structure with a 61 × 61 grid division ($\Delta x = \Delta y = 0.05$ μm) was 3.3 seconds on a SUN Sparc 10/40 workstation.

3.5 References

[1] D. Li, H. van Brug, and H.J. Frankena: "Vectorial FD-BPM for 3-D waveguide structures," in *Linear and Nonlinear Integrated Optics*, Giancarlo C. Righini, David Yevick, Editors, Proc. SPIE **2212**, 36-46 (1994)

[2] *Cray-gebruikersgids*, SARA-publikatie 56, revisie A (1994).

[3] *SUN FORTRAN user's guide*, revision A (1991).

[4] M. Metcalf: *FORTRAN optimization*, Academic Press (1982).

[5] D. Schulz, C. Glingener, and E. Voges, "Novel generalized finite difference beam propagation method," submitted to *J. Quantum Electron.*.

[6] R.P. Ratowsky and J.A. Fleck, Jr., "Accurate numerical solution of the Helmholtz equation by iterative Lanczos reduction," *Optics Lett.*, **16**, 787-789 (1991).

[7] W.P. Huang, C.L. Xu, S.T. Chu, and S.K. Chaudhuri: "The finite-difference vector beam propagation method: analysis and assessment," *J. Lightwave Technol.*, **10**, 295-304 (1992).

[8] M.D. Feit and J.A. Fleck, Jr.: “Computation of mode properties on optical fiber waveguides by a propagating beam method,” Appl. Opt., **19**, 1154-1164 (1980).

[9] F.J. Harris: “On the use of windows for harmonic analysis with the discrete Fourier transform,” *Proc. IEEE*, **66**, 51-84 (1978).

[10] H. Kogelnik: in *Integrated Optics*, edited by T. Tamir, the 2nd edition, Springer-Verlag , 28 (1979).

[11] A.R. Vellekoop, and M.K. Smit: "A small-size polarization splitter based on a planar optical phased array," *J. Lightwave Technol.*, **8**, 118-124 (1990).

[12] P. Albrecht, M. Hamacher, H. Heidrich, D. Hoffman, H.-P. Nolting, and C.M. Weinert: "TE/TM mode splitters on InGaAsP/InP," *IEEE Phot. Technol. Lett.*, **2**, 114-115 (1990).

[13] J.J.G.M. van der Tol, J.W. Pedersen, E.G. Metaal, Y.S. Oei, H. van Brug, and I. Moerman: "Mode evolution type polarization splitter on InP/InGaAsP," *IEEE Phot. Technol. Lett.*, **5**, 1412-1414 (1993).

[14] D. Li, H. van Brug, and H.J. Frankena: “Vectorial beam propagation simulation of a polarization splitter on InP/InGaAsP,” *The 21st European Conference on Optical Communication*; September, 1995, Brussels, Belgium (accepted).

[15] M. Schlak, C.M. Weinert, P. Albrecht, and H.-P. Nolting: “Tunable TE/TM-mode converter on (001)-InP-substrate,” *IEEE Phot. Technol. Lett.*, **3**, 15-16 (1991).

[16] Y. Shani, R. Alfernese, T. Koch, U. Koren, M. Oren, B.I. Miller, and M.G. Young: “Polarization rotation in asymmetric periodic loaded rib waveguides,” *Appl. Phys. Lett.*, **59**, 1278-1280 (1991).

[17] H. Heidrich, P. Albrecht, M. Hamacher, H.-P. Nolting, H. Schroeter-Janβen, and C.W. Weinert: “Passive mode converter with a periodically tilted InP/GaInAsP rib waveguide,” *IEEE Phot. Technol. Lett.*, **4**, 34-36 (1992).

[18] J. van der Tol, F. Hakimzadeh, J. Pedersen, D. Li, and H. van Brug: “A new short and low-loss passive polarization converter on InP,” *IEEE Phot. Technol. Lett.*, **7**, 32-34 (1995).

[19] J.J.G.M van der Tol, J.W. Pedersen, E.G. Metaal, Y.S. Oei, F.H. Groen and I. Moerman: “Efficient short passive polarization converter,” in *Proceedings of the 7th European Conference on Integrated Optics*, 319-322 (1995).

[20] D. Li, H. van Brug, H.J. Frankena, J.J.G.M. van der Tol, and J.W. Pedersen: "Vectorial beam propagation simulation of a novel polarization conversion waveguide structure," in *Fibre Optic Network Components*, S.I. Najafi, H. Porte, Editors, Proc. SPIE **2449**, 266-272 (1995)

[21] A.S. Sudbø: "Film mode matching: a versatile numerical method for vector mode field calculations in dielectric waveguides," *Pure Appl. Opt.*, **2**, 211-233 (1993).

Chapter 4

Design of tapered waveguide structures

In Chapter 1, it has been pointed out that the modal field of a single-mode fiber can be approximated by a Gaussian distribution, which has a beam diameter (1/e-width of the amplitude profile) of about 8-11 μm, while the zero-order modal field in an InGaAsP waveguide on an InP substrate is usually asymmetrical and has a mode size of about 1-3 μm. In order to improve the fiber-chip coupling efficiency, various methods have been proposed. A properly designed three-dimensional tapered waveguide structure could be an ideal tool for matching these fields. In this chapter, we use our fully vectorial FD-BPM to design tapered waveguide structures on InGaAsP/InP, which can serve as mode size transformers for an effective fiber-chip coupling.

4.1 Basic considerations

An ideal mode size transforming tapered waveguide should modify the field profile of an incident modal field adiabatically. This implies that the local mode at the output of the waveguide contains all the incoming power. An actual situation usually suffers from a certain power loss due to the gradual change of the waveguide cross section over a finite length; as the propagation proceeds, this causes power to radiate out of the structure.

The fiber-chip coupling efficiency depends on two kinds of power loss (reflection loss can be effectively decreased using e.g. anti-reflection coating): one is the propagation loss in the tapered waveguide structure; the other is the overlap loss due to a difference in modal fields between fiber and waveguide fields. Therefore, to obtain the highest possible coupling efficiency, 3D tapered structures are necessary, which means that waveguide structures are tapered both vertically and horizontally along the propagation direction such that the mode field of the chip waveguide is transformed to match the circularly symmetrical field of the single mode fiber. In addition, the propagation loss in the designed tapered structures should be low.

Another point in designing is that the structures should be technologically realizable. We have taken into account the following aspects:

> ***Easy coupling of the fiber to the chip***: Coupling configurations should have a workable alignment tolerance. Based upon this, we do not use tapered fibers, which entail stringent alignment requirements in all directions. Instead, we use flat-end single mode fibers, such that a simple butt coupling can be used. This requires that a field propagating through the chip should have a relatively large mode size in the coupling area.
>
> ***Possibility to fabricate***: Tapered structures should be fabricated with technologically realizable process steps at acceptable costs.
>
> ***Small size***: The total length of a tapered structure should be less than 1.5 mm, since in our laboratory a normal chip has a size of about 10 × 10 mm and other structures must also be placed on the chip.

4.2 Suggested structure

The starting point is an InGaAsP waveguide with a bandgap of 1.3 μm ($Q_{1.3}$) on an InP substrate ($In_{0.71}Ga_{0.29}As_{0.62}P_{0.38}$). The waveguide width and thickness are 2 and 0.6 μm, respectively. Considering the technological possibilities, horizontal and vertical tapering occurs in separate sections. The waveguide is tapered to a smaller thickness in the vertical direction, but horizontally the width increases.

Such a structure is sketched in Figure 4.1 and consists of five sections. The straight Section A connects to a waveguide on the chip and Section E couples to a single mode fiber. Sections B and D include laterally and vertically tapered profiles, respectively. The straight Section C is a transitional area between two tapers; it can be very short (even negative) and is designed for technological convenience, since the two profiles of the adjacent sections are made by different processes making a perfect direct connection difficult. Finally, the whole structure is covered with a thick high index layer, which should match the index of the substrate in order to obtain a symmetrical field distribution as required for an efficient fiber-chip coupling.

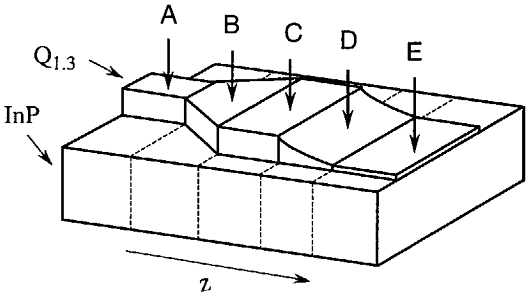


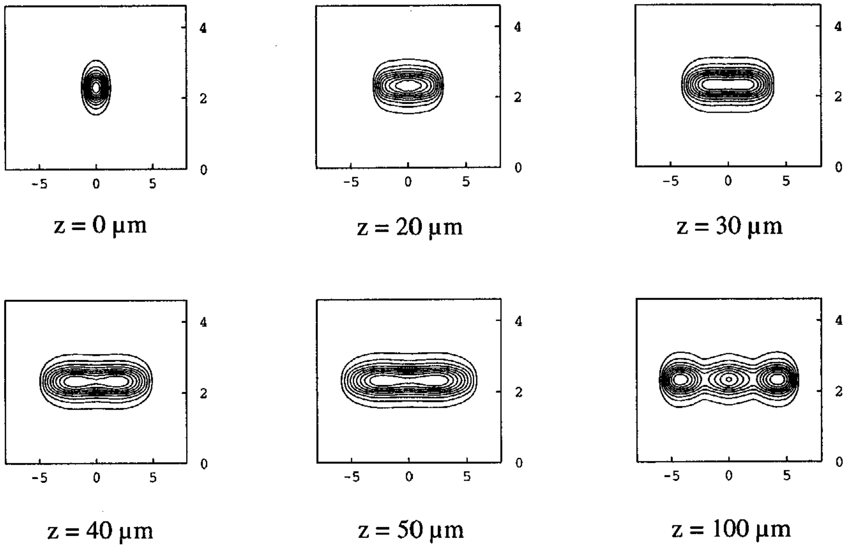
Figure 4.1 Schematic view of a 3D tapered waveguide

This cover layer is not drawn in Figure 4.1. We will see later that its index value and thickness have critical effects on the profile of the output field, thus dramatically influencing the coupling efficiency from the waveguide to a single mode fiber.

4.3 Simulation results

4.3.1 Laterally tapered section

Since Section A connects to the single mode chip waveguide, only the zero order guided mode will enter it. We start the simulation of the field propagation from the junction between Sections A and B by launching the HE_{00}-mode of Section A into the laterally tapered Section B, thus assuming Section A to be long enough such that possible transitional fields are radiated away. The effect of the lateral taper is to transform the chip mode field laterally such that the field at the end of this section has a lateral size matching that of the fiber. Based upon these considerations, the section is linearly tapered from 2 to 12 µm. The cover layer is chosen to have a thickness of 5 µm and a refractive index the same as that of the substrate (InP) in order to obtain a vertically symmetrical mode profile. The waveguide ($Q_{1.3}$) has a thickness of 0.6 µm. We first simulate the field propagation through the section with a length of 50 µm at 1.55 µm wavelength.



z = 0 µm

z = 20 µm

z = 30 µm

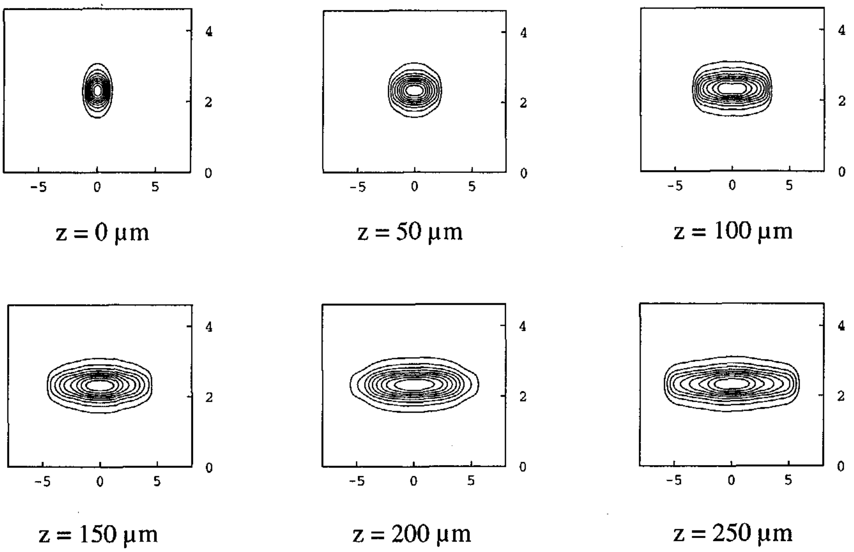
z = 40 µm

z = 50 µm

z = 100 µm

Figure 4.2 Evolution of iso-amplitude contours of E_y for the HE_{00}-mode coupled into the laterally tapered section. $\Delta x = \Delta z = 0.05$ µm; $\Delta y = 0.2$ µm; $\alpha = 0.65$, n_0 is variable. Section B has a length of 50 µm. After this, the field enters Section C.

The computational window has a size of 4.6 × 16 µm. We have chosen $\Delta x = \Delta z = 0.05$ µm; $\Delta y = 0.2$ µm. The scheme parameter is chosen as $\alpha = 0.65$. The reference propagation constant is variable as the propagation proceeds, changing from 13.341 μm^{-1} in the beginning to 13.401 μm^{-1} at the end. The reason to choose these values is because they correspond to the local effective refractive indices, yield a negligible spurious numerical loss and ensure the stability of the numerical scheme, as discussed in Sections 3.3.1 and 3.3.2 and verified by numerous calculations for different parameter values. The field evolution of the dominant component E_y is shown in Figure 4.2. We can see that the field changes as the waveguide widens. Higher order modes are seen to grow gradually from z = 30 µm onward. After z = 50 µm, the propagating field enters Section C, where the waveguide has a constant width of 12 µm. Higher order modes of this section can also be seen clearly in the field distribution at z = 100 µm. Since the presence of higher order modes lead to a low coupling efficiency with a single mode fiber, this should be avoided. Our results indicate that the length of Section B as 50 µm is too short to avoid those higher order modes.



z = 0 µm

z = 50 µm

z = 100 µm

z = 150 µm

z = 200 µm

z = 250 µm

Figure 4.3 Evolution of iso-amplitude contours of E_y for the HE_{00}-mode coupled into the laterally tapered waveguide. $\Delta x = \Delta z = 0.05$ µm; $\Delta y = 0.2$ µm; $\alpha = 0.65$, n_0 is variable. Section B has a length of 200 µm. After this, the field enters Section C.

Next, we simulate another laterally tapered section with the same parameters as in Figure 4.2 except that its length is now 200 µm. The results are shown in Figure 4.3. As the propagation proceeds, the field gradually changes without much evidence of higher order modes. After z = 200 µm, the field propagates further through Section C in which the waveguide width is constant at 12 μm. As a result, it can be seen that the field distribution (z = 250 μm) changes only slightly, which means that a stable field has almost been obtained after z = 200 μm. The propagation loss is calculated to be less than 0.1 dB after propagation over 500 μm, as shown in Figure 4.4. In addition, our computations show that the longer Section B is, the weaker the excited higher order mode fields will be. To ensure that excited higher order modes are negligible, the length of Section B should be longer than 200 μm for such a kind of structure (w_0 = 2 μm, w_1 = 12 μm).

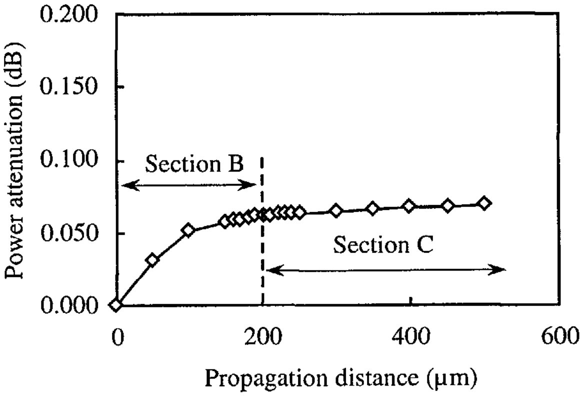


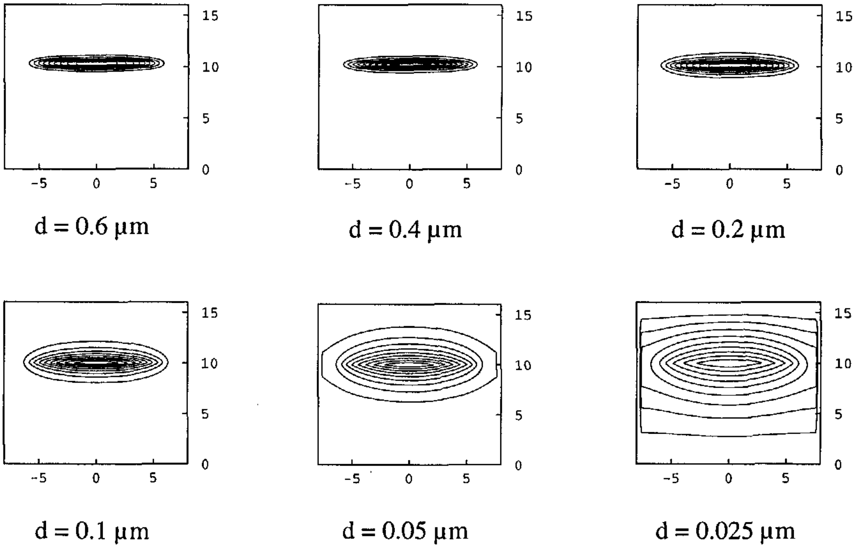
Figure 4.4 Relationship between the power attenuation and the propagation distance for the HE_{00}-mode coupled into the laterally tapered waveguide. Δx = 0.05 μm; Δy = 0.2 μm; Δz = 0.025 μm; α = 0.65, n_0 is variable. n_c = 3.169. Section B has a length of 200 μm. After this, the field enters Section C.

If the refractive index of the cover layer is 3.15 instead of 3.169, our calculations show that the field distribution hardly changes. Therefore, the laterally tapered section has a reasonable tolerance for index changes of the cover layer. In addition, the field distribution in a laterally tapered section is usually symmetrical in the horizontal direction. This kind of structure can be realized by conventional lithographic processes.

4.3.2 Vertically tapered section

The vertically tapered section (Section D) for the fiber-chip coupling is difficult to fabricate and has stringent requirements for the waveguide thickness and index value of the cover layer. To increase the mode size in the vertical direction considerably, the waveguide has to be tapered to a rather small thickness.

Figure 4.5 shows field contours of the dominant component E_y of the HE_{00} mode for structures with the same parameters but different waveguide thicknesses, calculated by the FEM-based program. The cover layer has a thickness of 5 μm and an index of 3.169 at 1.55 μm wavelength (same as the substrate). The waveguide width is chosen as 12 μm, in order to obtain a lateral mode size of about 10 μm width matching that of the fiber for a high coupling efficiency. The mode size in the vertical direction hardly changes if the waveguide thickness d ≥ 0.2 μm. However, if the waveguide thickness decreases even further, the mode size increases very rapidly. Therefore, the waveguide can be tapered fast in the beginning, but should be tapered more slowly later on in order to obtain a gradual change of the propagating field, ensuring a low radiation loss.



d = 0.6 μm

d = 0.4 μm

d = 0.2 μm

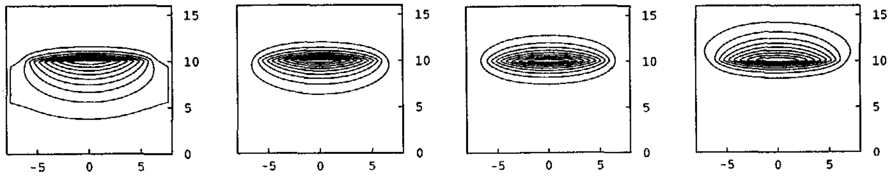
d = 0.1 μm

d = 0.05 μm

d = 0.025 μm

Figure 4.5 Influence of the waveguide thickness on the field distribution of the dominant component E_y of the HE_{00}-mode, calculated by the FEM-based program.

The cover index also becomes more important if the waveguide thickness is very small. Figure 4.6 shows the influence of the cover index on the field distribution of HE_{00} modes for a waveguide thickness of 0.075 μm. With a slight change of index value, the field distribution changes dramatically. The cover index must match that of the substrate in order to obtain a vertically symmetrical field distribution, which is important for an effective fiber-chip coupling.

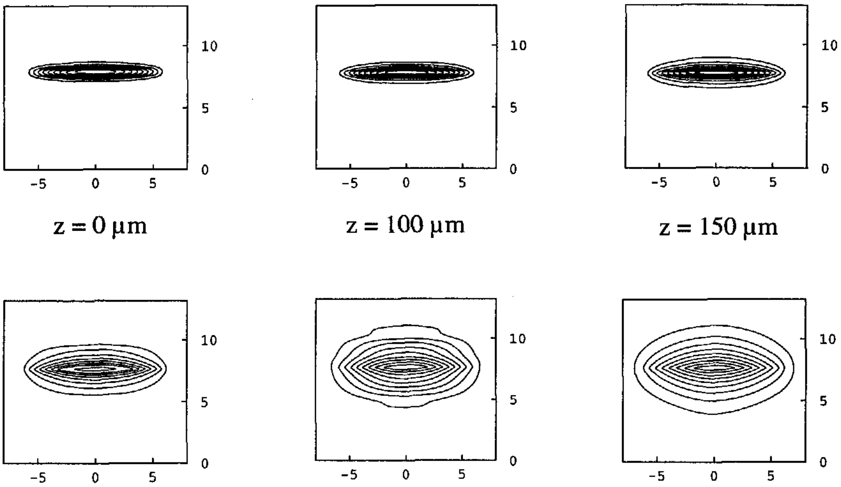


$n_c = 3.15$ $n_c = 3.16$ $n_c = 3.17$ $n_c = 3.18$

Figure 4.6 Influence of the refractive index value of the cover on the field distribution of the dominant component E_y of the HE_{00}-mode, calculated by the FEM-based program. The waveguide thickness is 0.075 μm.

We next simulate the field propagation in a vertically tapered section with a 5 μm thick cover layer of the same material as the substrate (InP). The waveguide is linearly tapered from d_0 = 0.6 μm to d_1 = 0.05 μm over a length of 200 μm. The waveguide width is 12 μm. We launch the HE_{00}-mode into the section from the thick end at the junction between C and D. The reference propagation constant is variable as the propagation proceeds, changing from 13.401 μm^{-1} in the beginning to 12.858 μm^{-1} in the end. The structure in the vertical direction is non-uniformly discretized since the waveguide thickness is very small at one end of the section, thus Δx is in the range of 0.025-0.2 μm. Δz = 0.025 μm; Δy = 0.2 μm; α = 0.65. The field evolution of the dominant component E_y is shown in Figure 4.7.

From Figure 4.7 it can be seen that the field gradually changes as the waveguide thickness becomes smaller. However, it changes slowly in the beginning. Only when the waveguide is tapered to a smaller thickness, the field expands more pronouncedly, which is consistent with the results of our FEM analysis as presented in Figure 4.5. After z = 200 μm, the propagating field enters Section E and the waveguide thickness of $d_l = 0.05$ μm remains constant. The field distribution, however, still changes towards that of the guided mode as the propagation proceeds (e.g. z = 750 μm), which indicates that where the taper terminates the field is not yet stable.



z = 0 μm

z = 100 μm

z = 150 μm

z = 180 μm

z = 200 μm

z = 750 μm

Figure 4.7 Evolution of iso-amplitude contours of E_y for the HE_{00}-mode coupled into the vertically tapered waveguide. Δx is non-uniform. $\Delta z = 0.025$ μm; $\Delta y = 0.2$ μm; $\alpha = 0.65$, n_0 is variable. $n_c = 3.169$. Section D has a length of 200 μm. After this, the field enters Section E.

Figure 4.8 shows the relationship between the power attenuation and the propagation distance. The taper starts at z = 0 and terminates at z = 200 μm. A rapid power attenuation occurs between z = 200 μm and 400 μm, where the field already propagates in Section E, which means that there the field is still not yet stable and radiation keeps escaping out of the computational window. The total propagation loss is 0.55 dB after propagation over 750 μm.

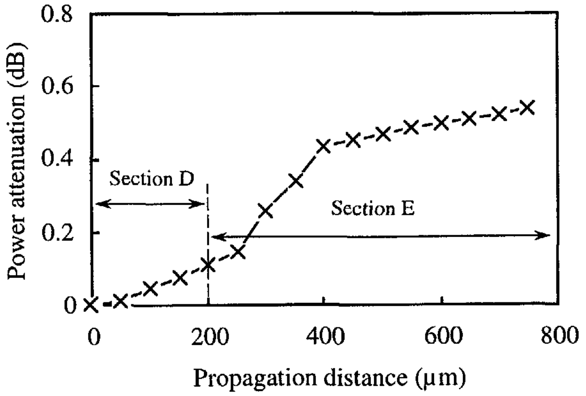


Figure 4.8 Relationship between the power attenuation and the propagation distance for the HE_{00}-mode coupled into the vertically tapered waveguide. Δx is non-uniform. $\Delta z = 0.025$ μm; $\Delta y = 0.2$ μm; $\alpha = 0.65$, n_0 is variable. $n_c = 3.169$. Section D has a length of 200 μm. After this, the field enters Section E.

From the results shown above, we arrive at the following conclusions. First, if the vertical taper is linearly tapered, an adiabatic taper (no radiation loss) will correspond to a rather large length, even if the total waveguide thickness change is only about 0.5 μm. This is very different from a laterally tapered section. Since the section length has a certain limit in order to restrict the total device size, a small propagation loss is inevitable. Second, the propagating field is not immediately stable after the vertically tapered section terminates as can be seen from the propagation loss, but is still varying over a certain distance in the next section. Within this distance, the wavefront of the total field will not be flat, leading to a low coupling efficiency with a fiber. Therefore, a minimal length for Section E, as in Figure 4.1, is necessary.

In fact, if the waveguide thickness is small, the fundamental mode is guided by a high index region, that is thin compared to the mode size. A small thickness variation will have a dramatic influence on the mode shape and mode width and thus will imply a considerable radiation loss [1]. Therefore, an adiabatic taper requires that the waveguide should be tapered rather slowly if the mode is close to cut-off. A semi-quantitative but, due to its simplicity, attractive criterion has been published by J.D. Love *et al* [2]. They point out that the local taper length-scale of an adiabatic taper has to be larger than the coupling length between the fundamental local mode and the dominant mode for radiation loss. In that case negligible radiation loss will occur. Our results indicate that a linear taper length of 200 μm is too short in the given example.

Figure 4.9 shows the relationship between the power attenuation and the propagation distance relating to the vertically tapered Section D with different lengths. They all start at z = 0 μm but terminate at z = 100, 200 and 300 μm, respectively. Subsequently, the waveguide thickness remains constant at 0.05 μm in Section E. The longer Section D is, the lower its power attenuation will be. For a section length of 300 μm, the total power loss is decreased from 0.55 to 0.41 over a total propagation distance of 750 μm (after z = 300 μm, the field enters Section E).

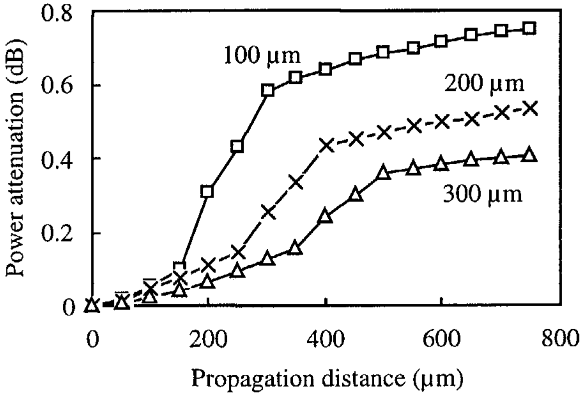


Figure 4.9 Relationship between the power attenuation and the propagation distance of the vertically tapered section D for different section lengths, for the HE_{00}-mode coupled into the waveguide. Δx is non-uniform. Δz = 0.025 μm; Δy = 0.2 μm; α = 0.65, n_0 is variable. n_c = 3.169. Section D has lengths of 100, 200 and 300 μm, respectively, and starts from z = 0 μm. After Section D, fields enter Section E.

However, the tapered section can not be too long, because it makes the total device size unacceptably long. Next, we change the tapered profile from a single linear section to three consecutive sections, while keeping the total section length unchanged at 300 μm.

The idea is to taper the structure fast in the beginning and slowly later on. The tapered profile is shown in Figure 4.10. The simulation results are shown in Figure 4.11.

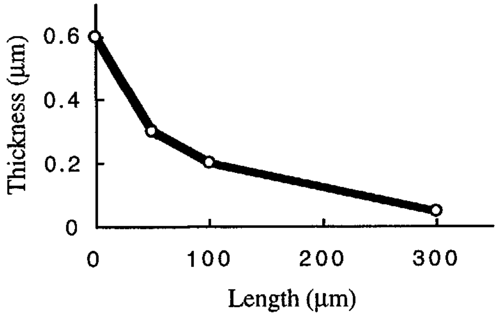


Figure 4.10 A tapered profile containing three linear parts.

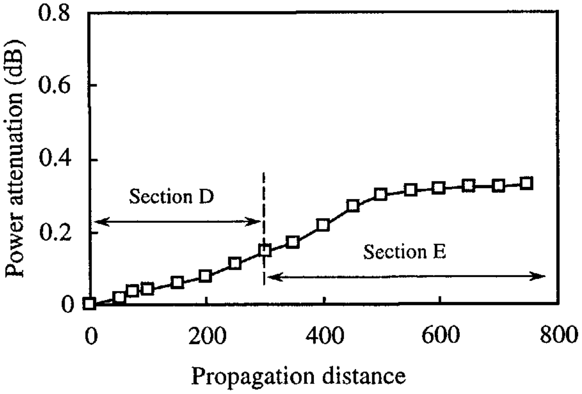
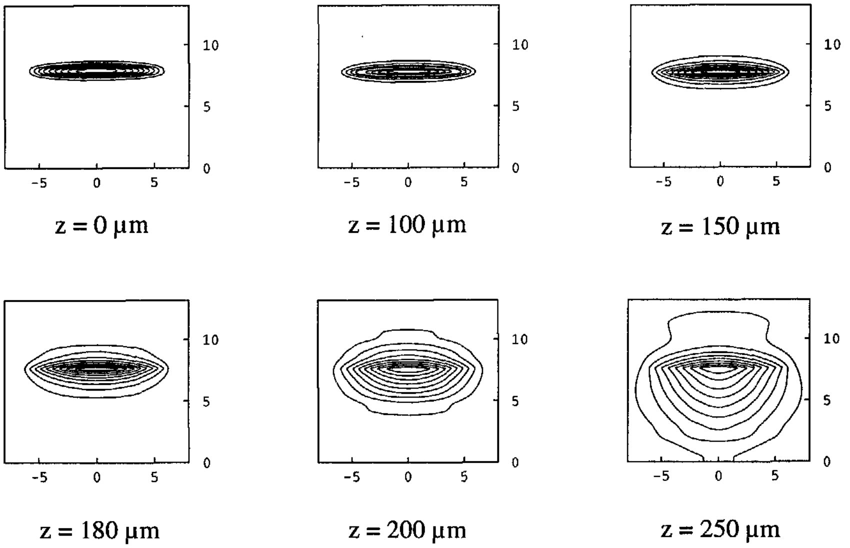


Figure 4.11 Relationship between the power attenuation and the propagation distance for the HE_{00}-mode coupled into the vertically tapered waveguide as shown in Figure 4.9. Δx is non-uniform. $\Delta z = 0.025$ μm; $\Delta y = 0.2$ μm; $\alpha = 0.65$, n_0 is variable, $n_c = 3.169$.

From Figure 4.10, we see that the waveguide is first tapered from 0.6 to 0.3 μm over a length of 50 μm, then from 0.3 to 0.2 μm also over 50 μm, and finally from 0.2 to 0.05 μm over 200 μm. In that case, its power attenuation decreases appreciably as compared to the single linear taper (see Figure 4.11). The total power loss decreases from 0.41 to 0.33. In fact, the attenuation can be decreased further by optimizing the tapered profile. However, such nonlinear profiles are usually difficult to realize by the available technologies, e.g. atomic beam etching.

Finally, we changed the index value of the cover layer from 3.169 to 3.16 and tapered the waveguide from 0.6 to 0.075 μm, while keeping all other parameters the same as in Figure 4.7. This way, the cover corresponds to a material choice different from InP, e.g. silicon nitride, which can be deposited by PECVD (see Section 5.3). The results are shown in Figure 4.12.



z = 0 μm

z = 100 μm

z = 150 μm

z = 180 μm

z = 200 μm

z = 250 μm

Figure 4.12 Evolution of iso-amplitude contours of E_y for the HE_{00}-mode coupled into the vertically tapered waveguide. Δx is non-uniform. $\Delta z = 0.025$ μm; $\Delta y = 0.2$ μm; $\alpha = 0.65$, n_0 is variable. $n_c = 3.16$. The taper length is 200 μm. After this, the field enters Section E.

In Figure 4.12, the field propagation is quite similar to that in Figure 4.7 up to z = 150 μm. Then, the field distribution in the vertical direction gradually becomes asymmetrical with decreasing waveguide thickness. Since a single mode fiber has a circularly symmetrical modal field, a low coupling efficiency between the modal fields of the fiber and the tapered structure is expected (see Section 4.4). Therefore, the cover index should match that of the substrate in order to obtain a very high fiber-chip coupling efficiency. This is the reason why we have to control precisely the refractive index of a deposited silicon nitride layer (see Section 5.3).

4.4 Fiber-chip coupling

In the following, we derive an expression for calculating the fiber-chip coupling efficiency.

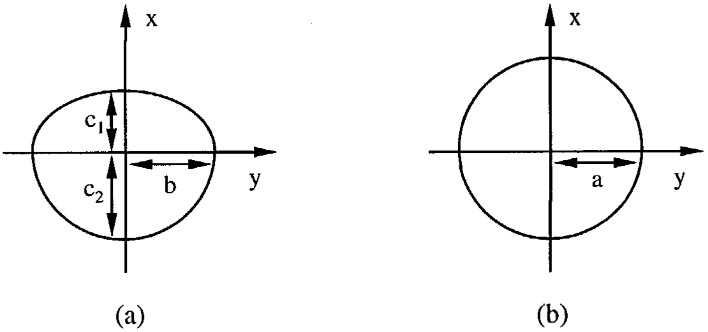


Figure 4.13 Schematic view of iso-amplitude contours with 1/e amplitudes of (a) the waveguide and (b) the fiber modal fields.

Usually, a zero-order waveguide modal field with a flat planar wavefront is symmetrical in the lateral direction, while vertically asymmetrical (see Figure 4.13a). The field profile can be expressed by a combination of two Gaussian profiles [3]:

$$G = G_0 \exp\left[-\frac{x^2}{c_j^2} - \frac{y^2}{b^2}\right], \tag{4.1}$$

where G_0 is the amplitude, b and c_j are half widths at 1/e amplitudes in the lateral and vertical directions, respectively. For $x > 0$ we use the subscript $j = 1$; for $x < 0$, we will write $j = 2$. Similarly, a horizontally or vertically polarized fiber modal field with a flat wave front can be expressed by a single Gaussian profile, thus:

$$F = F_0 \exp\left[-\frac{(x-\Delta)^2 + y^2}{a^2}\right], \tag{4.2}$$

where a is its half width at 1/e amplitude and Δ is an introduced vertical offset between the fiber and the waveguide.

The coupling efficiency can be derived as (see Appendix E):

$$\eta = \frac{2b(f_1 + f_2)^2}{(a^2 + b^2)(c_1 + c_2)}, \tag{4.3}$$

with

$$f_1 = \frac{a\, c_1}{\sqrt{a^2 + c_1^2}} \exp\left(-\frac{\Delta^2}{a^2 + c_1^2}\right)\left[1 - \mathrm{erf}\left(-\frac{c_1 \Delta}{a\sqrt{a^2 + c_1^2}}\right)\right], \tag{4.4a}$$

$$f_2 = \frac{a\, c_2}{\sqrt{a^2 + c_2^2}} \exp\left(-\frac{\Delta^2}{a^2 + c_2^2}\right)\left[1 + \mathrm{erf}\left(-\frac{c_2 \Delta}{a\sqrt{a^2 + c_2^2}}\right)\right], \tag{4.4b}$$

where erf denotes the error function. Practically, a tapered waveguide usually has an asymmetrical modal field in the vertical direction. Therefore, a vertical offset is necessary. As an example, if $a = 5$ µm, $b = 5$ µm, $c_1 = 1$ µm and $c_2 = 4$ µm, the calculated coupling efficiency is shown in Figure 4.14. The maximal coupling efficiency is 78% (i.e. 1.1 dB loss), requiring a vertical offset of 1.5 µm.

From Eq.(4.3), we can now predict coupling efficiencies between a fiber and designed tapers. If there is no cover layer grown on the structure as indicated in Figure 4.1, thus $c_1 = 0$, the maximal coupling efficiency can amount to 74%, obtained for a vertical offset of 2.5 µm and with $a = b = c_2$, provided that no propagation loss occurs. Similarly, if a cover with the proper refractive index and thickness is supplied, the maximal efficiency can reach 100% with $a = b = c_1 = c_2$ and no propagation loss. Thus, a properly designed cover plays an important role in optimizing the coupling efficiency.

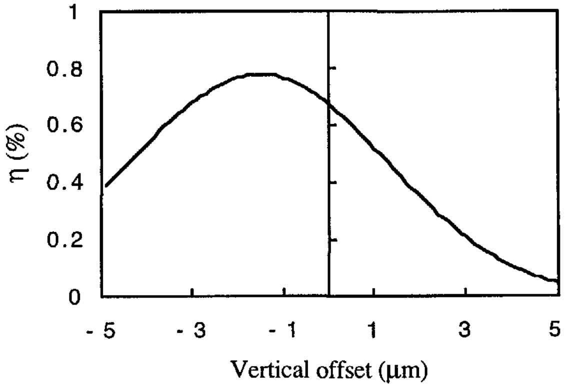


Figure 4.14 Relationship between coupling efficiency and vertical offset for a particular coupling configuration.

Suppose that the fiber has a 1/e amplitude radius of 5 μm. Since the lateral mode size of the $Q_{1.3}$ waveguide can be controlled by changing the waveguide width through lithography, we choose b = 5 μm. For a vertical waveguide taper from 0.6 to 0.05 μm with a cover layer of 5 μm ($n_c = 3.169$) and a length of 300 μm, $c_1 = 1.8$ μm, $c_2 = 1.8$ μm, the coupling efficiency is calculated to be about 64% (1.94 dB loss). Since the propagation loss is 0.41 dB (see Figure 4.9), the total loss will be about 2.35 dB. This value is still high for our tapered waveguide structure and can further be decreased. The reason is that the vertical mode size is not big enough even if the waveguide height has been tapered down to 0.05 μm.

Next, we taper the waveguide to 0.025 μm over a length of 300 μm. After a similar calculation, we get $c_1 = 2.7$ μm, $c_2 = 3.6$ μm. However, our results show that the lateral mode size also expands considerably. Based upon simulation results, we found that the width of the taper in the lateral direction should be changed to 7 μm instead of 12 μm in order to obtain a lateral mode size of 10 μm to match that of the fiber. Then the coupling efficiency will be about 90% (0.45 dB loss) using a 0.4 μm offset. The propagation loss is about 0.5 dB, making the total loss about 1 dB.

We have seen that the waveguide should be tapered down to 0.025 μm to obtain a high coupling efficiency. This might be realized by the atomic beam etching technique, which will be discussed in Section 5.2.

4.5 Summary

Three-dimensionally tapered waveguide structures composed of five sections have been designed and discussed. For technological convenience, the laterally and vertically tapered sections are separated by a straight transition section. The cover layer should have a mode index matching that of the substrate in order to obtain a symmetrical mode field distribution in the vertical direction. In addition, the cover should have a considerable thickness, otherwise the mode size will still be small in the vertical direction. We have discussed a design with a cover thickness of about 5 μm. Also, the length of the laterally tapered Section B (see Figure 4.1) should be longer than 250 μm in order to avoid that noticeable higher order modes are excited. Then, the vertically tapered Section D, if linearly tapered, should be longer than 400 μm to cause a propagation loss of less than 0.4 dB. The straight waveguide Section E should be longer than 200 μm to stabilize the propagating field in order to obtain a higher coupling efficiency. Finally, Section C should be less than 100 μm, and has no noticeable influence on the total power loss. This way, our design has a total length of 1.1 mm. If the thickness of Section E is less than 25 nm, the total access loss including both coupling and propagation loss can be less than 1 dB.

Based upon the above discussions, we propose the following structure (see Table 4.1): The waveguide material is $In_{0.71}Ga_{0.29}As_{0.62}P_{0.38}$ and the substrate is InP. The cover has an index of 3.165±0.005 and a thickness of 5 μm, which can be InP or silicon nitride. Sections A, B and D have lengths of 100, 250 and 400 μm, respectively. The length of Section C is less than 100 μm, its actual length depending on the technological possibilities. The above lengths are allowed to be up to 50 μm larger. Sections A and E have widths of 2 and 7 μm, respectively, with a tolerance of ±0.1 μm. Sections B and D include linearly tapered profiles. Section C has a waveguide thickness of 600 nm with a tolerance of ±25 nm, while Section E has a waveguide thickness of 25 nm which should be precisely controlled within a tolerance of ±5 nm.

Table 4.1. Designed tapered waveguide structure.

Section	Length (μm)	Waveguide thickness (nm)	Waveguide width (μm)	Cover material
A	100	600	2	InP or silicon nitride
B	250	600	2 -> 7	
C	100	600	7	
D	400	600 -> 25	7	
E	250	25	7	

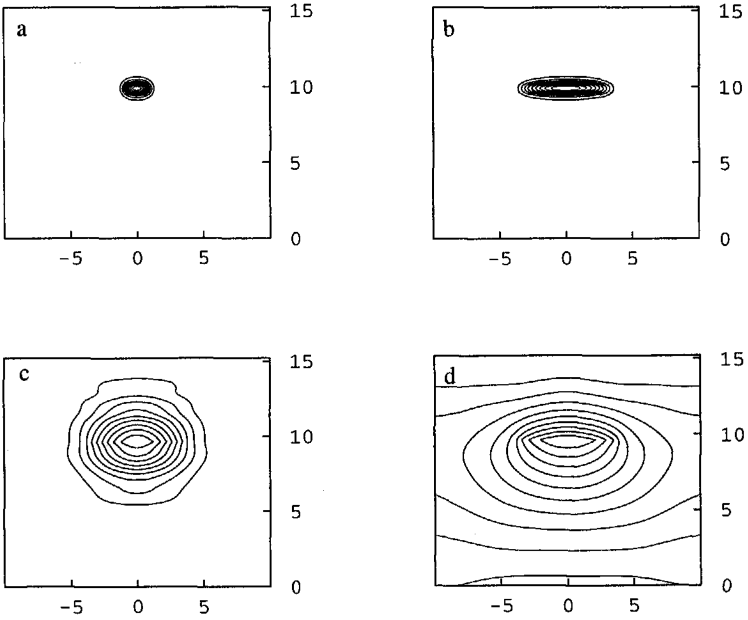


Figure 4.15 Evolution of iso-amplitude contours of E_y for the HE_{00}-mode coupled into the designed 3D tapered waveguide structure corresponding to (a) input, (b) behind the lateral taper, (c) at the end of the vertical taper and (d) output coupling to the fiber.

The propagating field from Sections A to E is shown in Figure 4.15, where a small input field has been transformed to a field matching that of the fiber. The total coupling loss is about 1 dB, while even if the fiber were directly butt-coupled to a waveguide of 12 μm width, in which case the lateral field sizes are properly matched, the loss would be still 7.2 dB. For a tapered fiber, its minimal coupling loss to the waveguide (with optimal lateral waveguide size) is usually larger than 3 dB [4], also requiring a stringent alignment tolerance. Therefore, the tapered waveguide structure leads to a high coupling efficiency and a simple coupling configuration.

4.6 References

[1] J. Haes, J. Willems and R. Baets: "Design of adiabatic tapers for high contrast step index waveguides," in *Linear and Nonlinear Integrated Optics*, Giancarlo C. Righini, David Yevick, Editors, Proc. SPIE **2212**,685-693 (1994).

[2] J.D. Love, W.M. Henry, W.J. Stewart, R.J. Black, S. Lacroix and F. Gonthier: "Tapered single-mode fibers and devices, Part 1: Adiabaticity criteria," *IEE Proc. J.*, **138**, 343-354 (1991).

[3] F. Irrea: "Improved analysis of coupling between 2-D optical waveguides and single mode fibers," *Alta Frequenza*, **57**, 493-495 (1988).

[4] G. Wenger, L. Stoll, B. Weiss, M. Schienle, R. Müller-Nawrath, S. Eichinger, J. Müller, B. Acklin and G. Müller: "Design and fabrication of monolithic optical spot size transformers (MOST's) for highly efficient fiber-chip coupling," *J. Lightwave Technol.*, **12**, 1782-1790 (1994).

Chapter 5

Fabrication of tapered waveguide structures

In our experiments, the fabrication of a three-dimensional tapered waveguide structure on InGaAsP/InP consists of the following steps:

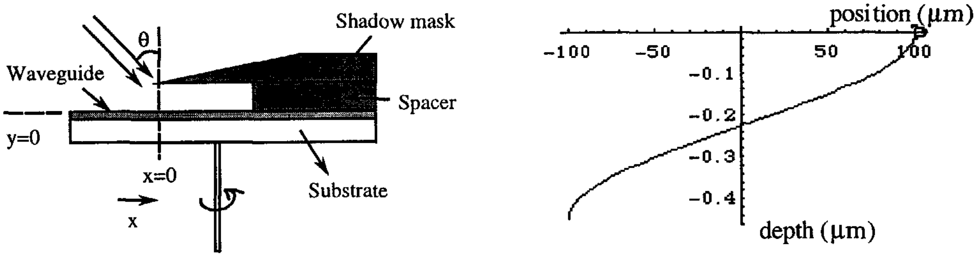
(1) realization of the vertically tapered profile,
(2) definition of the lateral pattern,
(3) cover layer growth over the structure.

Step (1) was realized by atomic beam etching combined with a shadow mask technique as will be described in Section 5.1. Step (2) can be performed through a conventional lithography process as summarized in Section 5.2. Step (3) could be carried out using the MOVPE technique, but two reasons have limited such a process. First, it poses critical requirements to the surface quality, which should be free of defects. This requirement is usually guaranteed by a wet chemical etching step before the MOVPE-process but leads to technological difficulties ensuring the tolerance of the vertically tapered profiles realized in Step (1). Secondly, MOVPE is a very expensive technique; extensive experiments are hard to afford. Therefore, we decided to develop a new cost-saving process to fabricate the high refractive index cover layer of silicon nitride by use of plasma-enhanced chemical vapor deposition (PECVD) instead of the expensive MOVPE technology, to be presented in Section 5.3. This development has indicated promising results and a practically acceptable process is shown to be possible.

5.1 Vertically tapered waveguides

5.1.1 Principle

The principle to realize a vertically tapered profile is schematically shown in Figure 5.1a. The chip is mounted on a holder which can be rotated at a certain speed around a vertical axis. A Si spacer is used to control the distance between a tapered Si shadow mask and the chip. If an atomic beam has a constant angle of incidence θ, a tapered profile will be obtained on the chip after atomic beam etching.



(a) (b)

Figure 5.1 Direct shadow etching.

With a straightforward mathematical derivation, the taper length L and the etching depth d(x) can be expressed as:

$$L = 2T\,\mathrm{tg}(\theta), \tag{5.1}$$

$$d(x) = \frac{d_{\max}}{\pi}\arccos\left(\frac{2x}{L}\right), \tag{5.2}$$

where T is the thickness of the spacer and θ denotes the incident angle of the atomic beam while $d_{\max}$ is the maximal etching depth. Thus, the taper length depends on the thickness of the spacer, while the taper profile depends on the maximal etching depth i.e. the etching time. For example, if the spacer has a thickness of 100 μm, the maximal etching depth is 0.5 μm and the angle of incidence of the atomic beam is 45°, the taper length will be twice the spacer thickness and a taper profile will be obtained as shown in Figure 5.1b.

5.1.2 Fabrication of Si taper masks

We have used two methods to fabricate Si taper masks.

Wet chemical etching: The procedure for this is symbolized in Figure 5.2. The (100) Si substrate is firstly covered with a Si_3N_4 layer by PECVD, then covered with another photoresist layer. The photoresist layer is opened following a certain pattern through exposure and development. Subsequently, the uncovered Si_3N_4 area is removed by reactive ion etching or wet chemical etching (40% HF). After removing the residual photoresist layer, the back side of the substrate is covered with another thick layer of Si_3N_4 (2 μm) which prevents etching from the backside. Then, the substrate is etched in a 30 wt% KOH solution at 85°C. The etching speed in the <100> direction is about 2.5 μm/min. Finally, the Si_3N_4 layer is removed from the taper masks by wet chemical etching (40% HF).

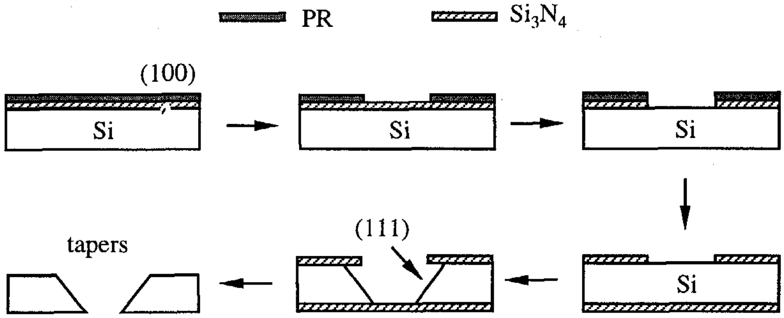


Figure 5.2 Si taper masks made by the wet chemical etching.

Si taper masks made by the above mentioned procedure have a constant taper angle of 54.7°, due to the slow etching speed in the <111> direction. Usually, the angle of incidence of the atomic beam is about 45°. The angle of 54.7° on the Si tapered masks may cause a slight change in the etched profile (see Section 5.1.3).

Mechanical polishing: We also have used a mechanical polishing to produce tapered Si-masks. The Si substrate is firstly mounted between two glass plates with tilted end-faces. After polishing, a Si taper is formed. The taper angle can be changed freely by adjusting the relative position of the glass plates. The masks produced had usually an angle of less than 40°. Taper masks made with this method have shown good quality. This method is preferred above the previous one because of its simplicity and versatility.

5.1.3 Atomic beam etching

The InGaAsP waveguide layer ($Q_{1.3}$) with a thickness of 0.6 μm is firstly grown on the InP substrate, then covered with a 0.3 μm thick InP protection layer, by MOVPE*. We used an argon (Ar) atomic beam to etch the sample. Before this etching, the InP protection layer has been removed by a selective wet chemical etching (1HCl : $4H_3PO_4$). Figure 5.3 shows a vertically tapered profile measured with a surface stylus profiler (α-step), where a spacer with a thickness of 100 μm has been used. It must be reminded that different scales occur along the axes in this figure. The maximal etching depth is 0.3 μm, while the taper length is about 500 μm.

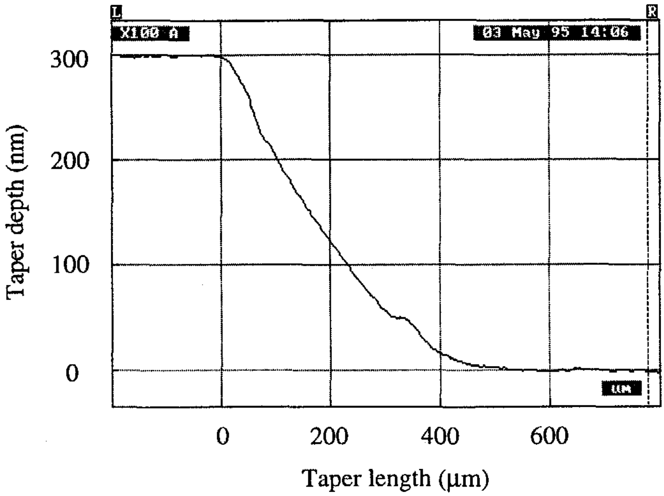


Figure 5.3 Vertically tapered profile measured with α-step profiler. The spacer has a thickness of 100 μm and the etching time was 60 min.

---

* Kindly carried out by IMEC, University of Gent, Belgium.

From Figure 5.3, it is seen that an almost linearly tapered profile has been obtained, which is different from the theoretically predicted profile as indicated in Figure 5.1b, where the bottom left end of the taper has an unwanted fast increase of the depth. In addition, the taper length in Figure 5.3 is much longer than 200 μm as was predicted according to Eq.(5.1). This occurs because, practically, the sample holder in Figure 5.1a has also a lateral periodic movement, causing the incident angle of the atomic beam to change periodically in the range of 45 ± 15°. Also, our Si taper mask has a taper angle slightly bigger than $90\text{-}\theta_{max}$ where θ_{max} is the maximal incident angle of the (not necessarily parallel) atomic beam. As a result, part of the atomic beam can not reach the sharp tip of the Si taper mask, due to blocking by its upper edge as shown in Figure 5.4.

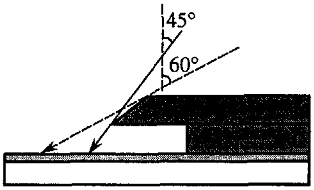


Figure 5.4 Influence of the small Si taper mask angle.

Consequently, the length of the tapered profile on the chip will be increased. This phenomenon will lead to a local depth change, which can be observed from Figure 5.3 in the neighborhood of a taper length of 300 μm. If the angle of the Si taper mask is smaller than $90\text{-}\theta_{max}$, this problem vanishes.

Figure 5.5 shows another vertically tapered profile made by use of a spacer thickness of 200 μm and the same atomic beam etching time. The length of the vertically tapered profile increases to about 800 μm, while its maximal depth stays unchanged. Thus, the required length can be obtained using a spacer with a controlled thickness. The etching speed was found to be about 5 nm/min. The maximal depth can be changed to meet the requirements by controlling the etching time.

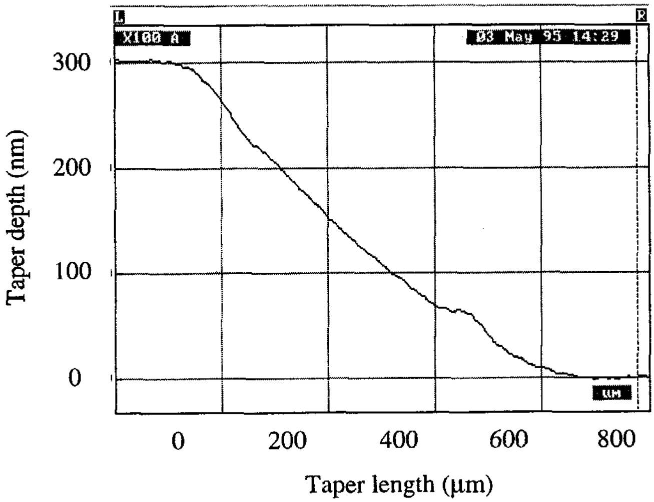


Figure 5.5 Vertically tapered profile measured with α-step profiler. The spacer has a thickness of 200 µm and the etching time is 60 min.

In Section 4.5, we designed a structure with a vertically tapered profile from 0.6 to 0.025 µm over a length of 400 µm. This can be realized using a spacer of about 75 µm thickness. If a bigger length does not influence the total device size, it has a positive effect on decreasing the propagation loss and thus is gladly accepted. This leads to a considerable technological convenience since the section length is allowed to be longer than the designed value. However, the thickness tolerance is only about 5 nm. We found that the etching speed is variable within the range of 5 ± 0.05 nm/min due to vacuum and beam current fluctuation of the available atomic beam etching equipment. This requires that the ambiance and equipment should be stabilized before etching starts. In addition, the surface roughness must be very small. We have inspected the surface also with a stylus profiler. The results are shown in Figures 5.6 and 5.7, indicating that the roughness increases slightly after atomic beam etching, over 2 and 3 to 4 nanometers, respectively. These results have been verified by alternative measurement using phase-shifting interference microscopy [1].

The experimental results show that a linearly vertically tapered profile can be realized by atomic beam etching combined with a silicon shadow mask. Its length can be controlled by the spacer thickness and its depth can be changed via the etching time. We also have

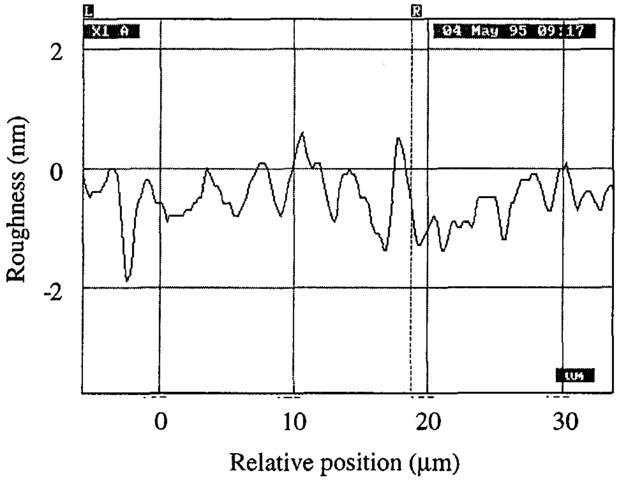


Figure 5.6 Surface roughness before the atomic beam etching measured with α-step profiler.

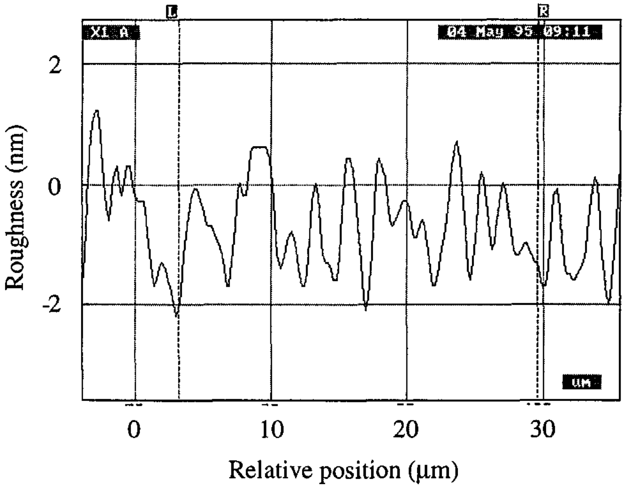


Figure 5.7 Surface roughness after the atomic beam etching for 60 min measured with α-step profiler.

inspected photoluminescence responses of surfaces before and after atomic beam etching. This is an extremely sensitive technique to detect surface lattice defects but can give qualitative answer only. As a result, there exist evident response differences, showing that the atomic beam has an influence on the lattice structure of surfaces. Since the defects can affect the further cover growth by MOVPE, they should be removed before such a growth, using wet chemical etching. However, we also grew a silicon nitride cover layer (see Section 5.3) by PECVD. In that case, the cover growth appears not to be affected by possible surface lattice defects.

5.2 Lateral definition

The lateral definition of three-dimensional tapered waveguide structures can be realized by a combination of photolithography and reactive ion etching. The basic photolithography process is summarized in the following:

Substrate cleaning: The surface of a substrate is in general polluted and oxidized. Therefore, cleaning is necessary. We use two cleaning steps to remove dust and other forms of contamination. One is ultrasonic cleaning in acetone, the other is O_2 plasma etching for 15 minutes (400 W) followed by wet chemical etching with 10% H_3PO_4 for 1 minute to remove oxides, e.g. In_2O_3.

Spin coating: Spin coating is the most widely used technique for applying photoresist to wafers. The resist is dispensed onto the surfaces of the wafers, which are then accelerated to provide a thin uniform film across wafer surfaces. We use Hoechst, type AZ-5214 positive photoresist.

Softbaking: Softbaking is used to remove the solvents present in the spin-coated resist film and is usually performed immediately after the coating process. We softbake our substrates at 100°C during 3 minutes.

Exposure: The goal of the exposure is to transmit a desired pattern on a mask as a latent image in the resist film. We use a Canon FPA-143 fine pattern projection aligner for exposure.

Development: In the development step the latent image is chemically transformed into a fixed one by removing solvable parts of the photoresist. Corresponding to AZ-5214, we use a PLSI solution (1PLSI : 1$H_2$0) as the developer. The developing time is 1 minute at room temperature. After developing, the wafers are rinsed and dried.

Postbaking: Postbaking is accomplished in the same manner as softbaking to remove any residual moisture from the developing operation and to improve the sticking of the resist to the wafer. We postbake our substrates at 120°C for 5 minutes.

Starting with a wafer on which a tapered profile has been produced by atomic beam etching, through the above lithography processes, the required horizontal pattern has been protected with a photoresist layer (Figure 5.8). The precise alignment of the pattern relative to the edge of the vertically tapered profile before the projection exposure is hard to realize, since the edge is not sharp. This is the reason why we have inserted Section C in our design (see Figure 4.1), the length of which depends on the technological possibilities. Subsequently, the uncovered area is etched to remove the InP top layer. Both wet chemical etching and reactive ion etching (RIE) can be used. However, wet chemical etching usually leads to a considerable undercut, which spoils the precise definition of a horizontal pattern. Consequently RIE is preferred, by which vertical walls have to be obtained.

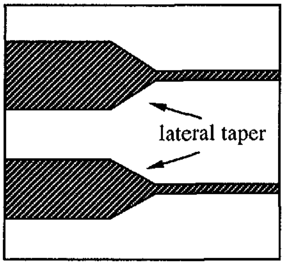


Figure 5.8 Top view of 3D tapers

5.3 Cover layer growth

A thick cover layer with a refractive index matching that of the substrate is necessary to obtain a vertically symmetrical field distribution. A thick InP cover layer can be deposited on an InGaAsP/InP wafer by MOVPE growth, provided that surface defects have been removed. Since the waveguide thickness of Section E in Figure 4.1 has a tolerance of about 5 nm only, the removal of defects will further increase the technological difficulties. In addition, MOVPE is an expensive technology. Since our 3D tapered waveguide structures were still in an experimental stage, numerous tests were required. Based upon these considerations, we developed a new process, i.e. using PECVD, in which an amorphous silicon nitride layer replaces the InP cover. The refractive index of silicon nitride depends on its composition and can be controlled within the range of 2.0 to 3.5.

5.3.1 Refractive index

The refractive index of the cover layer is a critical parameter for a tapered waveguide structure (see Section 4.3.2). We use a photospectrometer to determine the index values of fabricated silicon nitride layers. A silicon nitride layer is first deposited on an InP substrate by PECVD. Then, its reflection curve as a function of the wavelength is measured*. Typical reflection curves are shown in Figure 5.9.

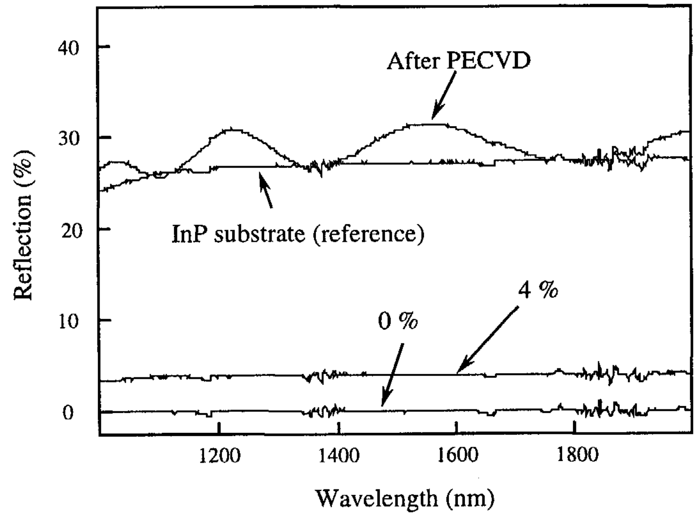


Figure 5.9 Reflection curve for a silicon nitride layer deposited on an InP substrate showing dispersion, together with reference curves.

---

[*] Kindly performed by Mr. A. Kunze.

A series of peaks are due to interference of contributions reflected at both film interfaces [2]. Both the index value and the thickness of the silicon nitride layer can be calculated from this dispersion curve. Usually, if the wavelength increases, the reflection decreases. In Figure 5.9, there exists a spurious dispersion since the reflected beam is not completely collected by the detector. However, it can be corrected later by use of a reference substrate, e.g. InP, which has a known refractive index and dispersion.

Our experiments show that the index of a silicon nitride layer can be changed by varying the gas flow ratio (N_2 : SiH_4). If the plasma power is 400 W, the deposition temperature 250 °C and the gas flow ratio is 250 : 195, the deposited layer has a refractive index of 3.12. If the gas flow ratio is 200 : 195, the index will be 3.16. The layer has a high growth rate of about 280 nm/min. If the plasma power is 100 W, both the index value and the deposition growth rate change. The index value becomes about 3.21 and the deposition rate 150 nm/min for a gas flow ratio of 250 : 195. The fabricated amorphous layer material can be indicated as Si_xN_{1-x}:H, where x exceeds 0.9.

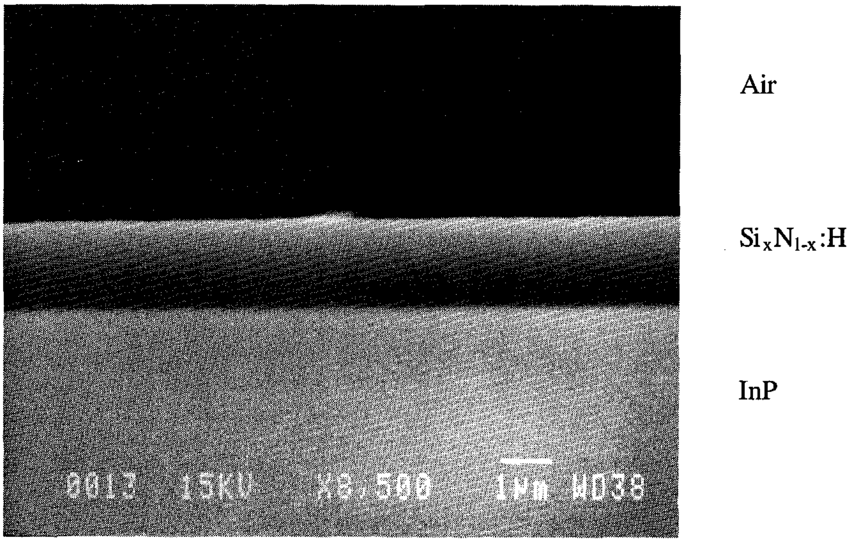


Figure 5.10 Morphology of the cover from SEM.

By optimizing the gas flow ratio, we have obtained an index value that has a deviation of less than 0.005 from that of the substrate, which is already acceptable for a practical waveguide taper.

Figure 5.10 shows a 1.8 μm thick silicon nitride layer on an InP substrate, obtained by scanning electron microscopy. With a plasma power of 400 W, depositing such a layer by PECVD takes less than 6 minutes.

5.3.2 Attenuation measurement

Ideally, the silicon nitride layer should not show absorption. Unfortunately, this is experimentally not realized. Therefore, an extra power loss in tapered structures is introduced by this cover layer. Suppose that the cover material has a bulk attenuation of 20 dB/cm. For a slab waveguide with a $Q_{1.3}$ layer of a certain thickness on an InP substrate, the extra power loss due to the cover can be calculated [3]. The results are shown in Figure 5.11, where the cover layer has a thickness of 900 nm and its index value is variable. Practically, we first deposit a silicon nitride layer on the wafer. Subsequently, the thickness and index of the cover are determined according to its dispersion curve. Then, the extra power loss due to the cover is measured. The bulk attenuation of the cover is finally calculated from the measured results.

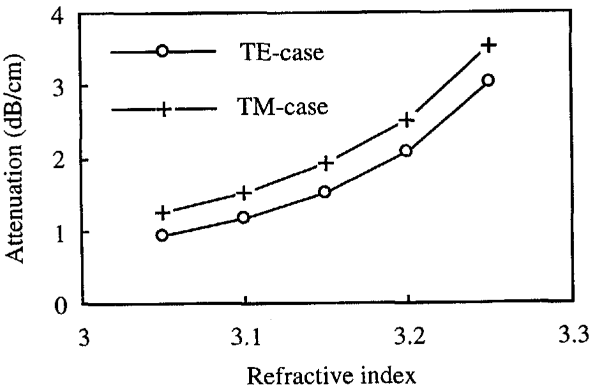


Figure 5.11 Extra power attenuation due to a cover layer of 900 nm thickness for a bulk attenuation in the cover of 20 dB/cm.

Starting with the wafer as described in Section 5.1.3, we first remove the InP top layer by use of a selective wet chemical etchant ($1HCl : 4H_3PO_4$). The etchant has an etching speed of about 400 nm/min for InP, while no etching occurs in InGaAsP ($Q_{1.3}$). Next, the wafer is masked by an aluminum plate which has a 9 × 6 mm open window. The silicon nitride layer is only deposited through this window. This arrangement is of convenience in the later attenuation measurements, allowing to compare attenuation values with that of an uncovered area on the same wafer. An input beam with a wavelength of 1.55 μm is coupled into the single transverse mode slab waveguide of the wafer by focusing on one endface. Since the beam expands in the lateral direction as the propagation proceeds, an extra cylinder lens is added in front of the focusing lens such that a laterally parallel beam is realized. The output beam is recorded on a television camera (Bosch) and its power is measured by a photo detector.

Table 5.1 Transmitted power (μW) for waveguides with a silicon nitride layer.

Position	TE case	TM case
1	4.6	6.8
2	4.6	7.0
3	4.4	5.6
4	4.6	5.6
5	4.6	6.5
6	4.3	5.2

We have deposited a silicon nitride layer with an index of 3.10 at 1.55 μm and a thickness of 0.9 μm on a wafer with a $Q_{1.3}$ layer of 0.6 μm thickness, utilizing a plasma power of 400 W*. Subsequently, the transmitted powers (μW) are measured. Table 5.1 shows the transmitted powers through the areas with the silicon nitride cover layer. Table 5.2 shows the transmitted powers through the areas that during the deposition of the cover have been masked by the aluminum plate. There are some variations in our transmission results for each area. This is because the cleaved endfaces of our chips suffered from minor damage during processing. The highest transmitted power values have been used in the subsequent calculations.

---

* Kindly performed by Ing. T. Scholtes and Dr.ir. L. Shi.

The extra power attenuation due to a silicon nitride layer over a given length can be calculated from $-10 \times \log_{10}\frac{P_c}{P_u}$, where P_c and P_u are the transmitted powers through the covered and uncovered areas, respectively. The transmitted powers for the TE and TM cases are about 4.6 and 7.0 μW through the deposited area (see Table 5.1), while 17.8 and 30.9 (see Table 5.2) through the uncovered area. Taking into account the window length of 9 mm, the calculated extra power attenuation is 6.53 and 7.17 dB/cm for the TE and TM cases, respectively. When the cover has a thickness of 900 nm and a refractive index of 3.10, a bulk attenuation of 20 dB/cm corresponds to an extra power attenuation of 1.17 and 1.54 dB/cm for the TE and TM cases, respectively (see Figure 5.11). From this, the equivalent real bulk attenuation value of the silicon nitride is calculated to be about 105 ± 10 dB/cm.

Table 5.2 Transmitted power (μW) for waveguides without a silicon nitride layer.

Position	TE case	TM case
7	17.0	30.9
8	13.3	21.5
9	13.5	23.4
10	17.8	29.9
11	16.4	23.2
12	14.6	23.2

The following two mechanisms might contribute to the high attenuation in the cover layer. First, the waveguide surface can be damaged by the plasma during the PECVD process. Secondly, the cover material could have a bandgap close to the wavelength of 1.55 μm we used, thus leading to a big absorption.

To investigate the first possibility, it is remarked that then the waveguide transmission after removing the PECVD-cover layer must be smaller than that in the area where no cover has been deposited. We removed the silicon nitride layer by wet chemical etching (30% KOH) at 80°C. The transmission results are shown in Table 5.3, where the transmitted powers in the covered area (positions 1-6) are similar to those in the protected area (positions 7-12). This means that the possible waveguide surface damage by the PECVD can be neglected.

Table 5.3 Transmitted power (μW) for waveguides without the silicon nitride layer.

	Position	TE case	TM case
cover removed	1	13.4	18.2
	2	13.4	17.7
	3	16.8	13.7
	4	15.1	12.2
	5	15.5	20.1
	6	17.9	16.1
never covered	7	7.5	8.4
	8	11.0	18.1
	9	11.7	15.6
	10	14.7	22.4
	11	12.7	14.8
	12	13.2	13.2

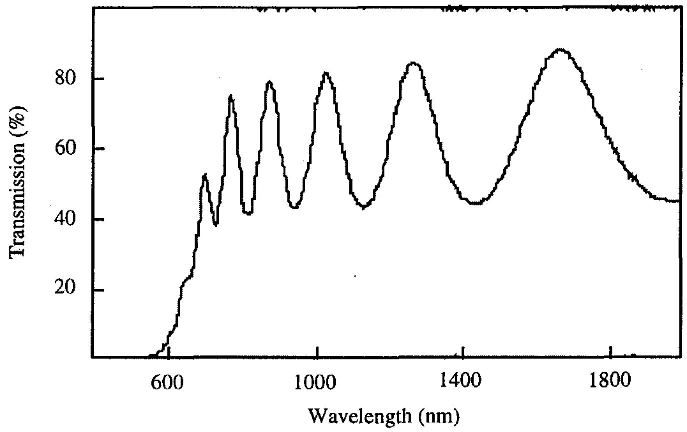


Figure 5.12 Transmission curve for the bandgap measurement

To investigate the second reason, we have deposited a silicon nitride layer on a Balzers glass substrate and measured the transmission properties. Figure 5.12 shows the measured results. The peaks are due to interference introduced by the film layer. The transmission decreases rapidly if the wavelength is below 700 nm, corresponding to the bandgap of the cover layer. This is far below the wavelength of 1.55 μm we used. Therefore, the high bulk attenuation of the obtained cover is not caused by the bandgap absorption. Thus, neither of the two mechanisms can be responsible for the measured high bulk attenuation and we conclude that the attenuation can be further decreased by optimization of the deposition process.

There are several parameters such as plasma power, gas concentration, operation temperature and pressure, which influence the material properties of the deposited layer, thus influencing the attenuation. Therefore, it can be expected that the attenuation of the cover layer can further be decreased by optimizing those process parameters.

The plasma power is believed to be an important parameter in this respect and has been investigated extensively. We have deposited silicon nitride layers by using a plasma power between 400 and 50 W. The high powers produce a relatively rough surface of the silicon nitride. The deposition speed decreases as the plasma power becomes smaller. The power of 50 W was found to be too low such that the deposition is not realizable. We have chosen 100 W for the next fabrication runs because this leads to layers with smooth surfaces.

As a next experiment, we removed the InP top layer from a central area on the chip by use of a combination of lithography and wet chemical etching. This area was chosen to be smaller than the window in the aluminum plate. Therefore, after PECVD deposition of silicon nitride through this plate, three kinds of layer structures were obtained on the same chip: InP/$Q_{1.3}$/InP, SiN/InP/$Q_{1.3}$/InP and SiN/$Q_{1.3}$/InP. The deposited silicon nitride layer had a thickness of 0.6 μm and an index of 3.21 at 1.55 μm, measured on the photospectrometer. Tables 5.4-6 show the measured transmission results for the three kinds of layer structures, respectively.

Comparing Tables 5.4 and 5.6, we detect an extra power loss of about 4.7 and 5.8 dB/cm. Table 5.6 corresponds to the area with the top InP layer but protected during the deposition. Since the removal of the InP top layer leads to a certain transmission decrease, the above power loss will be less if we replace the values in Table 5.6 with those for the area without the InP top layer. With a calculation similar to the one for Figure 5.11 where the silicon nitride layer had a bulk attenuation of 20 dB/cm, the extra power loss introduced by the cover without the InP top layer is found as 2.06 and 2.43 dB/cm for the

Table 5.4 Transmitted power (μW) through the SiN/$Q_{1.3}$/InP area.

Position	TE case	TM case
1	14.9	15.1
2	14.2	14.6
3	14.6	15.4
4	14.7	14.2
5	14.5	13.9
6	14.1	14.6
7	14.8	14.5
8	13.9	14.2

Table 5.5 Transmitted power (μW) through the SiN/InP/$Q_{1.3}$/InP area.

Position	TE case	TM case
9	28.3	40.3
10	29.1	40.0
11	29.6	39.8
12	28.2	39.9

Table 5.6 Transmitted power (μW) through the InP/$Q_{1.3}$/InP area.

Position	TE case	TM case
13	39.1	50.2
14	35.0	45.8
15	38.9	49.3
16	35.5	46.4
17	36.6	43.2
18	37.1	47.5
19	35.1	49.2
20	36.2	48.0

TE and TM cases, respectively. From this, the actual bulk attenuation of the silicon nitride layer is calculated to be slightly less than 45 dB/cm. Similarly, the extra 1 dB/cm transmission attenuation calculated from Table 5.5 as compared with the values from Table 5.6 is also due to the introduced power loss by the silicon nitride layer.

The above results show that the bulk attenuation of the cover has effectively been decreased through the PECVD optimization. A further decrease to less than 20 dB/cm is believed to be possible by further process optimization, e.g. by annealing.

Now, we predict the extra power loss introduced by the cover with a bulk attenuation of 45 dB/cm to our designed 3D tapered waveguide structure as described in Table 4.1. Based upon a calculation similar to that in Figure 5.11, the power attenuation in Sections A through C (450 μm in total) is about 0.2 dB, while 0.4 dB in Section D (400 μm) and 0.5 dB in Section E (250 μm). The overall loss introduced by the cover is about 1.1 dB. By adding the calculated radiation loss (0.5 dB) in the designed structure and the fiber-chip field overlap loss (0.45) dB, the total loss is about 2 dB.

In literature, tapered waveguide structures for fiber-chip coupling have been demonstrated [4-6]. All of them are based upon the expensive MOVPE technique in order to obtain an InP cover. Müller *et al* [4] presented a tapered waveguide structure with a transmission loss of 4.9 dB using ion milling combined with a tapered Al_2O_3 mask layer deposited on a $Q_{1.05}$ layer. Wegner *et al* [5] reported realization of 2.6 dB transmission loss (fiber-chip-lens-detector) by use of a combination of a rib n^--InP waveguide on an n^+-InP substrate and a tapered $Q_{1.3}$ waveguide. Brenner *et al* [6] demonstrated a tapered waveguide structure by use of a diffusion limited bromine etch [7] and low pressure MOVPE InP cover growth. They reported a very low total loss (including overlap loss and loss in the taper) in the range of 0.8 - 1.6 dB. However, they used a $Q_{1.05}$ waveguide which is not as strongly guiding as the $Q_{1.3}$ waveguide we used. This gives them relatively less difficulties to expand a propagating field adiabatically.

PECVD is a cost-saving technique, but the deposited cover layer still has a large bulk attenuation (45 dB/cm). Efforts have to be devoted to decrease this bulk attenuation in the future by optimizing PECVD parameters, e.g. deposition temperature, gas concentration and annealing. Furthermore, the tapered profile must further be optimized by use of our fully vectorial beam propagation method to decrease the radiation loss and the fiber-chip overlap loss. Other types of tapered waveguide structures such as laterally and vertically tapered profiles designed and realized in a single section could be considered. Extensive efforts are still needed to obtain practical 3D tapered waveguide structures with a high coupling efficiency to a single mode fiber (with a total loss smaller than 1 dB).

5.4 Summary

Fabrication of three-dimensionally tapered waveguide structures has been investigated. Vertically tapered profiles were obtained by atomic beam etching combined with a shadow mask. The length of the vertically tapered profile can be controlled by the spacer thickness, while its maximum depth changes with the etching time at a speed of about 5 nm/min. The lateral pattern can be realized by a photolithography process. A silicon nitride layer has been fabricated by PECVD, substituting an InP cover by MOVPE. Its index can be accurately controlled within a deviation of only 0.005 from that of InP. The bulk attenuation of the cover has been decreased from more than 110 dB to about 45 dB/cm.

5.5 References

[1] J. van Wingerden, H.J. Frankena and C. Smorenburg: "Linear approximation for measurement errors in phase shifting interferometry," *Appl. Opt.*, **30**, 2718-2729 (1991).

[2] L.C. Martin and W.T. Welford: *Technical Optics*, vol. 1, Sir Isaac Pitman & Sons Ltd, 231-234 (1966).

[3] J. de Jong: "A scheme for fast calculation of guided modes in planar waveguides," in *Education in Optics*, G.B. Altshuler, B.J. Thompson, Editors, Proc. SPIE **1603**, 515-522 (1991).

[4] G. Müller, B. Stegmüller, H. Westermeier and G. Wenger: "Tapered InP/InGaAsP waveguide structure for efficient fiber-chip coupling," *Electron. Lett.*, **27**, 1836-1838 (1991).

[5] G. Wenger, L. Stoll, B. Weiss, M. Schienle, R. Müller-Nawrath, S. Eichinger, J. Müller, B. Acklin and G. Müller: "Design and fabrication of monolithic optical spot size transformers (MOST's) for highly efficient fiber-chip coupling," *J. Lightwave*

Technol., **12**, 1782-1790 (1994).

[6] T. Brenner, M. Bachmann and H. Melchior: "Vertically tapered InGaAsP/InP waveguides for highly efficient coupling to flat-end single-mode fibers," *Appl. Phys. Lett.*, **65**, 798-800 (1994).

[7] T. Brenner and H. Melchior: "Local etch-rate control of masked InP/InGaAsP by diffusion-limited etching," *J. Electrochem. Soc.*, **141**, 1954-1958 (1994).

Appendix A

The scalar FFT- BPM

Since light is an electromagnetic wave phenomenon, its propagation properties follow from Maxwell's equations. Here, we restrict our discussions to an inhomogeneous dielectric medium which is linear, lossless, charge-free and isotropic. The Helmholtz equation for the vectorial electrical field **E**(x,y,z) in a Cartesian coordinate system has the following form:

$$\nabla^2 \mathbf{E}(x,y,z) + k_0^2 n^2(x,y,z)\mathbf{E}(x,y,z) = \nabla(\nabla \cdot \mathbf{E}), \tag{A.1}$$

where k_0 is wave number in vacuum and n(x,y,z) is the local refractive index. For a scalar treatment, the term $\nabla(\nabla \cdot \mathbf{E})$ is discarded and the vectorial field **E**(x,y,z) is replaced by its magnitude, i.e. E(x,y,z). Thus

$$\nabla^2 E(x,y,z) + k_0^2 n^2(x,y,z)E(x,y,z) = 0. \tag{A.2}$$

Assuming

$$E(x,y,z) = \varphi(x,y,z)\exp(i\beta z), \tag{A.3}$$

where $\varphi(x,y,z)$ is the complex wave amplitude and β is a properly chosen reference propagation constant, we get:

$$\frac{\partial^2\varphi(x,y,z)}{\partial x^2}+\frac{\partial^2\varphi(x,y,z)}{\partial y^2}+\frac{\partial^2\varphi(x,y,z)}{\partial z^2}+i2\beta\frac{\partial\varphi(x,y,z)}{\partial z}+ \qquad \text{(A.4)}$$
$$\left[k_0^2n^2(x,y,z)-\beta^2\right]\varphi(x,y,z)=0.$$

With the slowly varying envelope approximation, implying that

$$\left|\frac{\partial^2\varphi(x,y,z)}{\partial z^2}\right|<<\left|2\beta\frac{\partial\varphi(x,y,z)}{\partial z}\right|, \qquad \text{(A.5)}$$

we then obtain

$$\frac{\partial\varphi(x,y,z)}{\partial z}=\frac{i}{2\beta}\left\{\frac{\partial^2\varphi(x,y,z)}{\partial x^2}+\frac{\partial^2\varphi(x,y,z)}{\partial y^2}+ \right. \qquad \text{(A.6)}$$
$$\left.\left[k_0^2n^2(x,y,z)-\beta^2\right]\varphi(x,y,z)\right\}.$$

For a two-dimensional structure ($\frac{\partial\varphi}{\partial y}=0$), this simplifies further into:

$$\frac{\partial\varphi(x,z)}{\partial z}=\frac{i}{2\beta}\left\{\frac{\partial^2\varphi(x,z)}{\partial x^2}+\left[k_0^2n^2(x,y,z)-\beta^2\right]\varphi(x,z)\right\}. \qquad \text{(A.7)}$$

In the original BPM, an input field $\varphi(x,z_0)$ is propagated over a small distance Δz to obtain the field at $z_0+\Delta z$ by writing $\varphi(x,z_0+\Delta z) = \psi(x,z_0+\Delta z)\exp(i\Gamma)$, where Γ is a phase correction given by

$$\Gamma=\frac{k_0^2n^2-\beta^2}{2\beta}\Delta z. \qquad \text{(A.8)}$$

Thus, $\psi(x,z_0) = \varphi(x,z_0)$ and ψ satisfies

$$\frac{\partial\psi(x,z)}{\partial z}=\frac{i}{2\beta}\frac{\partial^2\psi(x,z)}{\partial x^2}. \qquad \text{(A.9)}$$

If $\Psi_m(z)$, with $m = 0, 1, \cdots, N-1$, denotes the discrete Fourier transform of $\psi(x,z)$ with respect to x and if $x_j = x_0 + j\Delta x$ for $j = 0, 1, \cdots, N-1$, we arrive at

$$\Psi_m(z_0) = \frac{1}{N}\sum_{j=0}^{N-1}\left[\psi(x_j, z_0)\exp\left(-\frac{i2\pi mj}{N}\right)\right]. \tag{A.10}$$

Combining (A.9) and (A.10), we further obtain*

$$\Psi_m(z_0+\Delta z) = \Psi_m(z_0)\exp\left(-\frac{i2\pi^2 m^2}{\beta N^2 \Delta x^2}\Delta z\right), \tag{A.11}$$

where the reference propagation constant β has been used to describe the local field propagation. Then, an inverse FFT is performed on $\Psi_m(z_0+\Delta z)$ to obtain $\psi(x_j,z_0+\Delta z)$

$$\psi(x_j, z_0+\Delta z) = \sum_{m=0}^{N-1}\left[\Psi_m(z_0+\Delta z)\exp\left(\frac{i2\pi mj}{N}\right)\right]. \tag{A.12}$$

Subsequently, $\varphi(x,z_0+\Delta z)$ can be calculated with $\varphi(x,z_0+\Delta z) = \psi(x,z_0+\Delta z)\exp(i\Gamma)$ by use of the phase correction.

The basic propagation algorithm is summarized as follows: to start this procedure, the field $\varphi(x,z_0)$ must be provided, which can be the field of a guided mode or any other incident field, then (1) performing an FFT on $\psi(x_j,z_0) = \varphi(x_j,z_0)$ including all grid nodes in the cross section $z = z_0$ to obtain $\Psi_m(z_0)$ from Eq.(A.10), (2) computing $\Psi_m(z_0+\Delta z)$ from Eq.(A.11), (3) performing an inverse FFT on $\Psi_m(z_0+\Delta z)$ to obtain $\psi(x_j,z_0+\Delta z)$ for the next cross section from Eq.(A.12), and (4) calculating $\varphi(x,z_0+\Delta z) = \psi(x,z_0+\Delta z)\exp(i\Gamma)$ by use of the phase correction Γ. This algorithm is applied repeatedly to obtain the field at any finite propagation distance.

---

[*] See R. Scarmozzino and R.M. Osgood, *J. Opt. Soc. Am. A*, **8**, 724-731 (1991)

Appendix B

The Crank-Nicholson scheme

We consider a partial differential equation of the form:

$$\frac{\partial u(x,y,z)}{\partial z} = v_1 \frac{\partial^2 u(x,y,z)}{\partial x^2} + v_2 \frac{\partial^2 u(x,y,z)}{\partial y^2} \tag{B.1}$$

where v_1 and v_2 are known constants. We consider its solution by the finite difference method, representing u(x,y,z) by its values at the discrete set of points $(x_h, y_j, z_l,)$, with

$$\begin{aligned} x_h &= x_0 + h\Delta x, \\ y_j &= y_0 + j\Delta y, \\ z_l &= z_0 + l\Delta z, \end{aligned} \tag{B.2}$$

where *h, j, l* = 0, 1, ···. If (B.1) is rewritten as

$$\frac{u_{h,j}^{l+1} - u_{h,j}^{l}}{\Delta z} = v_1 \left[\frac{u_{h+1,j}^{l+1} - 2u_{h,j}^{l+1} + u_{h-1,j}^{l+1}}{\Delta x^2} \right] + v_2 \left[\frac{u_{h,j+1}^{l+1} - 2u_{h,j}^{l+1} + u_{h,j-1}^{l+1}}{\Delta y^2} \right], \tag{B.3}$$

it is called an *implicit* scheme, which is unconditionally stable. However, there is always a numerical dissipation, leading to a spurious decrease of the field amplitudes if this scheme is used for the propagation of light.

On the other hand, Eq. (B.1) can be also written as

$$\frac{u_{h,j}^{l+1}-u_{h,j}^{l}}{\Delta z}=v_1\left[\frac{u_{h+1,j}^{l}-2u_{h,j}^{l}+u_{h-1,j}^{l}}{\Delta x^2}\right]+v_2\left[\frac{u_{h,j+1}^{l}-2u_{h,j}^{l}+u_{h,j-1}^{l}}{\Delta y^2}\right], \tag{B.4}$$

This is called an *explicit* scheme, which is unstable, leading to exponentially increasing answers.

However, if we form the average of the implicit and explicit schemes, i.e.

$$\begin{aligned}\frac{u_{h,j}^{l+1}-u_{h,j}^{l}}{\Delta z}=\frac{1}{2}\Bigg\{&v_1\left[\frac{u_{h+1,j}^{l+1}-2u_{h,j}^{l+1}+u_{h-1,j}^{l+1}}{\Delta x^2}\right]+v_2\left[\frac{u_{h,j+1}^{l+1}-2u_{h,j}^{l+1}+u_{h,j-1}^{l+1}}{\Delta y^2}\right]\\&+v_1\left[\frac{u_{h+1,j}^{l}-2u_{h,j}^{l}+u_{h-1,j}^{l}}{\Delta x^2}\right]+v_2\left[\frac{u_{h,j+1}^{l}-2u_{h,j}^{l}+u_{h,j-1}^{l}}{\Delta y^2}\right]\Bigg\},\end{aligned} \tag{B.5}$$

it is called a *Crank-Nicholson* scheme, which is stable and widely used. Numerical dissipation can still occur, depending on the considered differential equation and the choice of parameters.

Appendix C

The Coefficients for Eqs. (2.18)

In Section 2.2.1, a set of complicated equations (2.18) has been derived by a combination of (2.11)-(2.17) for the calculation of propagating fields in dielectric waveguide structures. Those equations can be written in matrix form. Their coefficients are functions of the refractive index, step sizes, the scheme parameter, the reference propagation constant, and the known field components at the previous step. The expressions relating to those coefficients are collected in the following:

$$A_1(h,j)=\frac{\alpha\Delta z}{2\beta\Delta x^2}\left[1+\frac{2(n(h+1,j,l+1)-n(h,j,l+1))}{n(h+1,j,l+1)+n(h,j,l+1)}\right],$$

$$A_2(h,j)=\frac{\alpha\Delta z}{2\beta\Delta x^2}\left[1+\frac{2(n(h-1,j,l+1)-n(h,j,l+1))}{n(h-1,j,l+1)+n(h,j,l+1)}\right],$$

$$A_3(h,j)=\frac{\alpha\Delta z}{2\beta\Delta y^2},$$

$$A_4(h,j)=\frac{\alpha\Delta z}{2\beta\Delta y^2},$$

$$A_5(h,j)=\frac{\alpha\Delta z}{2\beta}\left[k_0^2n^2(h,j,l+1)-\beta^2+\frac{2}{\Delta x^2}\left(\frac{n(h+1,j,l+1)-n(h,j,l+1)}{n(h+1,j,l+1)+n(h,j,l+1)}\right.\right.$$
$$\left.\left.+\frac{n(h-1,j,l+1)-n(h,j,l+1)}{n(h-1,j,l+1)+n(h,j,l+1)}\right)-\frac{2}{\Delta x^2}-\frac{2}{\Delta y^2}\right]+i,$$

$$B_1(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta x^2}\left[1+\frac{2(n(h+1,j,l)-n(h,j,l))}{n(h+1,j,l)+n(h,j,l)}\right],$$

$$B_2(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta x^2}\left[1+\frac{2(n(h-1,j,l)-n(h,j,l))}{n(h-1,j,l)+n(h,j,l)}\right],$$

$$B_3(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta y^2},$$

$$B_4(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta y^2},$$

$$B_5(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta}\left[k_0^2n^2(h,j,l)-\beta^2+\frac{2}{\Delta x^2}\left(\frac{n(h+1,j,l)-n(h,j,l)}{n(h+1,j,l)+n(h,j,l)}\right.\right.$$
$$\left.\left.+\frac{n(h-1,j,l)-n(h,j,l)}{n(h-1,j,l)+n(h,j,l)}\right)-\frac{2}{\Delta x^2}-\frac{2}{\Delta y^2}\right]+i,$$

$$B_6(h,j)=-\frac{\Delta z(n(h+1,j+1,l)+n(h,j+1,l)-n(h+1,j-1,l)-n(h,j-1,l))}{4\beta\Delta x\Delta y(n(h+1,j,l)+n(h,j,l))},$$

$$B_7(h,j)=-\frac{\Delta z(n(h,j-1,l)+n(h-1,j-1,l)-n(h,j+1,l)-n(h-1,j+1,l))}{4\beta\Delta x\Delta y(n(h-1,j,l)+n(h,j,l))},$$

$$B_8(h,j)=B_6(h,j)+B_7(h,j),$$

$$B_9(h,j)=-\frac{n(h+1,j,l+1)+n(h,j,l+1)-n(h+1,j,l)-n(h,j,l)}{2\beta\Delta x\left(n(h+1,j,l)+n(h,j,l)\right)},$$

$$B_{10}(h,j)=-\frac{n(h,j,l)+n(h-1,j,l)-n(h,j,l+1)-n(h-1,j,l+1)}{2\beta\Delta x\left(n(h-1,j,l)+n(h,j,l)\right)},$$

$$B_{11}(h,j)=B_9(h,j)+B_{10}(h,j),$$

$$C_1(h,j)=\frac{\alpha\Delta z}{2\beta\Delta x^2},$$

$$C_2(h,j)=\frac{\alpha\Delta z}{2\beta\Delta x^2},$$

$$C_3(h,j)=\frac{\alpha\Delta z}{2\beta\Delta y^2}\left[1+\frac{2\left(n(h,j+1,l+1)-n(h,j,l+1)\right)}{n(h,j+1,l+1)+n(h,j,l+1)}\right],$$

$$C_4(h,j)=\frac{\alpha\Delta z}{2\beta\Delta y^2}\left[1+\frac{2\left(n(h,j-1,l+1)-n(h,j,l+1)\right)}{n(h,j-1,l+1)+n(h,j,l+1)}\right],$$

$$\begin{aligned}C_5(h,j)=\frac{\alpha\Delta z}{2\beta}\Bigg[&k_0^2n^2(h,j,l+1)-\beta^2+\frac{2}{\Delta y^2}\left(\frac{n(h,j+1,l+1)-n(h,j,l+1)}{n(h,j+1,l+1)+n(h,j,l+1)}\right.\\&\left.+\frac{n(h,j\text{-}1,l+1)-n(h,j,l+1)}{n(h,j\text{-}1,l+1)+n(h,j,l+1)}\right)-\frac{2}{\Delta x^2}-\frac{2}{\Delta y^2}\Bigg]+i,\end{aligned}$$

$$D_1(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta x^2},$$

$$D_2(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta x^2},$$

$$D_3(h,j) = -\frac{(1-\alpha)\Delta z}{2\beta\Delta y^2}\left[1+\frac{2\big(n(h,j+1,l)-n(h,j,l)\big)}{n(h,j+1,l)+n(h,j,l)}\right],$$

$$D_4(h,j) = -\frac{(1-\alpha)\Delta z}{2\beta\Delta y^2}\left[1+\frac{2\big(n(h,j-1,l)-n(h,j,l)\big)}{n(h,j-1,l)+n(h,j,l)}\right],$$

$$D_5(h,j) = -\frac{(1-\alpha)\Delta z}{2\beta}\left[k_0^2 n^2(h,j,l)-\beta^2+\frac{2}{\Delta y^2}\left(\frac{n(h,j+1,l)-n(h,j,l)}{n(h,j+1,l)+n(h,j,l)}\right.\right.$$
$$\left.\left.+\frac{n(h,j-1,l)-n(h,j,l)}{n(h,j-1,l)+n(h,j,l)}\right)-\frac{2}{\Delta x^2}-\frac{2}{\Delta y^2}\right]+i,$$

$$D_6(h,j) = -\frac{\Delta z\big(n(h+1,j+1,l)+n(h+1,j,l)-n(h-1,j+1,l)-n(h-1,j,l)\big)}{4\beta\Delta x\Delta y\big(n(h,j+1,l)+n(h,j,l)\big)},$$

$$D_7(h,j) = -\frac{\Delta z\big(n(h-1,j,l)+n(h-1,j-1,l)-n(h+1,j,l)-n(h+1,j-1,l)\big)}{4\beta\Delta x\Delta y\big(n(h,j-1,l)+n(h,j,l)\big)},$$

$$D_8(h,j) = D_6(h,j)+D_7(h,j),$$

$$D_9(h,j) = -\frac{n(h,j+1,l+1)+n(h,j,l+1)-n(h,j+1,l)-n(h,j,l)}{2\beta\Delta y\big(n(h,j+1,l)+n(h,j,l)\big)},$$

$$D_{10}(h,j) = -\frac{n(h,j,l)+n(h,j-1,l)-n(h,j,l+1)-n(h,j-1,l+1)}{2\beta\Delta y\big(n(h,j-1,l)+n(h,j,l)\big)},$$

$$D_{11}(h,j) = D_9(h,j)+D_{10}(h,j),$$

$$G_1(h,j) = \frac{\alpha\Delta z}{2\beta\Delta x^2},$$

$$G_2(h,j)=\frac{\alpha\Delta z}{2\beta\Delta x^2},$$

$$G_3(h,j)=\frac{\alpha\Delta z}{2\beta\Delta y^2},$$

$$G_4(h,j)=\frac{\alpha\Delta z}{2\beta\Delta y^2},$$

$$G_5(h,j)=\frac{\alpha\Delta z}{2\beta}\left[k_0^2 n^2(h,j,l+1)-\beta^2+\frac{2\big(n(h,j,l+1)-n(h,j,l)\big)}{\alpha\Delta z^2 n(h,j,l+1)}\right.$$
$$\left.-\frac{2}{\Delta x^2}-\frac{2}{\Delta y^2}\right]+i,$$

$$H_1(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta x^2},$$

$$H_2(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta x^2},$$

$$H_3(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta y^2},$$

$$H_4(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta\Delta y^2},$$

$$H_5(h,j)=-\frac{(1-\alpha)\Delta z}{2\beta}\left[k_0^2 n^2(h,j,l)-\beta^2-\frac{2\big(n(h,j,l+1)-n(h,j,l)\big)}{(1-\alpha)\Delta z^2 n(h,j,l)}\right.$$
$$\left.-\frac{2}{\Delta x^2}-\frac{2}{\Delta y^2}\right]+i,$$

$$H_6(h,j) = -\frac{n(h+1,j,l+1) - n(h-1,j,l+1)}{2\beta\Delta x\, n(h,j,l+1)},$$

$$H_7(h,j) = -\frac{n(h-1,j,l) - n(h+1,j,l)}{2\beta\Delta x\, n(h,j,l)},$$

$$H_8(h,j) = -\frac{n(h,j+1,l+1) - n(h,j-1,l+1)}{2\beta\Delta y\, n(h,j,l+1)},$$

$$H_9(h,j) = -\frac{n(h,j-1,l) - n(h,j+1,l)}{2\beta\Delta y\, n(h,j,l)}.$$

Appendix D

The source code change

The program of the fully vectorial finite difference beam propagation method has been compiled and executed on the super computer Cray C98/4256. In the beginning, an unexpected low speed was obtained. The investigation on the source code revealed that the innermost *DO* loop in a CPU time-consuming subroutine had not been vectorized, which is critical for the calculation speed of this computer, i.e.

```
  ......
  DO N = 1, MAX
     ANORM = 0.0
     DO J = 2, JMAX-1
        DO L = 2, JMAY-1
           IF (MOD(J+L,2).EQ.MOD(N,2)) THEN
              RESID = A(J,L)*UX(J+1,L) + B(J,L)*UX(J-1,L) + C(J,L)*UX(J,L+1)
\                      + D(J,L)*UX(J,L-1) + E(J,L)*UX(J,L)-F(J,L)
              ANORM = ANORM + SQRT(RESID*CONJG(RESID))
              UX(J,L) = UX(J,L) - OMEGA*RESID/E(J,L)
           ENDIF
        END DO
     END DO
     IF ((N.GT.1).AND.(ANORM.LE.EEE*ANORMF)) GOTO 16
  END DO
  ......
```

The specialized personnel from SARA[*] suggested the following change:

```
......

DO N = 1, MAX
  ANORM = 0.0
  DO J = 2, JMAX-1
    IF (MOD(J,2).EQ.MOD(N,2)) THEN
      LB = 2
    ELSE
      LB = 3
    ENDIF
    DO L = LB, JMAY-1, 2
        RESID = A(J,L)*UX(J+1,L) + B(J,L)*UX(J-1,L) + C(J,L)*UX(J,L+1)
\                 + D(J,L)*UX(J,L-1) + E(J,L)*UX(J,L)-F(J,L)
        ANORM = ANORM + SQRT(RESID*CONJG(RESID))
        UX(J,L) = UX(J,L) - OMEGA*RESID/E(J,L)
    END DO
  END DO
  IF ((N.GT.1).AND.(ANORM.LE.EEE*ANORMF)) GOTO 16
END DO
......
```

As a result, the vectorization can be fully realized, leading to a considerable increase of the calculation speed on the Cray.

---

[*] Stichting Academisch Rekencentrum Amsterdam

Appendix E

The fiber-chip coupling efficiency

In Section 4.4, an approximate expression for calculating the fiber-chip coupling efficiency has been given in Eq.(4.3), i.e.

$$\eta = \frac{2b(f_1 + f_2)^2}{(a^2 + b^2)(c_1 + c_2)}, \tag{E.1}$$

with

$$f_1 = \frac{a\,c_1}{\sqrt{a^2 + c_1^2}} \exp\left(-\frac{\Delta^2}{a^2 + c_1^2}\right)\left[1 - \operatorname{erf}\left(-\frac{c_1\Delta}{a\sqrt{a^2 + c_1^2}}\right)\right], \tag{E.2a}$$

$$f_2 = \frac{a\,c_2}{\sqrt{a^2 + c_2^2}} \exp\left(-\frac{\Delta^2}{a^2 + c_2^2}\right)\left[1 + \operatorname{erf}\left(-\frac{c_2\Delta}{a\sqrt{a^2 + c_2^2}}\right)\right], \tag{E.2b}$$

The derivation is as follows. First, the coupling efficiency η can be expressed as

$$\eta = \frac{\left|\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left(FG^*\right)dx\,dy\right|^2}{\left|\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left(FF^*\right)dx\,dy\right|\left|\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty}\left(GG^*\right)dx\,dy\right|}, \tag{E.3}$$

where F and G are the scalar electric fields of the fiber and the waveguide, respectively:

$$F = F_0 \exp\left(-\frac{(x-\Delta)^2 + y^2}{a^2}\right), \tag{E.4}$$

$$G = G_0 \exp\left(-\frac{x^2}{c_j^2} - \frac{y^2}{b^2}\right), \tag{E.5}$$

with Δ as the vertical offset between the fiber and the waveguide. If $x > 0$, we use $j = 1$; if $x < 0$, we write $j = 2$ (see Figure 4.12). Then, we can obtain:

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(FF^*\right) dx\, dy &= F_0^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\exp\left(-\frac{2(x-\Delta)^2}{a^2} - \frac{2y^2}{a^2}\right)\right] dx\, dy \\
&= F_0^2 \int_{-\infty}^{\infty} \left\{\exp\left(-\frac{2y^2}{a^2}\right) \int_{-\infty}^{\infty} \left[\exp\left(-\frac{2(x-\Delta)^2}{a^2}\right)\right] dx\right\} dy \\
&= F_0^2 \frac{a\sqrt{\pi}}{\sqrt{2}} \int_{-\infty}^{\infty} \left[\exp\left(-\frac{2y^2}{a^2}\right)\right] dy \\
&= \frac{\pi a^2}{2} F_0^2.
\end{aligned} \tag{E.6}$$

Similarly, we have

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(GG^*\right) dx\, dy &= G_0^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left[\exp\left(-\frac{2x^2}{c_j^2} - \frac{2y^2}{b^2}\right)\right] dx\, dy \\
&= G_0^2 \int_{-\infty}^{\infty} \left\{\exp\left(-\frac{2y^2}{b^2}\right) \int_{-\infty}^{\infty} \left[\exp\left(-\frac{2x^2}{c_j^2}\right)\right] dx\right\} dy \\
&= G_0^2 \left(\frac{c_1\sqrt{\pi}}{2\sqrt{2}} + \frac{c_2\sqrt{\pi}}{2\sqrt{2}}\right) \int_{-\infty}^{\infty} \left[\exp\left(-\frac{2y^2}{b^2}\right)\right] dy \\
&= \frac{\pi b(c_1 + c_2)}{4} G_0^2.
\end{aligned} \tag{E.7}$$

Furthermore, we can get

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(FG^{*}\right)dx\,dy = F_0G_0\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\exp\left(-\frac{(x-\Delta)^2}{a^2}-\frac{y^2}{a^2}-\frac{x^2}{c_j^2}-\frac{y^2}{b^2}\right)\right]dx\,dy$$

$$= F_0G_0\int_{-\infty}^{\infty}\left\{\exp\left(-\frac{(x-\Delta)^2}{a^2}-\frac{x^2}{c_j^2}\right)\int_{-\infty}^{\infty}\left[\exp\left(-\frac{y^2}{a^2}-\frac{y^2}{b^2}\right)\right]dy\right\}dx$$

$$= F_0G_0\frac{ab\sqrt{\pi}}{\sqrt{a^2+b^2}}\int_{-\infty}^{\infty}\left[\exp\left(-\frac{(x-\Delta)^2}{a^2}-\frac{x^2}{c_j^2}\right)\right]dx \qquad \text{(E.8)}$$

$$= F_0G_0\frac{ab\pi}{2\sqrt{a^2+b^2}}(f_1+f_2),$$

where f_1 and f_2 are the expressions as shown in (E.2a) and (E.2b). Inserting (E.6-8) into (E.3), (E.1) is obtained.

Analysis, design and fabrication of tapered integrated optical structures

Implementation of the fully vectorial 3-D beam propagation method

Daoping Li

Summary

This thesis reports the development of a fully vectorial finite difference beam propagation method for the analysis of structures with varying cross-sections and a mathematical analysis of tapered integrated optical structures as required for fiber-chip coupling, together with its implementation in a computer program. Using this method, the design of an efficient fiber-chip coupling structure has been performed, combined with an analysis of its efficiency. Finally, the technology to produce such structures is discussed.

Tapered InGaAsP waveguide structures on an InP substrate, serving as mode size transformers, are promising configurations for an efficient fiber-chip coupling. To design satisfactory structures, a fully vectorial FD-BPM has been developed. The vectorial formulations are based upon the Maxwell equations and use the slowly varying envelope approximation. Waveguide structures are discretized by introducing grids. The above formulations are expressed in discrete form on the grid nodes by use of a finite difference scheme. The relaxation method is applied to solve the discrete expressions numerically. The wave propagation is calculated step by step along the longitudinal direction. A computer program is developed using the FORTRAN 77 language. The various aspects of the programming such as program structure, arrays, numerical considerations, double precision, debugging and optimization are discussed. A benchmark test on the SUN 10/40, the HP 735 and the Cray C98/4256 computers is carried out, showing the speed ratios: 1 : 2.6 : 35 after optimization. Acceleration procedures have been introduced, such as a modified Crank-Nicholson scheme, a decreased initial error, a reduced iteration parameter and transparent boundary conditions. Waveguide structures like a rib waveguide, a ridge waveguide, a polarization splitter and a polarization converter are simulated and evaluated, verifying the validity and accuracy of our vectorial FD-BPM. The optimal reference propagation constant, which corresponds to the least spurious

numerical loss, is found to be close to the propagation constant of the guided mode. There also exists a critical value α_c of the Crank-Nicholson scheme parameter α for each waveguide structure. If $\alpha < \alpha_c$, the scheme is unstable, leading to an exponentially increasing field in the propagation direction; if $\alpha > \alpha_c$, the scheme is stable but leads to a spurious numerical loss. Under the condition $\alpha > \alpha_c$, the bigger the value of α is, the larger its influence will be. If a non-optimal reference propagation constant β has been chosen, the influence of the scheme parameter on the results is stronger than when using an optimal β. Therefore, the reference propagation constant should be chosen carefully and α is chosen as small as possible (but $\alpha \geq \alpha_c$). Although big mesh sizes correspond to high simulation speeds, the mesh sizes should be chosen small enough to discretize the waveguide structure precisely, which is here for the presented structures about 0.01-0.05 μm in the vertical direction, 0.05-0.2 μm in the lateral direction and 0.01-0.2 in the longitudinal direction. The simulation speed of our vectorial FD-BPM depends on the mesh sizes and computational window size, typically in the order of 100 μm/hour along the propagation direction.

Designed 3D tapered waveguide structures include five sections. The first waveguide section is a straight $Q_{1.3}$ waveguide of 0.6 μm thickness and 2 μm width on an InP substrate, the second section is laterally tapered with an end width chosen such that its lateral mode size matches that of the fiber (10 μm) and with a length such that excited higher order modes are negligible at the end. Then again a straight transition section occurs, followed by a vertically tapered section. The insertion of the transition section between them is for technological convenience, since two tapered sections are realized by different process steps. The waveguide should be tapered to a small thickness of about 25 nm to obtain a large vertical mode size and the section length has to exceed 300 μm to lower the radiation loss below 0.5 dB. Since a small change of the waveguide thickness close to cut-off leads to a considerable mode size change, the thickness should have a small tolerance of about 5 nm to realize a high fiber-chip coupling efficiency. In addition, there is another straight waveguide section following the vertically tapered section, designed to obtain a flat wave front as required for an effective fiber-chip coupling. Finally, the whole structure is covered with a thick layer (5 μm) of high index material, matching that of the substrate in order to obtain a vertically symmetrical field distribution matching that of the fiber. Thus, the calculated total fiber-chip coupling loss can be below 1 dB.

The lateral pattern definition of the designed structures can be realized by a conventional lithography process. The vertically tapered section was realized by atomic beam etching combined with a shadow mask. There is a spacer between the mask and the chip. Linearly tapered profiles have been obtained, their lengths depend on the spacer thickness. Their depths are dependent on the etching time, the etching speed being about 5 nm/min. Thus, the required section length and the depth can be controlled by regulating the spacer thickness and the etching time. The surface roughness after atomic beam etching is in the order of 3-4 nanometers.

A new process to grow a high index silicon nitride layer by PECVD instead of an InP layer by MOVPE has been developed. The index can be controlled with a deviation of within ±0.005 from that of InP by regulating the gas flow ratio. The deposition speed ranges from 150 - 280 nm/min, depending on the plasma power. A silicon nitride layer of 5 μm thick can be deposited within one hour. In addition, the bulk attenuation of the cover layer has been effectively decreased to 45 dB/cm which corresponds to an extra power loss of 1 dB to the designed structure due to the cover absorption. Through further optimization, this new cost-saving process can be expected to reach a practically accepted level and have numerous applications in integrated optics.

Analyse, ontwerp en productie van schuin verlopende geïntegreerd-optische structuren

Implementatie van de volledig vectoriële driedimensionale bundelpropagatie methode

Daoping Li

Samenvatting

Dit proefschrift behandelt de eindige verschillen bundelpropagatie methode (FD-BPM) voor de analyse, in volledig vectoriële vorm, van structuren met een variërende doorsnede te zamen met de mathematische analyse van velden in schuin verlopende geïntegreerd-optische structuren zoals benodigd voor koppeling van een glasfiber aan een geïntegreerd-optische schakeling. Tevens behandelt dit werk de implementatie van deze berekeningen in een computer programma. Met behulp van dit programma is een efficiënte koppeling ontworpen, gecombineerd met een bepaling van het koppelrendement. Tenslotte wordt de technologie, benodigd voor dit soort structuren besproken.

Schuin verlopende InGaAsP golfgeleider structuren op een InP substraat, gebruikt als modale veld omzetter, zijn veelbelovend voor een efficiënte koppeling. Voor het ontwerp van bevredigende structuren is een volledig vectoriële FD-BPM ontwikkeld. De vectoriële formuleringen zijn gebaseerd op de vergelijkingen van Maxwell en maken gebruik van de benadering dat de omhullende van de golffunctie langzaam verandert. De golfgeleider structuren zijn ruimtelijk onderverdeeld door de introductie van een rooster, waardoor een discretisatie van de differentiaalvergelijkingen mogelijk wordt. De betreffende formuleringen zijn uitgedrukt in veldgrootheden in de roosterpunten met behulp van een eindige-verschillen schema. Voor de numerieke oplossing van de discrete vergelijkingen wordt de relaxatie-methode toegepast. De golfvoortplanting in de longitudinale richting wordt stap voor stap berekend. Het computer programma is geschreven in de programmertaal FORTRAN 77. Verscheidene aspecten van het programmeren, zoals structuur, arrays, numerieke methoden, dubbele precisie, foutzoeken en optimalisatie worden besproken. Een 'benchmark' test is uitgevoerd voor de computers SUN 10/40, HP 735 en Cray C98/4256. Na optimilisatie bleek het resultaat voor de

snelheidsverhoudingen 1 : 2.6 : 35. Versnellingsmethoden zijn geïntroduceerd, zoals een gemodificeerd Crank-Nicholson schema, een verbeterde startwaarde, een gereduceerde iteratie-parameter en transparante randvoorwaarden. Als controle op de betrouwbaarheid van onze FD-BPM zijn golfgeleider structuren, zoals een dijk golfgeleider, een begraven golfgeleider, een polarisatie-omzetter en een polarisatie-splitser gesimuleerd en geëvalueerd. Voor de optimale referentie-voortplantingsconstante, die hoort bij de laagste numerieke restfout, is vastgesteld dat deze zeer dicht bij de voortplantingsconstante van de geleide modus ligt. Er bestaat een kritische parameter α_c voor de parameter α in het Crank-Nicholson schema voor iedere golfgeleider structuur. Indien $\alpha < \alpha_c$, is het schema instabiel hetgeen leidt tot een exponentieel toenemend veld tijdens voortplanting; voor $\alpha > \alpha_c$ is het schema stabiel maar leidt tot (numeriek) verlies van voortgeplant vermogen. Onder de voorwaarde $\alpha > \alpha_c$ geldt dat hoe groter α is, des te groter de invloed ervan zal zijn. Indien een niet-optimale referentie voortplantingsconstante β gekozen is, dan zal de invloed van α sterker zijn dan in het geval van een optimale β. Hieruit kunnen we concluderen dat β zeer nauwkeurig bepaald moet worden en vervolgens α zo klein mogelijk, doch $\alpha \geq \alpha_c$. Ofschoon een grote maaswijdte in het rekenrooster een grote simulatie-snelheid zal opleveren, moet er voor gezorgd worden dat de mazen klein genoeg zijn om het veld nauwkeurig te kunnen discretiseren. Dit komt in de hier gepresenteerde structuren neer op 0.01-0.05 µm in de verticale richting, 0.05-0.2 µm in de laterale richting en 0.01-0.2 µm in de longitudinale richting. De simulatiesnelheid in longitudinale richting van onze vectoriële FD-BPM hangt af van het rooster en het rekenvenster en is typisch in de orde van 100 µm/uur.

De ontworpen schuin verlopende structuren voor koppeling bestaan uit vijf secties. De eerste sectie is een rechte $Q_{1.3}$ golfgeleider met een dikte van 0.6 µm en een breedte van 2 µm op een InP substraat. De tweede sectie verloopt schuin in de laterale richting, waarbij de eindbreedte zo gekozen is dat de veldbreedte gelijk wordt aan die van een fiber (10 µm) en met een lengte zodat geen hogere orde modi geëxciteerd worden. De derde sectie is weer een rechte golfgeleider, gevolgd door de vierde die in de vertikale richting schuin verloopt. Het toevoegen van de derde sectie heeft technologische redenen; de twee schuin verlopende secties worden in twee verschillende processtappen vervaardigd. De golfgeleider moet in sectie vier tot een dikte van 25 nm teruggebracht worden om een voldoende groot verticaal veldprofiel te krijgen en de lengte moet groter zijn dan 300 µm om de afstraalverliezen beneden de 0.5 dB te houden. Omdat een kleine verandering in de dikte, dicht bij de afsnijdings-omstandigheden, tot een grote verandering in de modus-afmeting leidt moet de dikte met een nauwkeurigheid van ±5 nm geproduceerd worden.

De vijfde sectie is een rechte golfgeleider, zodanig ontworpen dat een vlak golffront verkregen wordt. Tenslotte wordt de totale structuur bedekt met een dikke laag (5 μm) materiaal met hoge brekingsindex, bij voorkeur met een index gelijk aan die van het substraat om een symmetrisch veld te verkrijgen. Het verlies in de totale koppeling voor die configuratie is bepaald op minder dan 1 dB.

De laterale structuur kan met behulp van conventionele fotolithografische technieken vervaardigd worden. De vertikale helling is gerealiseerd door atoombundel-etsen gecombineerd met een schaduw-masker. Er is een afstandstuk tussen dit masker en het geïntegreerde circuit. Lineaire profielen zijn vervaardigd, waarbij de lengte bepaald wordt door de dikte van dit afstandstuk. De helling hangt af van de etstijd (de etssnelheid is ongeveer 5 nm/min). Concluderend kunnen we stellen dat de benodigde sectielengte en helling ingesteld kunnen worden door keuze van de dikte van het afstandstuk en de etstijd. De verkregen oppervlakteruwheid na atoombundel-etsen is circa 3-4 nm.

Voor het verkrijgen van een dikke afdeklaag met een hoge brekingsindex is een nieuwe techniek ontwikkeld waar bij een laag van siliciumnitride met behulp van PECVD gegroeid wordt, in plaats van InP door MOVPE. De brekingsindex kan ingesteld worden met een maximale afwijking van ±0.005 ten opzichte van die van InP door het regelen van de gastoevoeren. De depositiesnelheid ligt in de buurt van 150-280 nm/min, afhankelijk van het plasmavermogen. Een laag siliciumnitride met een dikte van 5 μm kan binnen een uur gedeponeerd worden. Door verbetering van de procesgang is de absorptie van de deklaag verlaagd tot 45 dB/cm, hetgeen overeen komt met een extra verlies in de ontworpen structuur van 1 dB ten gevolge van deklaag-absorptie. Door verdere optimalisatie kan van deze techniek, die sterk kosten besparend is, verwacht worden dat een in de praktijk acceptabel niveau bereikt wordt waarvoor vele toepassingen zullen bestaan binnen de geïntegreerde optica.

锥形集成光学结构的分析、设计和制备

全矢量束传播法的研发及应用

综　述

本书报道全矢量差分束传播法及其计算机软件的研发过程，该方法可用于分析变截面的波导结构。运用这一方法，我们设计了锥形纤维芯片偶合结构，并预测其效率。最后，我们报道和讨论此类结构的制造工艺。

以ＩｎＰ为衬底的ＩｎＧａＡｓＰ波导结构，通过转变导模的尺寸，实现纤维芯片间的高效偶合。为了设计此类结构，必须开发一个全矢量的束传播法，从而计算波导结构中的传播情形。公式推导基于Ｍａｘｗｅｌｌ方程组，并运用慢变场形近似。波导结构用网格代替，通过差分原理来表达导出的公式，并用松弛法求解。波的传播被沿长度方向逐步计算。开发的计算机软件使用ＦＯＲＴＲＡＮ语言。讨论了与编程有关的许多问题，如：程序安排、数组、数值处理、双精度、查错和优化。程序分别在三种计算机上运行，包括ＳＵＮ　１０／４０、ＨＰ　７３５工作站和Ｃｒａｙ　Ｃ９８／４２５６超级计算机，结果速度比为１：２．６：３５。提出了几个加速步骤，如：修改的Ｃｒａｎｋ－Ｎｉｃｈｏｌｓｏｎ几何、减少的原始偏差、缩小的迭代参数和透明的边界条件。通过如下波导结构的模拟：条形、锥形、极化分离和极化转变结构，验证了全矢量束传播法的准确性。对于每一波导结构，存在一个最佳参考传播常数，恰好近似于导模的传播常数。另外，也存在一个临界差分参数。如果小于这个值，差分几何将不稳定，导致传播场不切实的指数增加；如果大于这个值，尽管稳定，但传播场有一个数值方法本身带来的假性衰减。在大于临界参数的前提下，所取参数越大，影响相对越明显，特别是在参考传播常数没有选取最佳值时，则影响进一步加剧。反之，差分参数的影响则减弱。因此，参考传播常数要正确选取，同时差分参数要尽可能小。虽然网格尺寸较大时，计算速度增加，但网格尺寸的选取应保证波导结构能被细密地网格化，通常纵向尺寸０．０１－０．０５微米、横

向０．０５－０．２微米、轴向０．０１－０．２微米。计算速度取决于网格尺寸和计算窗尺寸的选取，通常以每小时１００微米的计算速度进行传播模拟。

设计的三维锥形波导结构包含五个部分：第一部分波导０．６微米厚2微米宽，轴向不变；第二部分横向由２微米至１０微米逐步拓宽，长度选取要保证产生的高级模可以忽略不计；第三部分轴向不变，引进的目的是为了便于制备；紧接着是第四部分，纵向由０．６微米至０．０２５微米逐步变薄，其长度选取应超过４００微米以保证辐射损失低于０．５分贝。由于接近ｃｕｔ－ｏｆｆ时波导厚度的极小变化可引起波导模尺寸的明显变化，厚度的允许偏差应小于5纳米。最后一个部分仍然轴向不变，其目的是为了获得平面波前，以提高纤维芯片偶合效率。

设计结构的横向实现应用常规的照相制版工艺，而纵向减薄则通过原子束剥蚀和锥形掩模的组合。在掩模与芯片之间放置一垫片，线性变薄的波导形状已经获得，其长度取决于垫片厚度，而深度取决于剥蚀时间，剥蚀速度为每分钟5纳米。因此，可以通过调节垫片厚度和剥蚀时间来控制纵向波导部分的形状。原子束剥蚀后的表面粗糙度约为３－４纳米。

我们开发了一个用等离子增强化学气相沉积（ＰＥＣＶＤ）制备高折射率氮化硅层的新工序，从而代替由金属氧化物气相外延（ＭＯＶＰＥ）生长的ＩｎＰ层。氮化硅层的折射率可通过改变气体比来调节。沉积速度约为每分钟１５０－２８０纳米，取决于等离子功率。５微米厚的氮化硅层可在一小时内制备完成。另外，氮化硅层的功率衰减已被降低到每厘米４５分贝以下，从而使由于氮化硅层吸收而引入设计结构的额外功率损耗低于１分贝。通过进一步的优化，该工序可望达到实用水准，将在集成光学领域具有广泛的应用前景。

Index

Biography

Daoping Li was born on March 5, 1963 in Anhui, China. After graduation from Suzhou No.1 High School, Anhui province in 1979, he entered Zhejiang University and obtained B.Sc. and M.Sc. from Dept. of Materials Science and Engineering in 1983 and 1986, respectively. Then he became a staff member as an assistant lecturer (9/86 - 9/88) and lecturer (9/88 - 5/90) in the same university. Subsequently, he came to the Netherlands working first for KEMA N.V. (5/90 - 5/91) and then for the TNO Institute of Applied Physics (5/91 - 9/91) as a researcher. Since September 1991, he joined the Optics Research Group, to carry out research as an "Assitent in Opleiding". The subject of this investigation concerned integrated optics, developing a fully vectorial 3-D beam propagation method and performing the analysis, design and fabrication of tapered integrated optical structures for fiber-chip coupling.