



University
of Glasgow

Masoudi, Husain Muhammad Y. (1995) *Parallel numerical methods for analysing optical devices with the BPM*. PhD thesis.

<http://theses.gla.ac.uk/2858/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Parallel Numerical Methods for Analysing Optical Devices with the BPM

A Thesis Submitted to the
Faculty of Engineering of the University of Glasgow
for the Degree of Doctor of Philosophy

by

Husain Muhammad Y. Masoudi

Copyright © 1995 by Husain M. Y. Masoudi.

In the name of God, Most Gracious, Most Merciful.

“God is the Light of heavens and the earth. The parable of His Light is as if there were a Niche and within it a Lamp: The Lamp enclosed in Glass: The glass as it were A brilliant star: Lit from a blessed Tree, An Olive, neither of the East nor of the West, Whose Oil is well-nigh Luminous, Though fire scarce touched it: Light upon Light! God doth guide whom he will To His Light: God doth set forth Parables For men (people): and God doth know all things.”

**The Holy Qura'an
Al-Noor chapter (35)**

*To my mother the one who taught me how to love.
To my father the one who taught me how to be patient.
To my wife Ghadeer for caring and continuous support.
To my little son Hesham.
To all my family with respect.*

Acknowledgements

Thanks are due to the Department of Electronics and Electrical Engineering at the University of Glasgow for giving me the opportunity to work on this research. I would like to express my deep appreciation to my supervisor Professor John M. Arnold for his constant help and encouragement through out this work. I would like also to thank the lecturers of the Optoelectronics Group of this Department and my colleagues for providing useful research discussion environment. Thanks are also to all my friends for their support.

The support of King Fahd University of Petroleum and Minerals (Dhahran, Saudi Arabia) while I am in Glasgow is gratefully acknowledged. Thanks also to the Edinburgh Parallel Computing Centre (EPCC) at the University of Edinburgh for the access to the Connection Machine.

Abstract

In this work, some developments in the theory of modelling integrated optical devices are discussed. The theory of the Beam Propagation Method (BPM) to analyse longitudinal optical waveguides is established. The BPM is then formulated and implemented numerically to study both two and three-dimensional optical waveguides using several Finite-Difference (FD) techniques. For the 2-D analysis, comparisons between the performance of the implicit Crank Nicholson (CN), the explicit Real Space (RS) and the Explicit Finite-Difference (EFD) are made through systematic tests on slab waveguide geometries. For three-dimensional applications, two explicit highly-parallel three-dimensional FD-BPMs (the RS and the EFD) have been implemented on two different parallel computers, namely a transputer array (MIMD type) and a Connection Machine (SIMD type). To assess the performance of parallel computers in this context, serial computer codes for the two methods have been implemented and a comparison between the speed of the serial and parallel codes has been made. Large gains in the speed of the parallel FD-BPMs have been obtained compared to the serial implementations; both methods, in their parallel form, can execute, per propagational step, a large problem containing 10^6 discretisation points in a few seconds. In addition, a comparison between the performance of the transputer array and the Connection Machine in executing the two FD-BPMs has been discussed. To assess and compare the two methods, three different rib waveguides and three different directional couplers have been analysed and the results compared with published results. It has been concluded from testing these methods that the parallel EFD-BPM is more efficient than the parallel RS-BPM. Then, the linear parallel EFD-BPM was extended to model nonlinear second harmonic generation process in three-dimensional waveguides, where the source field is allowed to deplete, using the transputer array and the Connection Machine. The new nonlinear version of the EFD-BPM has the same features of the linear counterpart in using two separate computational windows, one for the fundamental field and the other for the second harmonic field. The implementation and the application of this method to a nonlinear rib waveguide using the Quasi-Phase-Matching technique are discussed in this work. It has been concluded that the new nonlinear parallel EFD-BPM is accurate, efficient and very useful for the analysis of large complicated optical devices.

Publications

From the work of this thesis the following have been published or submitted for publication

- [1] Masoudi, H. M. and Arnold, J. M., "Parallel-Processing on a Transputer Array for the Numerical Modelling of Rib Waveguides", Presented at the UK. National Radio Science Conference at Leeds, April 1993.
- [2] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Methods", *IEEE Photonics Technology Letters*, 6, 848-850, July 1994.
- [3] Masoudi, H. M. and Arnold, J. M., "Parallel Three-Dimensional Finite-Difference Beam Propagation Methods" in the *International Journal of Numerical Modelling*, 8, 95-107, March-April 1995.
- [4] Masoudi, H. M. and Arnold, J. M., "Parallel-Processing Finite-Difference Beam Propagation Methods", Presented at the *International Symposium on Guided-Wave Optoelectronics: Device Characterization, Analysis and Design*, IEEE and Optical Society of America in Polytechnic University, New York, October 1994.
- [5] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Method For The Analysis of Second Harmonic Generation", *IEEE Photonics Technology Letters*, 7, 400-402, April 1995.
- [6] Masoudi, H. M. and Arnold, J. M., "Modelling Second-Order Nonlinear Effects in Optical Waveguides Using a Parallel-Processing Beam Propagation Method", submitted to the *IEEE Journal of Quantum Electronics*, October 1994.
- [7] Masoudi, H. M. and Arnold, J. M., "Three-Dimensional Numerical Modelling of the Second-Order Cascaded Effect in a Semiconductor Rib Waveguide Using Parallel-Processing", presented at the *Optical Society of America Topical meeting: Nonlinear Guided Waves and Their Applications*, Dana Point, California, February 1995.
- [8] Masoudi, H. M. and Arnold, J. M., "Numerical Modelling of Second Harmonic Generation in an Optical Waveguide Using Parallel-Processing Finite-Difference Beam Propagation Method", presented at the *URSI Electromagnetic Theory Symposium*, St. Petersburg, Russia, May 1995.

Symbols

Symbol	Description	Value	Dimension
A	Transverse cross sectional area	-	m ²
\mathbf{a}_{norm}	Normal unit vector	-	-
\mathbf{B}	Vector magnetic flux density	-	T
c	Speed of light in vacuum	3×10^8	m/s
\mathbf{D}	Vector electric displacement	-	C/m ²
δ	Delta function	-	-
Δz	A small longitudinal step in the z-direction	-	m
Δx	The mesh spacing in the x-direction	-	m
Δy	The mesh spacing in the y-direction	-	m
Δk	The wave vector mismatch	-	1/m
\mathbf{E}	Vector electric field	-	V/m
E	Scalar electric field	-	V/m
\mathcal{E}	Normalised Electric field	-	V/m
ϵ	Permittivity	-	F/m
ϵ_0	Free space permittivity	8.85×10^{-12}	F/m
ϵ_r	Relative permittivity equals to $(1+\chi)$	-	-
Φ	A scalar electric field without a reference phase	-	V/m
ϕ	A parabolic scalar electric field of Φ	-	V/m
\mathbf{H}	Vector magnetic field	-	A/m
H	Scalar magnetic field	-	A/m
\mathcal{H}	Normalised magnetic field	-	A/m
\mathbf{J}	Current density	-	A/m ²
j	Complex numbers notation	$(\sqrt{-1})$	-
k_0	Wave vector in vacuum	-	1/m
k_x	Wave vector in the direction x (transverse)	-	1/m
k_y	Wave vector in the direction y (transverse)	-	1/m
k_z	Wave vector in the direction z (propagation)	-	1/m
Λ	The grating period of second order nonlinearity	-	m
χ	Linear susceptibility	-	-
$\chi^{(2)}$	Second order nonlinear susceptibility	-	m/V
$\chi^{(3)}$	Third order nonlinear susceptibility	-	m ² /V ²
\mathcal{L}	Line shape function	-	-
L_c	Coupling length	-	m
λ	Wavelength	-	m
∇^2	The Laplacian operator	-	1/m ²
∇_{\perp}^2	The transverse Laplacian operator	-	1/m ²
∇_x^2	The x Laplacian operator	-	1/m ²
∇_y^2	The y Laplacian operator	-	1/m ²
M_x	The total number of mesh points in the x-direction	-	-
M_y	The total number of mesh points in the y-direction	-	-
μ	Permeability	-	H/m
μ	Micro	10^{-6}	prefix

Parallel Numerical Methods for Analysing Optical Devices with the BPM

μ_0	Free space permeability	$4\pi \times 10^{-7}$	H/m
μ_r	Relative permeability	-	-
n	Refractive index equals to $\sqrt{\epsilon_r}$	-	-
n_0	A reference refractive index	-	-
n_{eff}	Effective index	-	-
ω	Angular frequency (radians)	-	s^{-1}
\mathbf{P}	Nonlinear Polarisation	-	V/m
P	Power (for two dimensional waveguides)	-	W/m
P	Power (for three dimensional waveguides)	-	W
ψ	One dimensional z-dependent electric field	-	V/m
v	Normalised frequency	-	-
\mathbf{r}	Vector space location (x,y,z)	-	m
$\text{Re}()$	The real part of the quantity in the bracket	-	-
ρ	Charge density	-	C/m^3
\mathbf{S}	Poynting vector	-	W/m^2
\mathbf{T}	The Transpose of a matrix	-	-
t	Time	-	s
(x,y,z)	Cartesian co-ordinate system	-	m
$w(z)$	Window Function in the longitudinal direction	-	-
Z	Total longitudinal distance	-	m
*	Complex conjugate	-	-

Abbreviations

Term	Description
1-D	One Dimension
2-D	Two Dimensions
3-D	Three Dimensions
ADI	Alternating Direction Implicit
BC	Boundary Conditions
BPM	Beam Propagation Method
CM	Connection Machine
CM-FORTRAN	Connection Machine FORTRAN
CN	Crank Nicholson
EFD	Explicit Finite Difference
EIM	Effective Index Method
FD	Finite Difference
FDTD	Finite Difference Time Domain
FE	Finite Element
FFT	Fast Fourier Transform
MEFD	Modified Explicit Finite Difference
MIMD	Multiple Instruction Multiple Data
NGM	Number of Guided Modes
NP	Number of Processors
QPM	Quasi Phase Matched
RSM	Real Space Method
SHG	Second Harmonic Generation
SIMD	Single Instruction Multiple Data
SIPIM	Shifted Inverse Power Iteration Method
TBC	Transparent Boundary Conditions
TE	Transverse Electric field
TM	Transverse Magnetic field

Table of Contents

<i>Acknowledgements</i>	iii
<i>Abstract</i>	iv
<i>Publications</i>	v
<i>Symbols</i>	vi
<i>Abbreviations</i>	viii

Chapter 1 Introduction	1
1.1 General	1
1.2 The Objective of the Thesis	2
1.3 Organisation of the Thesis	5
REFERENCES	7

Part I: Theory

Chapter 2 Modelling Optical Devices	10
2.1 Introduction	10
2.2 The Wave Equation	11
2.3 The Dielectric Slab Waveguide	15
2.3.1 Transverse Electric (TE) Guided Modes	17
2.3.2 Transverse Magnetic (TM) Guided Modes	19
2.3.3 Guided Mode Computations	20
2.3.4 Guided Modes Power Flow	22
2.3.5 Radiation Modes	23
Orthogonality of modes	24
2.4 Modelling Three-Dimensional Waveguides	26
2.4.1 The Effective Index Method	28
2.4.2 Solutions to the wave equation	30
2.4.3 Solving the matrix eigenvalue numerically	33
REFERENCES	33

Chapter 3 The Beam Propagation Method (BPM)	36
3.1 Introduction	36
3.2 The Parabolic Equation	39
3.3 The BPM Solutions	42
3.3.1 The Fast Fourier Transform (FFT-BPM)	43
3.3.2 The Finite Difference BPM (FD-BPM)	43
3.3.2.1 The Alternating Direction Implicit BPM (ADI-BPM) ..	44
3.3.2.2 The Real Space method (RS-BPM)	45
3.3.2.3 The Explicit Finite Difference Method	49
3.4 The Spectral Method	50
REFERENCES	53

Chapter 4 Parallel Computing	57
4.1 Introduction	57
4.2 The Parallel Machines	58
4.3 Parallel Programming	63
4.3.1 Programming under PARIX	66
4.3.2 Programming under CM-FORTRAN	68
REFERENCES	71

Part II: Implementations (Analysis and assessments)

Chapter 5	Two-Dimensional Beam Propagation Methods	73
5.1	Introduction	73
5.2	Two-Dimensional BPM Implementations	74
5.2.1	Formulation	74
5.2.1.1	The Crank Nicholson method (CN-BPM)	75
5.2.1.2	The Real Space method (RS-BPM)	78
5.2.1.3	The Explicit Finite-Difference method (EFD-BPM)	80
5.3	Analysis and Assessments	82
5.4	The Modified EFD-BPM	96
5.5	Applications	99
5.6	Conclusion	104
	REFERENCES	105
Chapter 6	Parallel Three-Dimensional Beam Propagation Method	107
6.1	Introduction	107
6.2	The Alternating Implicit Method (ADI-BPM)	109
6.2.1	Formulation	109
6.2.2	Parallel implementations	112
6.2.3	Speed	116
6.3	The Explicit Finite-Difference Methods	117
6.3.1	Equations	117
6.3.2	Parallel implementations	118
6.3.3	Speed	121
6.4	Analysis and Assessments	126
	Examples	140
6.5	Discussion and Conclusion	144
	REFERENCES	146
Part III: Nonlinear Optical Waveguides		
Chapter 7	Parallel Second Harmonic Generation BPM	148
7.1	Introduction	148
7.2	Formulation	149
7.2.1	The Second Harmonic Generation EFD-BPM	150
7.2.2	The simplified 1-D Second Harmonic Generation	153
7.3	Parallel Implementations	155
7.4	Analysis and Assessments	157
	The Phase-Matched Case	158
	The Quasi Phase-Matched Case	159
7.5	Conclusion	167
	REFERENCES	167
Chapter 8	Summary and Conclusion	170
	Future Work	172
	REFERENCES	173

Chapter 1

Introduction

1.1 General

Lightwave technology has become increasingly important to the future of the telecommunication field due to the high bandwidth of light (≈ 100 THz). The inventions, first the laser in 1960 as an optical source and later the low-loss optical fiber as a key communication link, have made optical communication becomes a reality [1-7]. However, due to losses and dispersion, the optical signal has to be regenerated (re-amplified) along the path of transmission. Then it was realised that a complete optical system has to be integrated upon a single substrate to allow for efficient optical signal processing [1]. Following these lines the idea of integrated optics emerged. Such complicated circuits of integrated optics use optical waveguides to execute various different functions. In most practical applications, these waveguides have a rectangular geometry (e.g. rib waveguides), where the main purpose of designing the rib waveguide structure is to confine the optical field in two dimensions, vertically by refractive index changes and horizontally by geometrical changes. In some devices the rectangular waveguide has to split into two other waveguides forming a branch coupler (Y-junction) such as the Mach-Zehnder switch, which is the basic element of many integrated optical signal processing devices, and the junction of a ring laser [1]. The practical reason for designing the branch coupler waveguide is to divide the input power into two output waveguides or, conversely, to combine the power from two input waveguides.

In the current state of integrated optics, switching of light between two output channels is executed using the electro-optic effect, where an external dc electric field is used to change the refractive index of the medium resulting in a phase change, over a certain distance, to the guided wave [8]. However, the involvement of electronics in optical communication degrades greatly the use of the allowed bandwidth of light due to the speed limitation of electronics and opto-electronics devices [9]. It seems that large bandwidth is only obtainable using all-optical integration where the modulation and the switching have

to be executed in the optical domain. In this context nonlinear optics plays a very important role for developing an all-optical switch [8,9]. Initially, most of the research was focused on intensity-dependent nonlinear materials in which the wave produces its own phase shift and hence selects, according to its power level, its own output path [8]. Such an effect, where the refractive index of the material depends on the intensity of the wave, has been demonstrated using the $\chi^{(3)}$ nonlinear response of the material, sometimes referred to as the Kerr-effect [8,9]. $\chi^{(3)}$ here is the third-order nonlinear susceptibility of the material. Recently, theoretical and experimental studies of the $\chi^{(2)}$ response of the material showed that a large nonlinear phase shift to the wave can be obtained, where $\chi^{(2)}$ here is the second-order nonlinear susceptibility [9-12]. In the past, $\chi^{(2)}$ response was mainly used to generate a wave that oscillates with double the original frequency (Second Harmonic Generation (SHG)). Initial comparison studies between the two responses of $\chi^{(2)}$ and $\chi^{(3)}$, for GaAs material, showed that the $\chi^{(2)}$ switching device may be more efficient than the $\chi^{(3)}$ counterpart [11].

1.2 The Objective of the Thesis

All these developments in the area of integrated optics show that there is a great need for theoretical studies in order to understand and improve these devices. In addition, modelling avoids the fabrication of devices by trial and error, and therefore reduces the cost and the time of fabrication [13]. From the previous discussion, we can distinguish two classes of linear three-dimensional optical devices, the directionally non-varying waveguides (z-invariant), such as the rib waveguides, and the directionally varying waveguides (z-variant), such as the Y-junction. z-invariant waveguides can be modelled *only* numerically, because the rectangular geometry of the dielectric waveguide does not have an exact analytical solution. Therefore, efforts have been devoted to try to find the optimum solution using numerical techniques [14]¹. On the other hand, modelling z-variant waveguides is much more complicated than the previous class of waveguides, because they contain variations in the z-direction and this leads to a very involved analysis [13]. For example, the Y-junction is not a three-port device but a four-port one, where the three arms are connected with guided modes and the fourth is connected with radiation modes. With a field incident upon the junction from the combined waveguide, it is required to find the power transmitted to the two output waveguides, the reflected power in the input waveguide and the power radiated from the junction into the surrounding medium. Modelling z-variant waveguides using Coupled Mode Theory is very challenging, because

¹ A review to some of these techniques is given in chapter 2.

it is not easy to deal with the continuous mode spectrum of radiation modes. However, if the device is enclosed inside an artificial box then these radiation modes can be approximated [1]. But the size of the artificial box is a very critical parameter in solving these kinds of problems; the larger the box, the more the method converges to the correct results. In addition, inaccurate calculation of radiation modes may lead to misleading results. On the other hand, nonlinear waveguides are very difficult to model using analytically-based techniques like the Coupled Mode theory, and even more difficult when the devices have geometrical and/or material change in all three directions. Other methods, based on numerical analysis, are much better suited to such devices. In this context, numerical modelling of integrated optics devices is of great importance. In addition, the continuous rapid development of hardware helped to produce new powerful computers specifically for the purpose of numerical studies. This has also helped to produce/add new numerical techniques to improve the accuracy and the efficiency of the solution.

One of the numerical methods that can be used to study optical devices is the Finite-Difference Time-Domain technique (FDTD) [16-20]. The method was first used by Yee in 1966 [16] and later has been used by many workers to model electromagnetic wave interaction with arbitrary structures [17]. Recently, the technique was used to model two-dimensional optical devices [18-20]. For a time independent refractive index, the FDTD can be employed by discretising, using centred finite-difference approximations, the electric and the magnetic fields of the full set of Maxwell's equations and uses a time stepping mechanism, using again the centred finite-difference approximation for time derivatives, to update the discrete field values [17]. The time-stepping is continued until the sinusoidal steady state is achieved at each point. This technique is conditionally stable, where the time step size depends on the space grid size [17]. As the space grid size decreases, the time step size must also decrease for the method to remain stable. For practical applications the time step is very small, resulting in a very long simulation time. While this method combines many features of scattering problems, it requires enormous computer resources even for simple two-dimensional waveguides [19]. In addition, in many applications of cw operation, the transient response of the problem is not very important; however, the computation of this response, using the FDTD, is essential in order to find the steady state response, and that takes most of the computing time in a FDTD simulation.

Another attractive method to analyse complicated structures of integrated optics is the Beam Propagation Method (BPM). From the large volume of literature on this method, one can conclude that the method has established itself as a prime numerical tool for the

analysis of two and three-dimensional linear and nonlinear optical devices [15]². This is mainly due to its simplicity and applicability to a variety of optical devices. This method depends on the idea of marching the optical field along the direction of propagation using a one-way propagation principle. The main advantage of the BPM over the Coupled Mode Analysis is that it predicts both the guided and the radiation modes in the same formalism and its propagational ability makes it a good tool in the analysis of longitudinally varying devices. The original method was based on the FFT algorithm; however, recently new techniques to improve the accuracy and the efficiency of the method have been reported, mainly based on the Finite-Difference (FD) approximation. Broadly speaking, the FD-BPM proves to be more efficient than the classical FFT-BPM in terms of computational time and flexibility. The most popular FD-BPM is an implicit method based on the Alternating Direction Implicit (ADI) approximation. On the other hand, the explicit FD-BPMs are becoming more popular due to their simplicity and high efficiency. The explicit Real Space (RS) method, depends on splitting the finite-difference matrix operator into two matrix operators where each matrix contains small sub-matrices that are solved analytically. The Explicit Finite-Difference (EFD) technique is another attractive method for the BPM, and is a very simple technique to implement.

Generally the BPM is very efficient for the analysis of two-dimensional optical structures but consumes a lot of time when applied to three-dimensional devices, due to the following reasons³. Three-dimensional problems always lead to very large computational tasks. In addition, practical integrated optical circuits contain multiple coupled linear and non-linear waveguides which are generally longer than several hundreds of wavelengths. On top of that, the existence of large-contrast media will force the BPM to use small transverse mesh sizes and/or small longitudinal step sizes for convergence or stability reasons. When all of these are added together, the computational problem becomes very large, and conventional serial computers are not adequate for this kind of problem. Therefore, it is very natural to turn to parallel processing implementations in order to achieve practical run times for realistic devices [21-27]. On the other hand, parallel computers are built for the purpose of speeding up the execution time of large computational problems. In addition, the very intensive research activity in the area of high performance computation shows that most standard mathematical application can gain speed in the parallel environment [28-29]. However, not all serial methods are readily adapted to the parallel environment, and care needs to be exercised in selecting the

² See also the references of chapter 3.

³ As mentioned before, practical devices of integrated optics have three-dimensional geometries.

appropriate method for parallel implementation. Generally, problems that can be divided into smaller sub-problems where each is spatially localised, to minimise the communication cost, are most suitable for parallel implementation. Therefore, this work was motivated to investigate and develop efficient numerical BPMs for the analysis of three-dimensional linear and nonlinear optical waveguides using parallel-processing computers. In the nonlinear domain, waveguides that contain second order nonlinear processes of $\chi^{(2)}$ of the material will be considered, as an example of one of the demanding topics of the current nonlinear optics research investigation [9-12]. It has to be noticed that analysing waveguides containing $\chi^{(2)}$ requires slightly more than double the computation effort of the linear counterpart in which parallel computers are very useful for efficient modelling. To the best of my knowledge, the first numerical methods for the analysis of 3-D linear optical devices using large scale parallel computers were first reported from the output of this thesis [21-23]. In addition the first efficient parallel BPM for the analysis of Second Harmonic Generation in 3-D optical waveguides is also published using the results of this work [24-27].

1.3 Organisation of the Thesis

This thesis consists of three parts in addition to an introduction and a conclusion; part I is Theory, part II is Implementations (Analysis and assessments) and part III is Nonlinear optical waveguides. The theory part consist of three chapters, the implementation part consists of two chapters and the nonlinear optical waveguides part contains only one chapter. The description of these chapters is as follows:

Chapter 1 is this introductory chapter to the entire work that provides a general survey on the current issues of lightwave technology and modelling integrated optical devices and also gives the motivation of this study.

Chapter 2 is an introductory chapter to the theory of linear optical waveguides. The analysis in this chapter is restricted to z-invariant waveguides. The linear three-dimensional wave equation is derived from Maxwell's equations for general linear waveguides. The theory of guided modes of two-dimensional slab waveguides is also given including TE and TM modes. In addition, radiation modes of two-dimensional waveguides are also discussed. Then the rest of the chapter deals with modelling three-dimensional rectangular optical waveguides with a review of some of the techniques used to approximate the solution of 3-D waveguides.

Chapter 3 contains the theory of the Beam Propagation Method. The start of the chapter gives a literature review of the method, then the derivation of the parabolic equation from the Helmholtz equation and the relationships between the eigenvalues and the eigenvectors

of the two equations are shown. The formulation of the classical FFT-BPM and three different FD-BPM techniques for the analysis of 3-D waveguides are described. The chapter also shows two different techniques to compute the modes' propagation constants and their three-dimensional field distributions from the BPM field for z-invariant waveguides.

Chapter 4 describes two different advanced parallel computers; the Multiple Instruction Multiple Data (MIMD) machine and the Single Instruction Multiple Data (SIMD) machine. The MIMD is a PARSYTEC super-cluster transputer array and the SIMD is a Connection Machine (CM-200). These computers are used for the implementation of the linear and the non-linear BPM of this work. The chapter concentrates on the differences between these two parallel machines and how applications can be implemented on them. The chapter also describes two different compilers for developing parallel algorithms on these two computers.

Chapter 5 shows the formulation and the implementation of three FD-BPMs for the analysis of 2-D waveguides (slab geometry). These methods are the Crank Nicholson, the Real Space and the Explicit Finite Difference BPMs. Detailed comparisons between the accuracy and the efficiency of the three techniques are shown. A new modified EFD-BPM that improve the stability of the EFD-BPM is discussed at the end of the chapter. The chapter also contains the application of the FD-BPM to analyse practical directional couplers and Y-junctions.

Chapter 6 first shows an approximate attempt to parallelise the three-dimensional ADI-BPM on the MIMD and the SIMD machines. Then the chapter shows the detailed implementations of the two explicit BPMs (the RS and the EFD) on the transputer array and the CM-200 for the analysis of 3-D waveguides. In addition the chapter shows comparisons between the performance of these parallel methods. Then these methods have been used to analyse practical devices of 3-D nature (rib waveguides and directional couplers) where comparisons between their accuracy are made and their results have been compared with published data of serial techniques. Simulation examples for the propagation of light inside 3-D three-core directional couplers are also shown.

Chapter 7 investigates the phenomenon of SHG. In this chapter the parallel EFD-BPM is extended to study 3-D waveguides in the presence of a second order nonlinearity $\chi^{(2)}$. The chapter shows the derivation of the nonlinear coupled wave equations both in 3-D (for the SHG-EFD-BPM) and in 1-D, which is commonly used in the literature to validate experimental results and to study possible concepts for optical devices. The chapter also shows the implementation of the SHG-EFD-BPM on the two parallel machines used in the linear implementations. Then the parallel SHG-EFD-BPM is used to analyse semiconductor rib waveguides containing a second order nonlinearity. In one of the

simulation, Quasi Phase Matching (QPM) technique has been used to model the relative phase between the fundamental and the second harmonic fields. In addition the parallel SHG-EFD-BPM results have been compared with the solution of the reduced 1-D coupled wave equations where the Runge-Kutta numerical technique has been used for the solution of the 1-D model.

Chapter 8 contains a summary and a conclusion for the entire thesis work with suggestions to future work of this study.

REFERENCES

- [1] Lee, D. L., "Electromagnetic Principles of Integrated Optics", John Wiley & Sons, New York, 1986.
- [2] Yariv, A. and Yeh, P., "Optical Waves in Crystals", John Wiley & Sons, 1984.
- [3] Cherin, A. H., "An Introduction to Optical Fibers", McGraw-Hill Intr., 1983.
- [4] Marcuse, D., "Theory of Dielectric Optical Waveguides", Academic Press, San Diego, 1991.
- [5] Adams, M. J., "An Introduction to Optical Waveguides", John Wiley & Sons, 1981.
- [6] Snyder, A. W. and Love, J. D. "Optical Waveguide Theory", Chapman and Hall, London, 1983.
- [7] Tamir, T., "Guided-Wave Optoelectronics", Springer-Verlag, New York, 1990.
- [8] Stegeman, G. I. and Wright, E. M., "All-Optical Waveguide Switching", *Opt. Quant. Elec.*, **22**, 95-122, 1990.
- [9] Ironside, C. N., "Ultra-Fast All-Optical Switching", *Contem. Phys.*, **34**, 1-18, 1993.
- [10] DeSalvo, R., Hagan, D. J., Sheikh-Bahae, M., Stegeman, G. and Van Stryland, E. W., "Self-Focusing and Self-Defocusing by Cascaded Second-Order Effects in KTP", *Opt. Lett.*, **17**, 28-30, 1992.
- [11] Ironside, C. N., Aitchison, J. S. and Arnold, J. M., "An All-Optical Switch Employing the Cascaded Second-Order Nonlinearity Effect", *IEEE J.Q.E.*, **29**, 2650-2654, 1993
- [12] Hutchings, D. C., Aitchison, J. S. and Ironside, C. N., "All-Optical Switching Based on Nondegenerate Phase Shifts From a Cascaded Second-Order Nonlinearity", *Opt. Lett.*, **18**, 793-795, May 1993.
- [13] Baets, R., Kaczmarek, P. and Vankwikelberge, P., "Design and Modelling of Passive and Active Optical Waveguide Devices", Waveguide Optoelectronics (chapter 2), Edited by March, J. H. and De La Rue, R. M., Kluwer Academic Publisher, 1992.

- [14] Saad, M. S. , "Review of Numerical Methods for the Analysis of Arbitrary-Shaped Microwave and Optical Dielectric Waveguides", *IEEE Trans. MTT*, **33**, 894-899, 1985.
- [15] Yevick, D., "A Guide to Electric Field Propagation Techniques for Guided-Wave Optics", *Opt. Quant. Elect.*, **26**, 185-197, 1994.
- [16] Yee, K. S., "Numerical Solution of Initial Boundary Value Problems Involving Maxwell's Equations in Isotropic Media", *IEEE Trans. Antennas. Prop.*, **AP-14**, 302-307, 1966.
- [17] Taflove, A., "Review of the Formulation and Application of the Finite-Difference Time-Domain Method for Numerical Modelling of Electromagnetic Wave Infractions With Arbitrary Structures", *Wave Motion*, **10**, 547-582, 1988.
- [18] Chu, S. and Chaudhuri, K., "A Finite-Difference Time-Domain Method for the Design and Analysis of Guided-Wave Optical Structures", *J. Lightwave Tech.*, **7**, 2033-2038, 1989.
- [19] Haung, W. P., Chu, S., Goss, A., and Chaudhuri, K., "A Scalar Finite-Difference Time-Domain Approach to Guided-Wave Optics", *IEEE Phot. Techn. Lett.*, **3**, 524-526, 1991.
- [20] Jamid, H. A. and Al-Bader, S. J., "Finite-Difference Time-Domain Approach to Nonlinear Guided Waves", *Elect. Lett.*, **29**, 83-84, 1993.
- [21] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Methods", *IEEE Phot. Tech. Lett.*, **6**, 848-850, 1994.
- [22] Masoudi, H. M. and Arnold, J. M., "Parallel Three-Dimensional Finite-Difference Beam Propagation Methods", *Int. J. of Num. Mod.*, **8**, 95-107, 1995.
- [23] Masoudi, H. M. and Arnold, J. M., "Parallel-Processing Finite-Difference Beam Propagation Methods", Presented at the *International Symposium on Guided-Wave Optoelectronics: Device Characterization, Analysis and Design*, IEEE and Optical Society of America in Polytechnic University, New York , October 1994.
- [24] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Method For The Analysis of Second Harmonic Generation", *IEEE Photonics Technology Letters*, **7**, 400-402, April 1995.
- [25] Masoudi, H. M. and Arnold, J. M., "Modelling Second-Order Nonlinear Effects in Optical Waveguides Using a Parallel-Processing Beam Propagation Method", submitted to the *IEEE J. of Quant. Elect.*, October 1994.
- [26] Masoudi, H. M. and Arnold, J. M., "Three-Dimensional Numerical Modelling of the Second-Order Cascaded Effect in a Semiconductor Rib Waveguide Using Parallel-Processing", presented at the *Optical Society of America Topical meeting: Nonlinear Guided Waves and Their Applications*, Dana Point, California, February 1995.

- [27] Masoudi, H. M. and Arnold, J. M., "Numerical Modelling of Second Harmonic Generation in an Optical Waveguide Using Parallel-Processing Finite-Difference Beam Propagation Method", presented at the *URSI Electromagnetic Theory Symposium*, St. Petersburg, Russia, May 1995.
- [28] Modi, J. J., "Parallel Algorithms and Matrix Computation", Oxford Series Book, 1988.
- [29] Lewis, T. and El-Rewini, H., "Introduction to Parallel Computing", Printce Hall Int., New Jersey, 1992.

Part I

Theory

Chapter 2

Modelling Optical Waveguides

2.1 Introduction

In this chapter, we start with the basic equations of electromagnetic theory which are based on the solution of Maxwell's equations that describe precisely the propagation and the distribution of electromagnetic energy in dielectric waveguides. In order to model the complicated theory of nonlinear waveguides, it is essential initially to understand the analysis of the simpler linear counterparts. Therefore, this part of the thesis will deal with practical linear two and three dimensional optical waveguides. In addition, this chapter is written as an introduction to the rest of this thesis where the information of this chapter is the starting point to derivations or a reference for the later chapters. Further, we will restrict the analysis, in this chapter, to waveguides that are uniform along the direction of propagation (z-invariant).

The derivation of the linear three-dimensional vector wave equation from Maxwell's equations is given in detail in the next section, where these equations can be considered as a fundamental mathematical tool that describes the full behaviour of electromagnetic fields in dielectric media. Then, in section 2.3, the analysis of an asymmetric slab waveguide is considered, based on the solution of the simplified wave equation. The slab waveguide consists of three regions of different dielectric constants in which light can be confined in the middle layer (one dimensional confinement). Although the treatment of two-dimensional waveguides is simple, the concept of the analysis can be extended to understand the more complicated theory of three-dimensional waveguides. For practical waveguide theory, generally it is required to find all information about guided fields in the waveguide and the condition of guidance. The derivations of the field distributions and the eigenvalue equations of the TE and the TM guided modes of the slab waveguides are given in section 2.3. In addition, radiation modes of the slab waveguide are discussed, which is very important for waveguides that have discontinuities in the direction of propagation

because the total transverse field at any position of the waveguide can be written as a superposition of all orthogonal transverse modes.

Section 2.4 contains a discussion about different types of practical rib waveguides (three-dimensional waveguides) which are the skeleton of integrated optics circuits. As a closed form analytical solutions of this type of waveguide are not usually known, the section contains a review of some methods used to approximate the solution of rib waveguides, ranging from simple and fast to accurate and time consuming. The limited Effective Index Method (EIM), known for its simplicity, is discussed in section 2.4 as an extension to the solution of slab waveguides. Then the section shows some of the more rigorous analysis used, based on the solution of the wave equation in its vectorial, semi-vectorial and scalar forms, using the well-known finite-difference approximation.

2.2 The Wave Equation

We start with Maxwell's equations [1-6]. These equations are very accurate mathematical representations that describe the propagation of electromagnetic energy in materials from a macroscopic perspective. They combine relations between all electromagnetic fields and can be described in differential form as

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} \quad (2.1a)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (2.1b)$$

$$\nabla \cdot \mathbf{D} = \rho \quad (2.1c)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (2.1d)$$

where respectively, $\mathbf{E}(\mathbf{r},t)$ and $\mathbf{H}(\mathbf{r},t)$ are the electric and the magnetic field vectors, $\mathbf{D}(\mathbf{r},t)$ and $\mathbf{B}(\mathbf{r},t)$ are the electric and magnetic displacement vectors and $\mathbf{J}(\mathbf{r},t)$ and $\rho(\mathbf{r},t)$ represent the current and the charge sources. \mathbf{r} and t represent the vector space location (x,y,z) and the time variable respectively. In this thesis, we restrict the analysis for source free cases in which \mathbf{J} and ρ are set to zero. Each of the vector fields, in Eq. 2.1, contains three unknown quantities (a total of 12 unknowns). For isotropic materials the magnetic flux density vector \mathbf{B} has a direct relation with the magnetic field vector \mathbf{H} through the scalar magnetic permeability μ and can be written as

$$\mathbf{B} = \mu\mathbf{H} \quad (2.2)$$

where $\mu = \mu_0\mu_r$, μ_r being the relative magnetic permeability that can be approximated as unity for non-magnetic material (optical waveguides), and μ_0 is the free space magnetic permeability. Also for isotropic materials the electric displacement \mathbf{D} can be written in a Taylor series as a function of the electric field \mathbf{E} , as [7,8]

$$\mathbf{D}(\mathbf{r}, t) = \epsilon_0 \left[\mathbf{E}(\mathbf{r}, t) + \chi\mathbf{E}(\mathbf{r}, t) + \chi^{(2)}\mathbf{E}^2(\mathbf{r}, t) + \chi^{(3)}\mathbf{E}^3(\mathbf{r}, t) + \dots \right] \quad (2.3a)$$

or

$$\mathbf{D}(\mathbf{r}, t) = \epsilon_0(1 + \chi)\mathbf{E}(\mathbf{r}, t) + \epsilon_0\mathbf{P}(\mathbf{r}, t) \quad (2.3b)$$

where ϵ_0 is the free space permittivity, χ is the scalar linear susceptibility, $\chi^{(2)}$ and $\chi^{(3)}$ are the second and the third order nonlinear susceptibilities respectively. \mathbf{P} in Eq. 2.3b is the nonlinear polarisation vector¹. The relation in (2.3) is valid for instantaneous response of the material [7].

Now let us write the first two equations (Eqs. 2.1a and 2.1b) for the linear case after substituting \mathbf{D} and \mathbf{B} in Eqs. 2.3 and 2.2 respectively; then we arrive at

$$\nabla \times \mathbf{H} = \epsilon_0\epsilon_r \frac{\partial \mathbf{E}}{\partial t} \quad (2.4a)$$

$$\nabla \times \mathbf{E} = -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \quad (2.4b)$$

where $\epsilon_r(x,y,z)$ is the relative permittivity equals to $(1+\chi)$. As a common practice, we will derive an equation that contains only the electric field \mathbf{E} by eliminating the magnetic field \mathbf{H} . A similar equation could be derived for the magnetic field using the same approach below. If we take the curl ($\nabla \times$) of Eq. 2.4b as

$$\begin{aligned} \nabla \times \nabla \times \mathbf{E} &= \nabla \times \left\{ -\mu_0 \frac{\partial \mathbf{H}}{\partial t} \right\} \\ &= -\mu_0 \frac{\partial}{\partial t} \{ \nabla \times \mathbf{H} \} \end{aligned} \quad (2.5)$$

and using the following vector identity for the double curl in Eq. 2.5 acting on the vector \mathbf{E} as

¹Later in this thesis, we will consider only $\chi^{(2)}$ response of the material.

$$\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} \quad (2.6)$$

and substituting Eq. 2.4a in (2.5), then we can rewrite (2.5) in terms of the electric field only as

$$\begin{aligned} \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E} &= -\mu_0 \frac{\partial}{\partial t} \left\{ \epsilon_0 \epsilon_r \frac{\partial \mathbf{E}}{\partial t} \right\} \\ &= -\mu_0 \epsilon_0 \epsilon_r \frac{\partial^2 \mathbf{E}}{\partial t^2} \end{aligned} \quad (2.7)$$

Now we can use Eq. 2.1c for a source free field ($\rho = 0$), which can be written in this form

$$\nabla \cdot \mathbf{D} = \nabla \cdot (\epsilon_r \mathbf{E}) = \mathbf{E} \cdot \nabla \epsilon_r + \epsilon_r \nabla \cdot \mathbf{E} = 0$$

or

$$\nabla \cdot \mathbf{E} = -\mathbf{E} \cdot \frac{\nabla \epsilon_r}{\epsilon_r} \quad (2.8)$$

substituting (2.8) in (2.7) we get the following equation

$$\nabla^2 \mathbf{E} + \nabla \left(\mathbf{E} \cdot \frac{\nabla \epsilon_r}{\epsilon_r} \right) = \mu_0 \epsilon_0 \epsilon_r \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (2.9)$$

For a monochromatic field that oscillates with an angular frequency ω , the electric and the magnetic fields can be written as²

$$\begin{bmatrix} \mathbf{E}(\mathbf{r}, t) \\ \mathbf{H}(\mathbf{r}, t) \end{bmatrix} = \begin{bmatrix} \mathbf{E}(\mathbf{r}) \\ \mathbf{H}(\mathbf{r}) \end{bmatrix} e^{j\omega t} \quad (2.10)$$

Then the second derivative with respect to t in Eq. 2.9 can be replaced with $-\omega^2$, leading to the following vectorial wave equation where, as throughout this work, the time dependence term ($e^{j\omega t}$) will be removed from equations as a matter of convenience

$$\nabla^2 \mathbf{E} + \nabla \left(\mathbf{E} \cdot \frac{\nabla n^2}{n^2} \right) + k_0^2 n^2 \mathbf{E} = 0 \quad (2.11)$$

² $j = \sqrt{-1}$.

where

$$n = \sqrt{\epsilon_r}$$

$$k_0 = \frac{\omega}{c} = \omega \sqrt{\mu_0 \epsilon_0} = \frac{2\pi}{\lambda} \quad ,$$

$n(x,y,z)$ is the refractive index, c is the speed of light in vacuum and λ is the wavelength of the wave. Eq. 2.11 is a very complicated and involved equation as the three components of the electric field are coupled through the second term which can be expanded in Cartesian co-ordinates³ as

$$\begin{aligned} \nabla \left(\mathbf{E} \cdot \frac{\nabla n^2}{n^2} \right) &= \mathbf{a}_x \frac{\partial}{\partial x} \left[\frac{1}{n^2} E_x \frac{\partial n^2}{\partial x} + \frac{1}{n^2} E_y \frac{\partial n^2}{\partial y} + \frac{1}{n^2} E_z \frac{\partial n^2}{\partial z} \right] \\ &+ \mathbf{a}_y \frac{\partial}{\partial y} \left[\frac{1}{n^2} E_x \frac{\partial n^2}{\partial x} + \frac{1}{n^2} E_y \frac{\partial n^2}{\partial y} + \frac{1}{n^2} E_z \frac{\partial n^2}{\partial z} \right] \\ &+ \mathbf{a}_z \frac{\partial}{\partial z} \left[\frac{1}{n^2} E_x \frac{\partial n^2}{\partial x} + \frac{1}{n^2} E_y \frac{\partial n^2}{\partial y} + \frac{1}{n^2} E_z \frac{\partial n^2}{\partial z} \right] \end{aligned} \quad (2.12)$$

where \mathbf{a}_x , \mathbf{a}_y and \mathbf{a}_z are unit vectors in the x , y and z directions respectively. In order to find a solution for Eq. 2.11, all three components must be solved simultaneously. For a homogenous medium this term will be zero, and the dropping of this term simplifies the equation because the three components of the electric field are decoupled and the following Helmholtz equation

$$\nabla^2 \mathbf{E} + k_0^2 n^2 \mathbf{E} = 0 \quad (2.13)$$

results. The equation could be written for any of the electric field components separately as

$$\left. \begin{aligned} \nabla^2 E_x + k_0^2 n^2 E_x \\ \nabla^2 E_y + k_0^2 n^2 E_y \\ \nabla^2 E_z + k_0^2 n^2 E_z \end{aligned} \right\} = 0 \quad (2.14)$$

A similar approach to the derivation of Eq. 2.13 will lead to a similar equation for the magnetic field \mathbf{H} as

³ The Cartesian co-ordinate system has been used because it is very convenient for the analysis of most integrated optics devices.

$$\nabla^2 \mathbf{H} + k_0^2 n^2 \mathbf{H} = 0 \quad (2.15)$$

or

$$\left. \begin{aligned} \nabla^2 H_x + k_0^2 n^2 H_x \\ \nabla^2 H_y + k_0^2 n^2 H_y \\ \nabla^2 H_z + k_0^2 n^2 H_z \end{aligned} \right\} = 0 \quad (2.16)$$

For dielectric waveguides which consist of different homogenous media, the fields at the interface between two layers must obey Boundary Condition (BC) requirements. In other words, the solution of a certain problem that consists of two (or more) media could be solved locally by means of Maxwell's equations (or the derived wave equation) and then coupled together using BC. BC equations can be derived from the integral form of Maxwell's equations [1]. Let us consider two dielectric media with different refractive indices n_1 and n_2 . All the fields at the interface between the two media must be linked through

$$\mathbf{a}_{\text{norm}} \times (\mathbf{E}_2 - \mathbf{E}_1) = 0 \quad \Rightarrow \mathbf{E}_2^{\text{tang}} = \mathbf{E}_1^{\text{tang}} \quad (2.17a)$$

$$\mathbf{a}_{\text{norm}} \times (\mathbf{H}_2 - \mathbf{H}_1) = 0 \quad \Rightarrow \mathbf{H}_2^{\text{tang}} = \mathbf{H}_1^{\text{tang}} \quad (2.17b)$$

$$\mathbf{a}_{\text{norm}} \cdot (\mathbf{B}_2 - \mathbf{B}_1) = 0 \quad \Rightarrow \mathbf{B}_2^{\text{norm}} = \mathbf{B}_1^{\text{norm}} \quad \Rightarrow \mathbf{H}_2^{\text{norm}} = \mathbf{H}_1^{\text{norm}} \quad (2.17c)$$

$$\mathbf{a}_{\text{norm}} \cdot (\mathbf{D}_2 - \mathbf{D}_1) = 0 \quad \Rightarrow \mathbf{D}_2^{\text{norm}} = \mathbf{D}_1^{\text{norm}} \quad \Rightarrow n_2^2 \mathbf{E}_2^{\text{norm}} = n_1^2 \mathbf{E}_1^{\text{norm}} \quad (2.17d)$$

where the subscripts 1 and 2 refer to medium n_1 and n_2 respectively. \mathbf{a}_{norm} is a unit vector normal to the interface between the two media, "norm" and "tang" refer to the normal and the tangential components of the fields. From the above boundary conditions, we may observe that the tangential components of \mathbf{E} and \mathbf{H} and the normal of \mathbf{H} must be continuous across the interface of the two dielectric media. On the other hand, the normal of \mathbf{D} is continuous at the boundary but the normal of \mathbf{E} is discontinuous.

2.3 The Dielectric Slab Waveguide

In this section we will be dealing with the analytical solution of one-dimensional confinement of light using three layers (slab waveguide). For simplicity, we will consider materials which are linear, lossless and piece-wise homogeneous. Although the slab waveguide mathematical treatment is simple and physically easy, the analytical solution of this waveguide helps to understand other more complicated structures (e.g. three-

dimensional and/or nonlinear, etc.). In addition, the slab waveguide theory could be easily extended to multi-layer waveguides solutions using the same concept of this section [2]. When we say that we are interested to find the analytical solutions of a waveguide, this means the physical distribution of the electromagnetic fields inside and outside the waveguide, which must satisfy Maxwell's equations. In other words, it is required to find the modes that can exist in the waveguide. A mode is defined here as any transverse field distribution whose magnitude is independent of position along the direction of propagation.

Fig. 2.1 shows a dielectric slab waveguide consisting of three homogeneous and lossless layers: the cladding, the guiding and the substrate, where respectively the refractive indices of these layers are n_c , n_g and n_s .

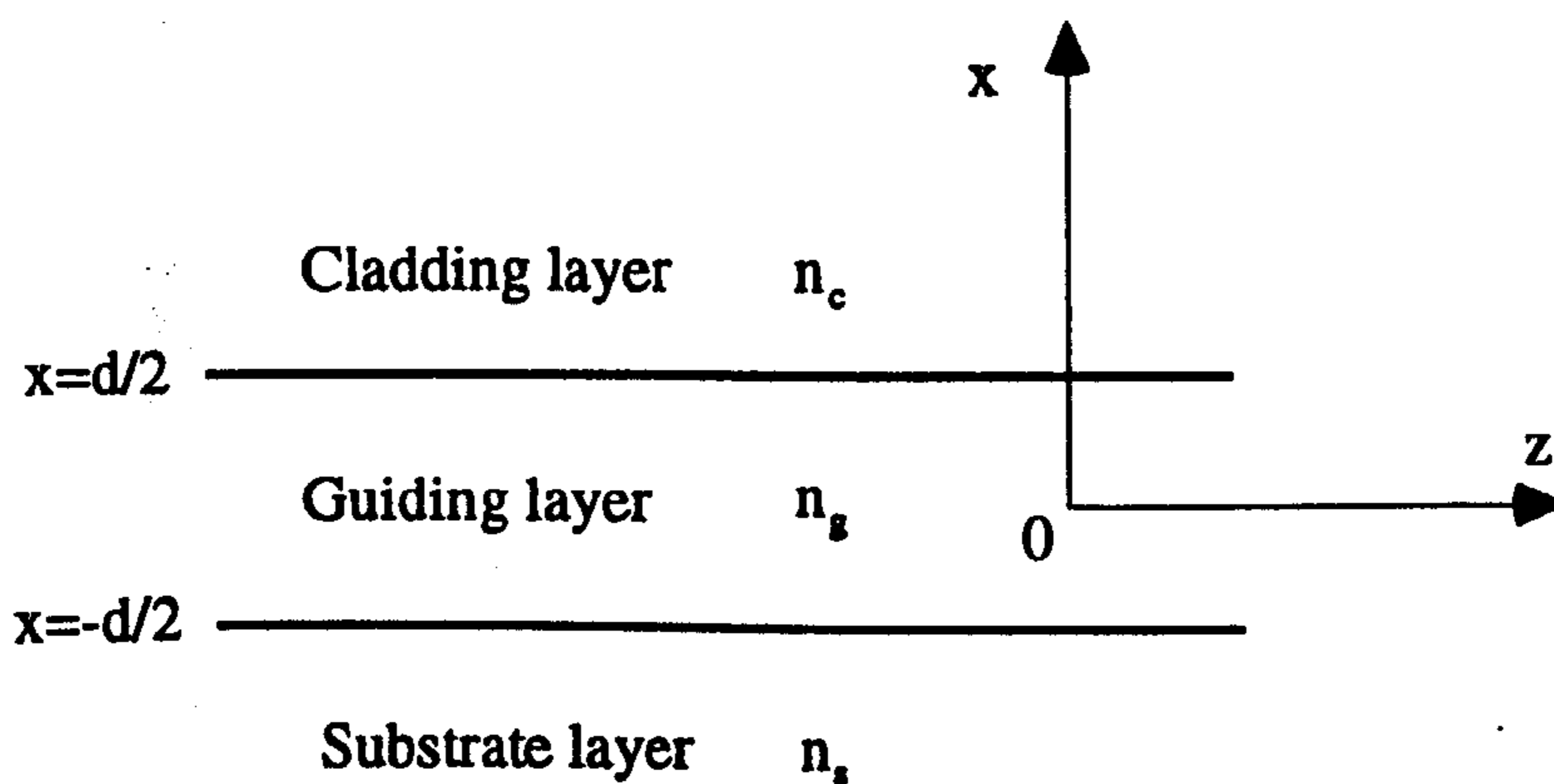


Fig. 2.1 The dielectric slab waveguide.

It has been assumed that z is the direction of propagation where the light does not encounter any changes. The guiding layer where the guided (or trapped) light exist is surrounded by two semi-infinite layers (the guide extends to infinity in the positive and negative y -direction) with refractive indices lower than the value of the refractive index of the guiding layer. Typical values of the refractive indices, for semiconductor materials (GaAs), are $n_c = 1$ (air), $n_g = 3.44$ and $n_s = 3.4$ and a thickness d of the order of $1 \mu\text{m}$. Since we are concerned with only two-dimensional geometry, all derivatives with respect to y in the wave equation can be set to zero $\left(\frac{\partial}{\partial y} = 0, \text{ and } \frac{\partial^2}{\partial y^2} = 0\right)$. Then the wave equation in (2.13) can be written, for each region of Fig. 2.1, as [2]

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} + \frac{\partial^2 \mathbf{E}}{\partial z^2} + k_0^2 n_r^2 \mathbf{E} = 0 \quad (2.18)$$

where n_r denotes the refractive index of one of the three layers. Further we may assume that the variation of the fields in the positive z -direction as $e^{-jk_z z}$ where k_z is the longitudinal propagation constant. We may rewrite Eq. 2.18 for the three regions separately as

$$\frac{\partial^2 \mathbf{E}_c}{\partial x^2} - r_c^2 \mathbf{E}_c = 0 \quad (\text{The cladding region}) \quad (2.19a)$$

$$\frac{\partial^2 \mathbf{E}_g}{\partial x^2} + r_g^2 \mathbf{E}_g = 0 \quad (\text{The guiding region}) \quad (2.19b)$$

$$\frac{\partial^2 \mathbf{E}_s}{\partial x^2} - r_s^2 \mathbf{E}_s = 0 \quad (\text{The substrate region}) \quad (2.19c)$$

where

$$\begin{aligned} r_c^2 &= k_z^2 - k_0^2 n_c^2 \\ r_g^2 &= k_0^2 n_g^2 - k_z^2 \\ r_s^2 &= k_z^2 - k_0^2 n_s^2 \end{aligned}$$

Note that in Eq. 2.19, the term $e^{j(\omega t - k_z z)}$ has been dropped from the equation as a matter of convenience. At this stage, we have to distinguish between the two polarisations that can be supported in slab geometry; the Transverse Electric (TE) field and the Transverse Magnetic (TM) field.

2.3.1 Transverse Electric (TE) Guided Modes

In TE waves, E_y , H_x , and H_z are the only non zero field components, and we may write Eq. 2.19 in terms of E_y only; from Maxwell equations (Eq. 2.4b), we can obtain H_x , and H_z from the value of E_y as

$$H_x = -\frac{k_z}{\omega \mu_0} E_y \quad (2.20a)$$

$$H_z = \frac{j}{\omega \mu_0} \frac{\partial E_y}{\partial x} \quad (2.20b)$$

It is important to note, from the above, that the electric field is always tangential to the interfaces between the dielectric layers of the slab waveguide. Under guiding condition it is required to have most of the power confined to the central region of the slab waveguide

(guided region of Fig. 2.1). This implies that oscillatory behaviour in the middle layer and exponential decay (evanescent) behaviour in the outer regions are required to satisfy this requirement, and the solution of Eq. 2.19. This requirement can be enforced on the coefficients of Eq. 2.19 as r_c^2, r_g^2 and $r_s^2 \geq 0$ and combined to give the following inequalities⁴

$$k_o n_g \geq k_z \geq k_o n_s \geq k_o n_c \quad (2.21)$$

Using the above, we may write the guided mode solutions for the asymmetric slab waveguide for the electric field E_y as [2]

$$E_y = \begin{cases} \left[A \cos(r_g d/2) + B \sin(r_g d/2) \right] e^{-r_c(x-d/2)} & x \geq d/2 \\ A \cos(r_g x) + B \sin(r_g x) & -d/2 \leq x \leq d/2 \\ \left[A \cos(r_g d/2) - B \sin(r_g d/2) \right] e^{r_s(x+d/2)} & x \leq -d/2 \end{cases} \quad (2.22)$$

Eq. 2.22 has been written so that the boundary condition requirements for the continuity of the electric field E_y at the both interfaces of the slab waveguide are satisfied. If E_y is continuous, then from Eq. 2.20a H_x is continuous as well. The other requirement for boundary conditions to satisfy is the continuity of H_z at $x = d/2$ and $x = -d/2$. From Eq. 2.20b H_z can be written from E_y of (2.22) as

$$H_z = \frac{j}{\omega \mu_o} \begin{cases} -r_c \left[A \cos(r_g d/2) + B \sin(r_g d/2) \right] e^{-r_c(x-d/2)} & x \geq d/2 \\ r_g \left[-A \sin(r_g x) + B \cos(r_g x) \right] & -d/2 \leq x \leq d/2 \\ r_s \left[A \cos(r_g d/2) - B \sin(r_g d/2) \right] e^{r_s(x+d/2)} & x \leq -d/2 \end{cases} \quad (2.23)$$

The continuity of H_z at $x = d/2$ gives

$$-r_c \left[A \cos(r_g d/2) + B \sin(r_g d/2) \right] = r_g \left[-A \sin(r_g d/2) + B \cos(r_g d/2) \right]$$

⁴ Note that it has been assumed that $n_g > n_c$.

and at $x = -d/2$ gives

$$r_g \left[A \sin(r_g d/2) + B \cos(r_g d/2) \right] = r_s \left[A \cos(r_g d/2) - B \sin(r_g d/2) \right]$$

From the above two relations, one can eliminate the ratio A/B and with some tedious but straight forward algebra we arrive at the following relation

$$\tan(dr_g) = \frac{r_g(r_s + r_c)}{r_g^2 - r_s r_c} \quad (\text{TE}) \quad (2.24)$$

Eq. 2.24 is the eigenvalue (characteristic) equation for the TE guided modes where the only unknown quantity is the propagation constant k_z (k_z is embedded in r_g , r_s , and r_c ; see below Eq. 2.19).

2.3.2 Transverse Magnetic (TM) Guided Modes

Using a similar approach to the derivation of the TE guided modes, we can obtain the guided modes for the TM polarisation. For this polarisation the only non-zero field components are H_y , E_x , and E_z . The wave equation in (2.19) can be written in terms of H_y only, and we may also assume that the fields vary as $e^{-jk_z z}$ in the z -direction. From Maxwell's equation (Eq. 2.4a), we can obtain E_x , and E_z from the value of H_y using

$$E_x = \frac{k_z}{\omega \epsilon_0 n_r^2} H_y \quad (2.25a)$$

$$E_z = \frac{-j}{\omega \epsilon_0 n_r^2} \frac{\partial H_y}{\partial x} \quad (2.25b)$$

where n_r refers to the refractive index of the layers n_c , n_g and n_s . The H_y solution for the TM guided modes can be written similarly to the E_y of the TE in (2.22) as [2]

$$H_y = \begin{cases} \left[C \cos(r_g d/2) + D \sin(r_g d/2) \right] e^{-r_c(x-d/2)} & x \geq d/2 \\ C \cos(r_g x) + D \sin(r_g x) & -d/2 \leq x \leq d/2 \\ \left[C \cos(r_g d/2) - D \sin(r_g d/2) \right] e^{r_s(x+d/2)} & x \leq -d/2 \end{cases} \quad (2.26)$$

Eq. (2.26) has been written so that H_y is continuous across the two interfaces of the slab waveguide (at $x = d/2$ and $-d/2$). From Eq. 2.25b it follows that E_z could be written as

$$E_z = \frac{-j}{\omega\epsilon_0} \begin{cases} -\frac{\Gamma_c}{n_c^2} [C \cos(r_g d/2) + D \sin(r_g d/2)] e^{-r_c(x-d/2)} & x \geq d/2 \\ \frac{\Gamma_g}{n_g^2} [-C \sin(r_g x) + D \cos(r_g x)] & -d/2 \leq x \leq d/2 \\ \frac{\Gamma_s}{n_s^2} [C \cos(r_g d/2) - D \sin(r_g d/2)] e^{r_s(x+d/2)} & x \leq -d/2 \end{cases} \quad (2.27)$$

The continuity of E_z in Eq. 2.27 at $x = d/2$ gives

$$-\frac{\Gamma_c}{n_c^2} [C \cos(r_g d/2) + D \sin(r_g d/2)] = \frac{\Gamma_g}{n_g^2} [-C \sin(r_g d/2) + D \cos(r_g d/2)]$$

and at $x = -d/2$ gives

$$\frac{\Gamma_g}{n_g^2} [-C \sin(r_g d/2) + D \cos(r_g d/2)] = \frac{\Gamma_s}{n_s^2} [C \cos(r_g d/2) - D \sin(r_g d/2)]$$

Eliminating the ratio C/D from the above two equations and with some algebraic manipulations we arrive at

$$\tan(dr_g) = \frac{\Gamma_g n_g^2 (n_c^2 r_s + n_s^2 r_c)}{n_s^2 n_c^2 r_g^2 - n_g^4 r_s r_c} \quad (\text{TM}) \quad (2.28)$$

Eq. 2.28 is the eigenvalue (characteristic) equation for the TM guided modes where again the only unknown quantity is the propagation constant k_z .

2.3.3 Guided Mode Computations

For guided modes k_z 's are discrete values which can be computed using Eqs. 2.24 and 2.28. It is a usual practice to normalise the propagation constant k_z in terms of the free space propagation constant k_0 , as

$$k_{z,N} = n_{\text{eff},N} k_0, \quad (N = 0, 1, 2, \dots) \quad (2.29)$$

where $n_{\text{eff},N}$'s are the discrete normalised effective indices of guided modes. N is added to the definition of effective index just to refer to the mode number where 0 is the highest value of effective index (sometimes referred to as the fundamental mode or the 0th order mode). N also refers to the number of nodes in the transverse field distribution. From Eq. 2.21, the discrete values of the effective indices of guided modes that satisfy the eigenvalue equations should be

$$n_g > n_{\text{eff},N} > n_s \quad (2.30)$$

It is important to note that a waveguide supports a finite number of guided modes where the cut-off condition is always given by $n_{\text{eff},c} = n_s$ or $k_{z,c} = k_o n_s$.

The eigenvalue equations (2.24 for TE and 2.28 for TM) could be solved graphically to find the exact values of the propagation constants and the total number of guided modes [4,5]. From such a plot, the intersections between the left side of these equations (\tan function curves) and the curves of the right hand side of the equations indicate the discrete values of the propagation constants where the number of intersections is the total number of guided modes. On the other hand, simple numerical methods could be used to solve the eigenvalue equations (e.g. a zero finding routines like Muller's method [9]). In these methods a guess near the eigenvalue is needed as an input and the routine improves this value iteratively toward the true eigenvalue, within an achievable limit, until it converges. From the eigenvalue equations of both the TE and the TM we may write the following equation to find the total number of guided modes for the TE or the TM [2,5]

$$\text{Number of modes} = \left\{ \frac{1}{\pi} \left[v - \tan^{-1} \left\{ \eta \left(\frac{n_g^2 - n_c^2}{n_g^2 - n_s^2} \right)^{1/2} \right\} \right] \right\}_{\text{int}} \quad (2.31)$$

where the subscript "int" refers to the next largest integer of the curly bracket and v is the normalised frequency defined as

$$v = dk_o (n_g^2 - n_s^2)^{1/2} \quad (2.32)$$

and the parameter η in Eq. 2.31 is defined as

$$\eta = \begin{cases} 1, & \text{for TE modes} \\ n_g^2/n_c^2, & \text{for TM modes} \end{cases}$$

From the above equations of both the TE and the TM, we may conclude that there are three factors that determine the total number of guided modes supported by a slab waveguide. The first is the difference between the values of the refractive indices of the three layers where, for example, increasing the value of the refractive index of the guiding region will increase the total number of guided modes. The second factor is the thickness of the slab waveguide d where the total number of guided modes increases by increasing d . And the third is the wavelength of the field (or the frequency) where increasing the wavelength will decrease the total number of guided modes.

2.3.4 Guided Modes Power Flow

Once the propagation constants of guided modes are known, then we can compute the time-average Poynting power P carried by guided modes along the direction of propagation z . However, the slab waveguide considered in our analysis extends to infinity in the y -direction, so in this case we assume P as power per unit length (in the y -direction) with a unit W/m . The guided power flow P in the z direction is given by the integral over the waveguide cross section of the z -component of the Poynting vector S_z [2]

$$P = \int_{-\infty}^{\infty} S_z dx = \frac{1}{2} \int_{-\infty}^{\infty} \text{Re}(\mathbf{E} \times \mathbf{H}^*)_z dx \quad (2.33)$$

where P is a real and a positive quantity and the asterisk $*$ refers to the complex conjugate. For TE modes, the electric field \mathbf{E} is given by (2.22) (only the y -component of \mathbf{E} exists) and the magnetic field can be computed from E_y using (2.20a). Then (2.33), can be written as

$$P = -\frac{1}{2} \text{Re} \int_{-\infty}^{\infty} E_y H_x^* dx = \frac{k_z}{2\omega\mu_0} \int_{-\infty}^{\infty} |E_y|^2 dx \quad (\text{TE}) \quad (2.34a)$$

In a similar manner, the guided power flow in the z -direction for the TM modes can be computed using H_y in (2.26) and E_x in (2.25a)

$$P = \frac{1}{2} \text{Re} \int_{-\infty}^{\infty} H_y^* E_x dx = \frac{k_z}{2\omega\epsilon_0} \int_{-\infty}^{\infty} (1/n^2(x)) |H_y|^2 dx \quad (\text{TM}) \quad (2.34b)$$

$n(x)$ in (2.34) is a reminder that the refractive index is different in the three regions of the slab waveguide. It is convenient to normalise the field distribution to the power carried by the mode. First let us write the transverse fields as

$$E_y(x) = a\mathcal{E}_y(x) \quad (\text{For TE}) \quad (2.35a)$$

$$H_y(x) = a\mathcal{H}_y(x) \quad (\text{For TM}) \quad (2.35b)$$

where the fields \mathcal{E}_y and \mathcal{H}_y are the transverse fields, normalised so that the mode is carrying a power of 1W/m and are defined as [3]

$$\mathcal{E}_y(x) = \frac{E_y(x)}{\sqrt{\frac{1}{2} \left| \int_{-\infty}^{\infty} E_y(x) H_x^*(x) dx \right|}} \quad (\text{TE}) \quad (2.36a)$$

and

$$\mathcal{H}_y(x) = \frac{H_y(x)}{\sqrt{\frac{1}{2} \left| \int_{-\infty}^{\infty} H_y^*(x) E_x(x) dx \right|}} \quad (\text{TM}) \quad (2.36b)$$

and a 's are amplitude coefficients. Then we may write the power in (2.34) as

$$P = \frac{k_z}{2\omega\mu_0} \int_{-\infty}^{\infty} |E_y|^2 dx = \frac{k_z}{2\omega\mu_0} \int_{-\infty}^{\infty} |a|^2 |\mathcal{E}_y|^2 dx = |a|^2 \quad (\text{For TE}) \quad (2.37a)$$

$$P = \frac{k_z}{2\omega\epsilon_0} \int_{-\infty}^{\infty} (1/n^2(x)) |H_y|^2 dx = \frac{k_z}{2\omega\epsilon_0} \int_{-\infty}^{\infty} (1/n^2(x)) |a|^2 |\mathcal{H}_y|^2 dx = |a|^2 \quad (\text{For TM}) \quad (2.37b)$$

2.3.5 Radiation Modes

The number of guided modes of dielectric waveguides, as described previously, is always finite, but in order to provide a complete set of orthogonal modes, there must exist other modes, which also must satisfy both the solution of Maxwell's equations and the boundary conditions [2,4]. These modes do exist and are called radiation modes [2,4]. Unlike the guided modes, which have evanescent solutions outside the guiding region, radiation fields could satisfy the wave equation and the boundary conditions by having sinusoidal solutions outside the guiding region which extend to infinity. In addition, the time and the z variations for these fields, can be also described by the factor $e^{j(\omega t - k_z z)}$. Thus, we may say that these fields qualify with all the characteristics to be called modes, except that their energy is not confined to the guiding region but reaches to infinite

distance in the x-directions. The other main difference between radiation modes and guided modes is that the propagation constants of radiation modes are not discrete but form a continuum. At this stage we must distinguish between two types of radiation modes which may describe the slab waveguide solution (both as TE or TM); namely the substrate radiation modes and the cladding (cover) radiation modes. The first could be described by sinusoidal solutions in the guided and the substrate regions and evanescent waves in the cladding region while the second type has sinusoidal variations in all three regions. Fig. 2.2 shows a schematic diagram for the location of the effective indices of all possible modes in the asymmetric slab waveguide of Fig. 2.1. From the figure, we see that radiation modes' propagation constants are below the cut-off condition of guided modes (see section 2.3.3).

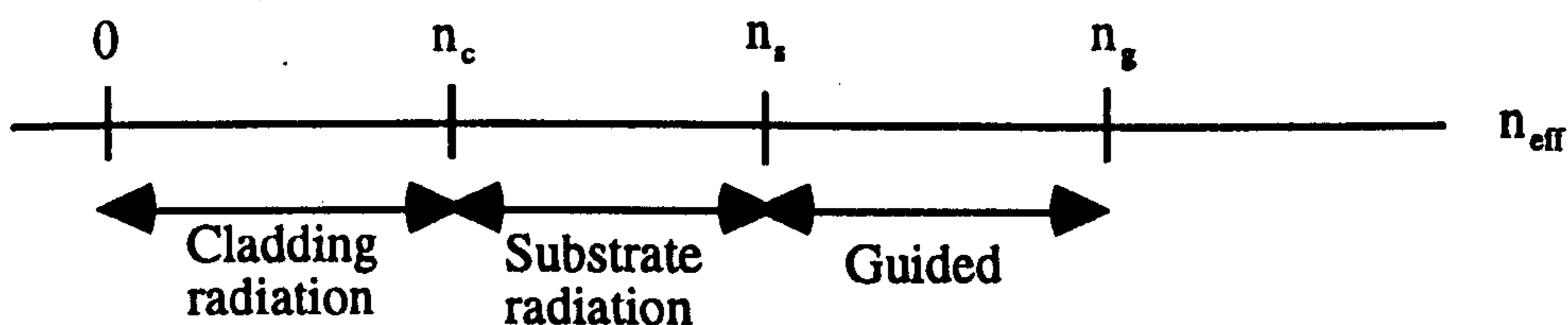


Fig. 2.2 The location of the effective indices of all possible modes in asymmetric slab waveguide.

Radiation modes are necessary to model dielectric optical devices, because they are part of the description of the total field solution of dielectric waveguides. In practice, the study of z-variant waveguides (waveguides that have discontinuities in the refractive index along the direction of propagation z) requires the solution of both guided and radiation modes. For example, guided modes travelling inside z-variant waveguides will excite an infinite number of radiation modes. This means that discontinuities in waveguides are the cause of transferring energy from guided to radiation modes and consequently the energy will travel away from the guided region toward the infinite space around the waveguide. For z-invariant waveguides, radiation modes will exist if the waveguide is excited with an initial input field distribution (e.g. at $z = 0$) different from those of the guided fields (e.g. Gaussian field).

Orthogonality of modes

One of the important relations between modes (guided and radiation of TE and TM) of waveguides is that all modes are orthogonal. This relation can be expressed as [3,6]

$$\int_A \mathbf{E}_i^v \times \mathbf{H}_i^r \cdot d\mathbf{A} = 0, \quad k_z^v \neq k_z^r \quad (2.38)$$

where t indicates the transverse field distribution, A is the transverse cross sectional area of the waveguide, and ν and τ refer to two different modes whose propagation constants are different. This equation (2.38) is written in such way that it can be used for three dimensional waveguides where A could contain x and y dimensions. For TE modes of the slab waveguide of section 2.3.1, the orthogonality relation can be simplified into

$$\int_{-\infty}^{\infty} \frac{-k_z^\nu}{\omega\mu_0} E_y^\nu E_y^{\tau*} dx = 0, \quad k_z^\nu \neq k_z^\tau \quad (2.39)$$

and for the TM of section 2.3.2, the relation is

$$\int_{-\infty}^{\infty} \frac{-k_z^\nu}{\omega\epsilon_0 n^2} H_y^\nu H_y^{\tau*} dx = 0, \quad k_z^\nu \neq k_z^\tau \quad (2.40)$$

In addition to the orthogonality relation of modes, both guided and radiation modes form a complete set. This feature is also important which indicates that any arbitrary transverse field at a constant z can be written as a superposition of guided and radiation modes as [5]

$$E_y(x) = \sum_N^{NGM} a_N \mathcal{E}_N(x) + \int_0^{k_0 n_1} b(k_z) E(x, k_z) dk_z \quad (\text{TE}) \quad (2.41a)$$

$$H_y(x) = \sum_N^{NGM} a_N \mathcal{H}_N(x) + \int_0^{k_0 n_1} b(k_z) H(x, k_z) dk_z \quad (\text{TM}) \quad (2.41b)$$

In Eq. 2.41 the discrete summation is over the guided modes and the continuous summation (integral) is over radiation modes. NGM denotes the number of guided modes, a 's are the amplitude coefficient of the guided modes and b 's the amplitude coefficient of radiation modes. We have to note that radiation modes can not be normalised to a finite power because the distribution of the fields extend to infinity in the x -directions of the slab waveguide.

Due to the necessity of radiation modes in the analysis of optical waveguides, some of these modes can be computed *approximately* using the idea of metallic waveguide theory, also the method helps to understand the physical behaviour of radiation modes in dielectric waveguides [4,5]. If we enclose the slab waveguide of Fig.2.1 with two metal walls (with a conductivity $\sigma = \infty$) at $x = d/2+L$ and $-d/2-L$ (see Fig. 2.3), then the problem is transformed into a metallic waveguide with an insert of dielectric waveguide in the

middle. In other words, the propagation constants k_z of radiation modes are now discrete rather than continuum. The summation integral of Eq. 2.41 is transformed into a discrete summation over the new discrete radiation modes. The slab waveguide geometry is obtained in the limit of allowing L to approach infinity. As L increases, the spacing between the discrete values of radiation modes propagation constants decreases and in the limit of infinite L they merge into a continuum.

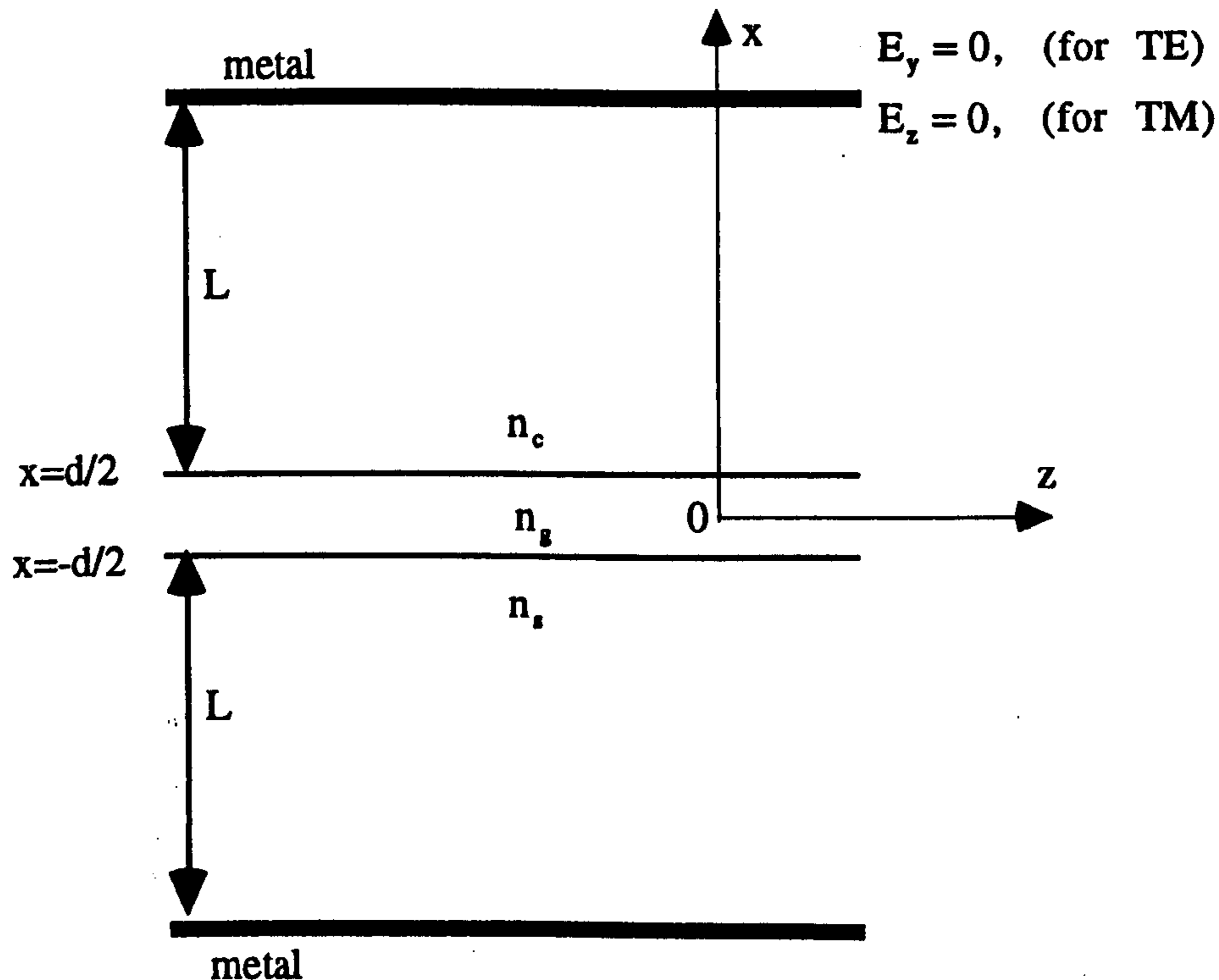


Fig.2.3 Two metal walls surrounds the slab waveguide of Fig. 2.1 to approximate the analysis of radiation modes.

2.4 Modelling Three-Dimensional Waveguides

Rectangular (rib) dielectric waveguides proved to be the practical way to control light for the use in optical processing circuits. One of the purposes of rib waveguides is to confine energy in both transverse directions x (horizontally) and y (vertically), in the y -direction by material changes and in the x -direction by geometrical changes. Although more complicated waveguides are used in many practical applications, the rib waveguide is considered to be the essential element in such devices. Fig. 2.4 shows a three-dimensional view of a typical dielectric rib waveguide. Similar to the operation of the slab waveguide for guiding light, the refractive index of the middle layer (the guided layer) n_s must be greater than the refractive indices of the outer regions n_c and n_s [2,4]. Practical devices are

always surrounded by air, where in such a case n_c can be set to the refractive index of air which is unity.

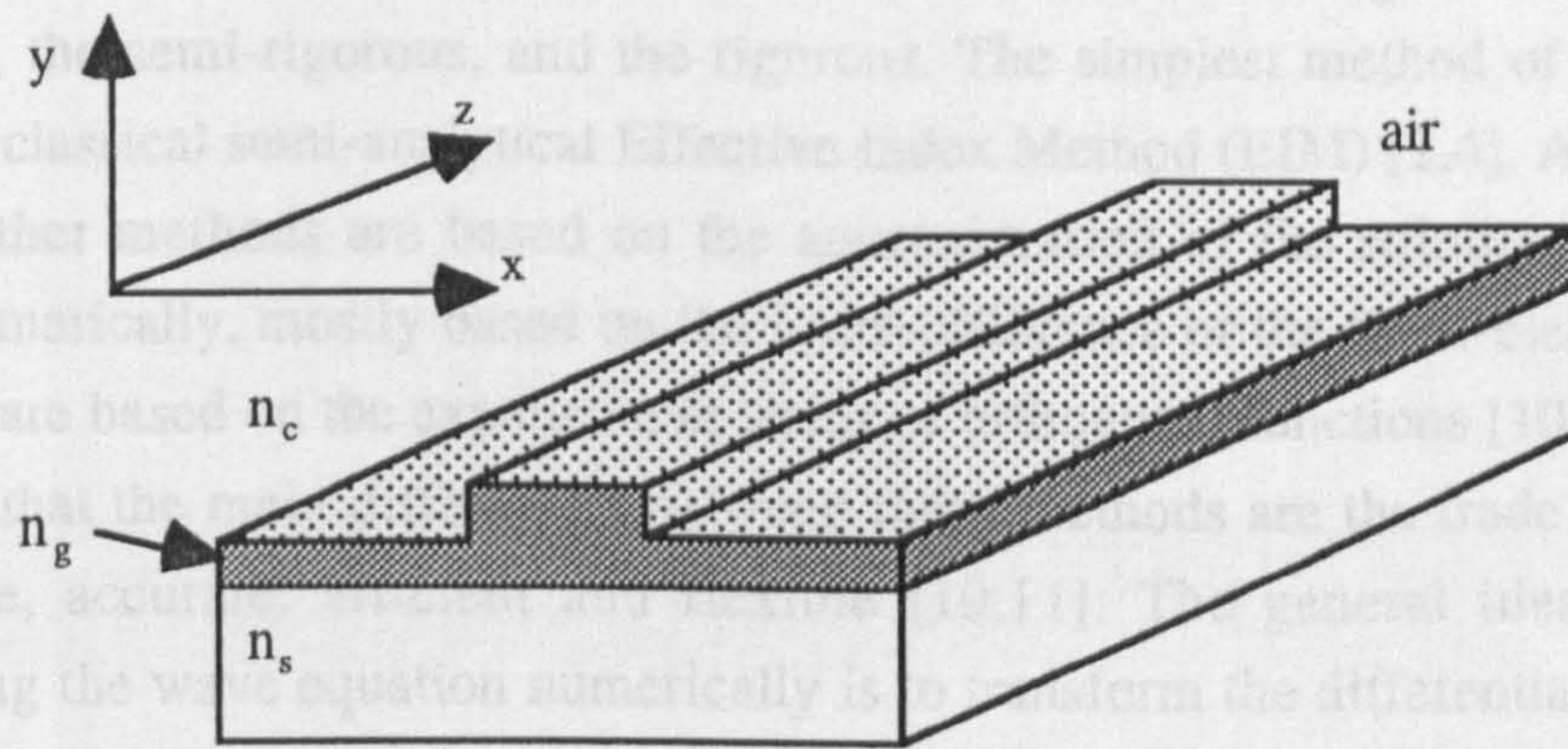


Fig. 2.4 Three dimensional view of a dielectric rib waveguide ($n_g > n_s > n_c$).

In this section we will be dealing with piece wise homogenous z-invariant rib waveguides. Z-invariant three dimensional waveguides are sometimes referred to as two dimensional variation waveguides. Fig 2.5 shows a cross sectional view of different common types of three-dimensional waveguides that exist in practical optical processing [2,4,5].

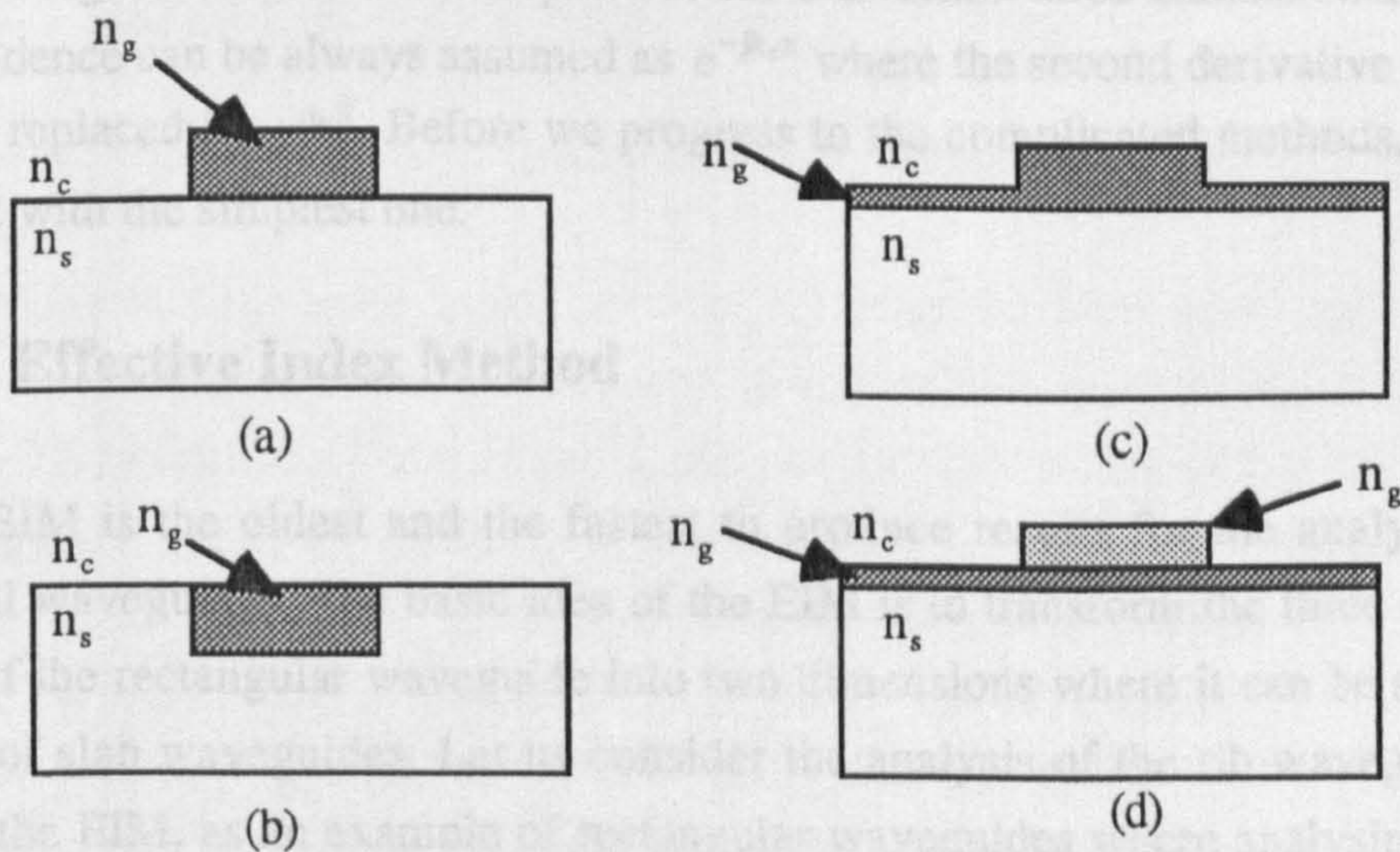


Fig. 2.5 Cross sectional view of different types of three dimensional waveguides geometries. (a) Strip waveguide, $n_g > n_s > n_c$. (b) Embedded strip waveguide, $n_g > n_s > n_c$ (c) Rib or ridge waveguide, $n_g > n_s > n_c$. (d) Strip loaded waveguide, $n_g > n_{gl} > n_s > n_c$.

The shape of three-dimensional optical waveguides can be classified into two types, the circular geometry of optical fibers and the rectangular geometry of rib waveguides (Figs. 2.4 and 2.5). The circular dielectric geometry can be solved analytically using Bessel functions [2,3], but modelling rectangular rib waveguides analytically is not possible

because there is no analytical function that can fit the rectangular refractive index distribution. Due to the great interest in modelling optical rib waveguides, there is an enormous number of techniques which can be divided into four categories, the simple, the approximate, the semi-rigorous, and the rigorous. The simplest method of all is the one based on the classical semi-analytical Effective Index Method (EIM) [2,4]. Apart from this technique, other methods are based on the approximation of the solution of Maxwell's equations numerically, mostly based on the finite-difference or the finite-element methods while others are based on the expansion in terms of orthogonal functions [10-25]. It can be summarised that the main differences between these methods are the trade off factors of being simple, accurate, efficient and flexible [10,11]. The general idea involved in approximating the wave equation numerically is to transform the differential equations of the wave equation into a matrix eigenvalue problem which can be solved using standard matrix manipulations [10]. The finite-difference approximation method is the most common technique used for a long time for the solution of microwave problems and other applications. In addition to its simplicity, it has some advantages over the finite-element approach of being free from producing spurious numerical fields and computationally more efficient [10,22]. In the following discussion we will review briefly some of the finite-difference techniques that have been used for the analysis of three-dimensional optical dielectric waveguides. As a common practice for z-invariant three dimensional waveguides the z-dependence can be always assumed as $e^{-jk_z z}$ where the second derivative with respect to z can be replaced by $-k_z^2$. Before we progress to the complicated methods, it is a good idea to start with the simplest one.

2.4.1 The Effective Index Method

The EIM is the oldest and the fastest to produce results for the analysis of three dimensional waveguides. The basic idea of the EIM is to transform the three dimensional geometry of the rectangular waveguide into two dimensions where it can be solved using the theory of slab waveguides. Let us consider the analysis of the rib waveguide in Fig. 2.5c using the EIM, as an example of rectangular waveguides where analysing the others follow similar approaches. Assume that the rib waveguide can be divided, in the x-direction, into three independent regions as in Fig. 2.6.

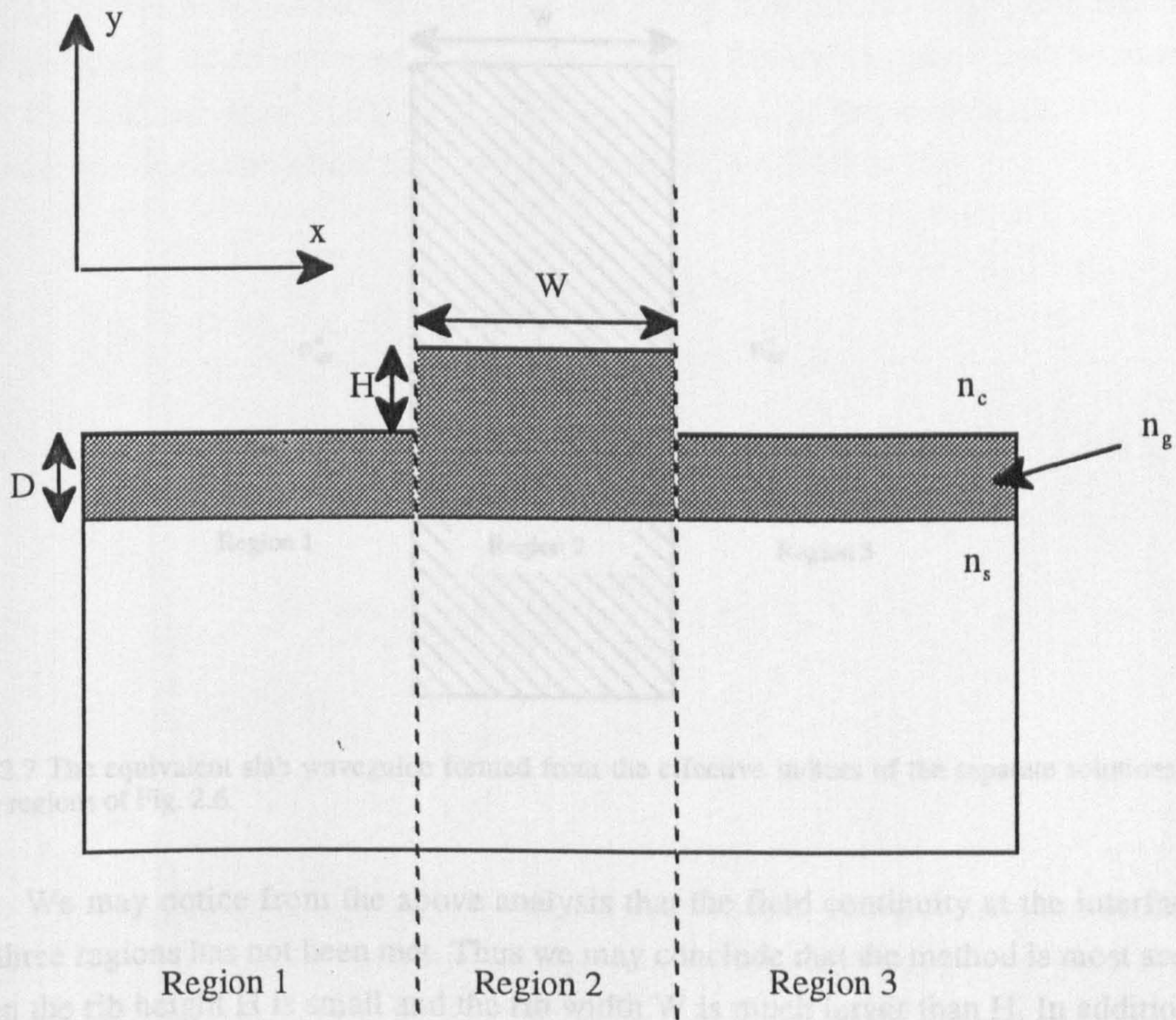


Fig. 2.6 The division of the rib waveguide into three slab waveguides using the EIM approach.

Also we assume that the three regions of Fig. 2.6 of three independent slab waveguides where each region can be solved using the theory of slab waveguide of the previous section. The outer regions (1 and 3) are identical slab waveguides with a width of D and the inner region is a slab waveguide with a width $D+H$. The EIM can be viewed from the scalar wave equation using the separation of variables (x and y) approach, where either polarisation TE or TM can be assumed. Let us call the effective indices of solving the two slab waveguides, the outer and the inner, of Fig. 2.6 as n_{eff}^o and n_{eff}^i respectively. The three regions can be considered now as a slab waveguide in the y -direction with a thickness W (see Fig. 2.7). This slab waveguide can be solved by considering the effective indices n_{eff}^o and n_{eff}^i as its refractive indices. The effective index resulting from solving the slab waveguide of Fig. 2.7 (in the y -direction) is the effective index of the rib waveguide based on the EIM approximation.

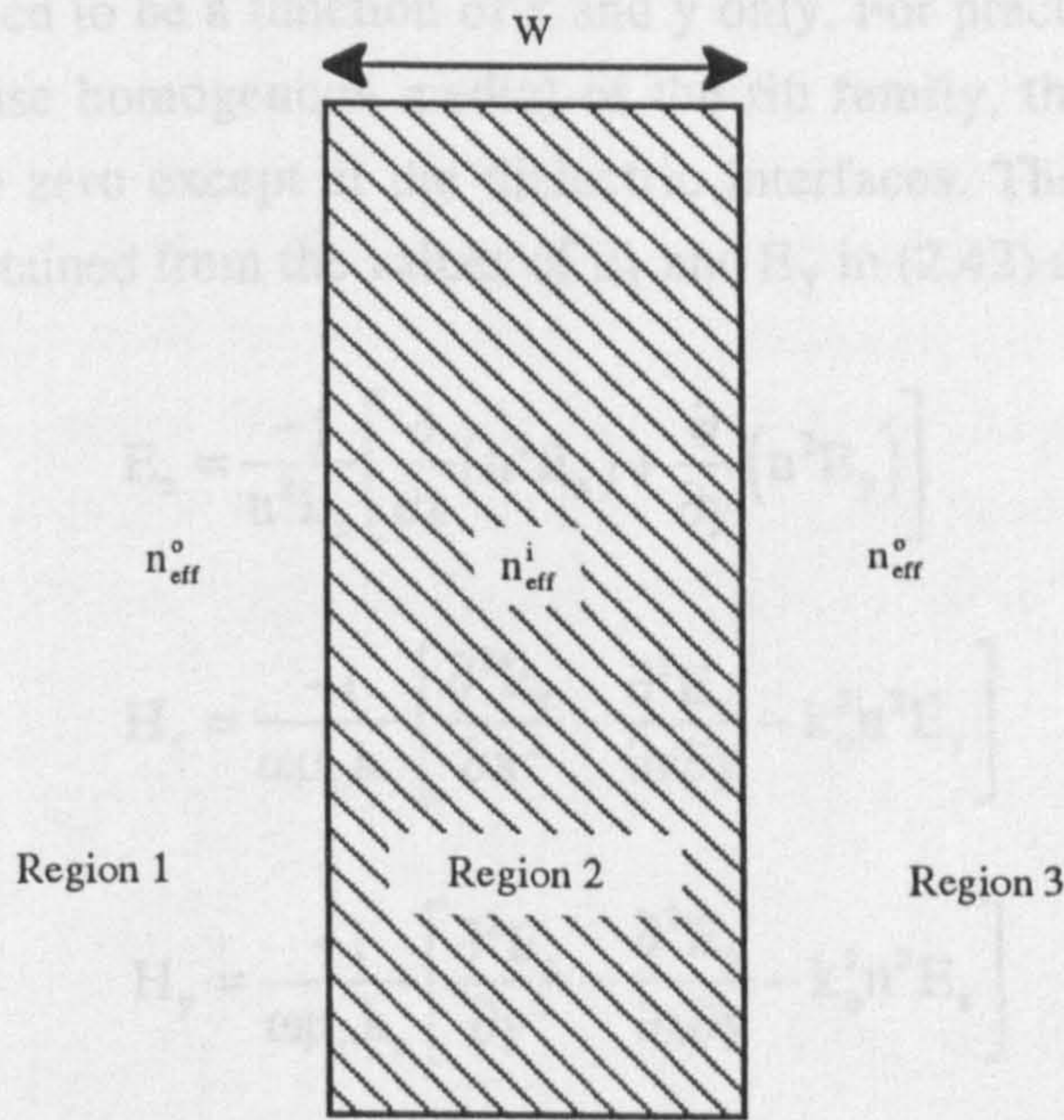


Fig. 2.7 The equivalent slab waveguide formed from the effective indices of the separate solutions of the three regions of Fig. 2.6.

We may notice from the above analysis that the field continuity at the interfaces of the three regions has not been met. Thus we may conclude that the method is most accurate when the rib height H is small and the rib width W is much larger than H . In addition, the method will break down if the outer regions of the rib waveguide do not support a mode. However, to get out of the break down situation one can replace the outer n_{eff}^o with the cladding refractive index n_c [19].

2.4.2 Solutions to the wave equation

We start with the rigorous approaches which are based on the numerical solution to the vector wave equation in Eq. 2.11. This equation contains three unknown fields (E_x, E_y, E_z) and can be written in terms of the following coupled equations that depend on only x and the y components of the vector field \mathbf{E} [12,15]

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial}{\partial x} \left(\frac{1}{n^2} E_x \frac{\partial n^2}{\partial x} + \frac{1}{n^2} E_y \frac{\partial n^2}{\partial y} \right) + (k_0^2 n^2 - k_z^2) E_x = 0 \quad (2.42a)$$

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial y^2} + \frac{\partial}{\partial y} \left(\frac{1}{n^2} E_x \frac{\partial n^2}{\partial x} + \frac{1}{n^2} E_y \frac{\partial n^2}{\partial y} \right) + (k_0^2 n^2 - k_z^2) E_y = 0 \quad (2.42b)$$

where $n(x,y)$ is assumed to be a function of x and y only. For practical three-dimensional waveguides (piece wise homogenous media) of the rib family, the third and the fourth terms of Eq. 2.42 are zero except at the dielectric interfaces. The remaining four field components can be obtained from the values of E_x and E_y in (2.42) as [12]

$$E_z = \frac{-j}{n^2 k_z} \left[\frac{\partial}{\partial x} (n^2 E_x) + \frac{\partial}{\partial y} (n^2 E_y) \right] \quad (2.43a)$$

$$H_x = \frac{-1}{\omega \mu_0 k_z} \left[\frac{\partial^2 E_y}{\partial x^2} - \frac{\partial^2 E_x}{\partial x \partial y} - k_0^2 n^2 E_y \right] \quad (2.43b)$$

$$H_y = \frac{-1}{\omega \mu_0 k_z} \left[\frac{\partial^2 E_x}{\partial y^2} - \frac{\partial^2 E_y}{\partial x \partial y} - k_0^2 n^2 E_x \right] \quad (2.43c)$$

$$H_z = \frac{j}{\omega \mu_0} \left[\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} \right] \quad (2.43d)$$

The coupled equations 2.42a and 2.42b could be solved numerically using finite difference approximations [15]. In order to discretise the optical fields (E_x and E_y) to get a finite number of equations, the optical device has to be limited in both directions (x and y), which could be done by placing the device inside an artificial box. The box should be large enough to contain all significant field points. This will lead to a finite number of field mesh points which correspond to the field equations. These equations can be written in the following matrix eigenvalue equation

$$A_v \begin{bmatrix} E_x \\ E_y \end{bmatrix} = k_z^2 \begin{bmatrix} E_x \\ E_y \end{bmatrix} \quad (2.44)$$

where A_v is a large real sparse matrix of a dimension of $2M$ by $2M$, where M is the total number of mesh points of the fields E_x or E_y . The subscript v of A_v denotes that the matrix belongs to the vector approach. A similar approach in terms of H_x and H_y fields could be found in [16].

The solution of the discretised vector wave equation in (2.44) will lead, eventually, to two groups of field polarisations, the quasi TE and the quasi TM. The quasi TE is considered when the E_x component is much larger than the E_y component and the quasi TM is considered when the E_y component is much larger than the E_x component. Under

the situation that one of the field component is much larger than the second component, then it can be assumed that the second one is zero which leads to a much simplified equation. This kind of assumption is called the semi-vectorial approximation which is based on the usual laboratory definition of horizontal and vertical polarisations of the optical field [17]. The semi-vectorial quasi-TE (horizontal) field assumes that the y field component of the electric field is zero ($E_x, 0, E_z$) while the semi-vectorial quasi-TM (vertical) field assumes that the x field component of the electric field is zero ($0, E_y, E_z$). This will lead to the decoupling of the two equations in (2.42) as

$$\frac{\partial^2 E_x}{\partial x^2} + \frac{\partial^2 E_x}{\partial y^2} + \frac{\partial}{\partial x} \left(\frac{1}{n^2} E_x \frac{\partial n^2}{\partial x} \right) + (k_o^2 n^2 - k_z^2) E_x = 0 \quad (\text{For quasi TE}) \quad (2.45a)$$

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial^2 E_y}{\partial y^2} + \frac{\partial}{\partial y} \left(\frac{1}{n^2} E_y \frac{\partial n^2}{\partial y} \right) + (k_o^2 n^2 - k_z^2) E_y = 0 \quad (\text{For quasi TM}) \quad (2.45b)$$

which can be solved independently. Again the finite difference approximations can be employed in the decoupled equations above, which will lead to the following eigenvalue equations for the quasi TE and the quasi TM [17]

$$A_{TE} E_x = k_z^2 E_x \quad (\text{For quasi TE}) \quad (2.46a)$$

$$A_{TM} E_y = k_z^2 E_y \quad (\text{For quasi TM}) \quad (2.46b)$$

where A_{TE} and A_{TM} are real non-symmetric band matrices. On the other hand, the semi-vectorial approach could be simplified further using the scalar approximation. This approximation is used when there is no interest in distinguishing explicitly between different polarisations. In addition, if the discontinuity of the refractive index at the interfaces between layers is small or the field amplitude is small where there is a large change in the refractive index, the third term of Eqs. 2.45a and 2.45b could be ignored leading to the scalar wave equation [10,17,19,22,23]

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + (k_o^2 n^2 - k_z^2) E = 0 \quad (2.47)$$

where E represent any component of the vector electric field \mathbf{E} . The application of the finite difference approximation to the scalar wave equation in (2.47) will give the following simple eigenvalue equation [17]

$$A_s E = k_z^2 E \quad (2.48)$$

where A_s is real symmetric band matrix.

2.4.3 Solving the matrix eigenvalue equation numerically

The question of solving the eigenvalue problem in (2.44), (2.46) or (2.48) numerically is an important issue because, this will reflect seriously in the overall accuracy and efficiency of the algorithm. We have to remember that the requirements for solving these eigenvalue equations are to find the eigenvalues (k_z) and the eigenvectors (the two dimensional field distributions). Generally, the methods of solving an eigenvalue equation numerically can be classified as matrix methods and iterative methods [9,26]. The matrix approach has the disadvantage of requiring to store or sometimes invert a large matrix. The iterative methods such as the relaxation method demand less storage facility and do not require an inversion of the matrix. The operation of the relaxation method is very simple; it determines the solution by starting with a guess for the eigenvector and improving upon it iteratively until convergence is reached, where the eigenvalue k_z also has to be guessed at the beginning and updated after a few iterations using, for example, the Rayleigh quotient equation [21]. The solution, using the relaxation method, converges to the dominant eigenmode, and it can be modified to compute higher order modes [15]. However, in order to find the m th mode all $m-1$ lower order modes have to be computed and stored in advance. Eventually, we expect that as the mode order increases, the accuracy decreases due to accumulating errors, and more storage memory is needed. On the other hand, the matrix methods like the Shifted Inverse Power Iteration Method (SIPIIM) for example [9,17,26] coupled with LU factorisation of the eigenvalue matrix, can be employed to solve for any eigenvalue (not necessarily the dominant one) and its corresponding eigenvector. The search for a particular eigenvalue and eigenvector using the SIPIIM is independent of other modes; a guess for the eigenvector and a guess near the eigenvalue must be provided to start the algorithm. Apparently, the accuracy from the SIPIIM for higher order modes is expected to be very high compared with the relaxation method since each mode is computed independently.

REFERENCES

- [1] Cherin, A. H., "An Introduction to Optical Fibers", McGraw-Hill Intr., 1983.
- [2] Adams, M. J. , "An Introduction to Optical Waveguides", John Wiley & Sons, 1981.

- [3] Snyder, A. W. and Love, J. D. "Optical Waveguide Theory", Chapman and Hall, London, 1983.
- [4] Lee, D. L., "Electromagnetic Principles of Integrated Optics", John Wiley & Sons, New York, 1986.
- [5] Marcuse, D., "Theory of Dielectric Optical Waveguides", Academic Press, San Diego, 1991.
- [6] Tamir, T., "Guided-Wave Optoelectronics", Springer-Verlag, New York, 1990.
- [7] Yariv, A. and Yeh, P., "Optical Waves in Crystals", John Wiley & Sons, New York, 1984.
- [8] Weitzman, P. S. and Osterberg, U., "A Modified Beam Propagation Method to Model Second Harmonic Generation in Optical Fibers", *IEEE J.Q.E.*, **29**, 1437-1443, 1993.
- [9] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. "The Art of scientific computing: numerical recipes in FORTRAN", Cambridge Uni. Press, 1992.
- [10] Saad, M. S. , "Review of Numerical Methods for the Analysis of Arbitrary-Shaped Microwave and Optical Dielectric Waveguides", *IEEE Trans. MTT*, **33**, 894-899, 1985.
- [11] Working Group I, Cost 216, "Comparison of Different Modelling Techniques for Longitudinally Invariant Integrated Optical Waveguides", *IEE Proc. Pt. J*, **136**, 273-280, 1989.
- [12] Marcuse, D. , "Solution of the Vector Wave Equation for General Dielectric Waveguides by the Galerkin Method", *IEEE J. Q. E.* , **28**, 459-465, 1992
- [13] Rahman, B. M. A., and Davies, J. B. , "Finite-Element Analysis of Optical and Microwave Waveguide Problems", *IEEE Trans. MTT*, **32**, 20-28, 1984.
- [14] Rahman, B. M. A., and Davies, J. B., "Vector-H Finite Element Solution of GaAs/GaAlAs Rib Waveguides", *IEE Proc. Pt. J*, **132**, 349-353, 1985.
- [15] Tayler, M. PhD. Thesis to be submitted to the Electronics and Electrical Eng. Dep. at the University of Glasgow, (under preparation), 1994.
- [16] Schultz, N., Bierwirth, K., Arndt, F., and Koster, U. , "Rigorous Finite-Difference Analysis of Coupled Channel Waveguides with Arbitrarily Varying Index Profile", *J. Lightwave Tech.*, **9**, 1244-1253, 1991.
- [17] Stern, M. S. , "Semivectorial Polarised Finite Difference Method for Optical Waveguides With Arbitrary Index Profiles", *IEE Proc. Pt. J.*, **135**, 56-63, 1988.
- [18] Vassallo, C. and Wang, Y. H. , "A New Semirigorous Analysis of Rib Waveguides", *J. Lightwave Tech.* , **8**, 56-65, 1990.

- [19] Robertson, M. J., Ritchie, S., and Dayan, P. , "Semiconductor Waveguides: Analysis of Optical Propagation in Single Rib Structures and Dielectric Couplers", *IEE Proc. Pt. J.*, **132**, 336-342, 1985.
- [20] Kendall, P. C., Adams, M. J., Ritchie, S., Robertson, M. J., " Theory for Calculating Approximate Values for the Propagation Constants of an Optical Rib Waveguide by Weighting the Refractive Indices", *IEE Proc. Pt. A*, **134**, 699-702, 1987.
- [21] Benson, T. M., Kendall, P. C., Stern, M. S., and Quinney, D. A. , "New Results for Rib Waveguide Propagation Constants", *IEE Proc. Pt. J.*, **136**, 97-102, 1989.
- [22] Schweig, E., and Bridges, W. , "Computer Analysis of Dielectric Waveguides: A Finite-Difference Method", *IEEE Tran. MTT*, **32**, 531-541, 1984.
- [23] Henry, C. H. and Verbeek, B. H. , "Solution of the Scalar Wave Equation for Arbitrarily Shaped Dielectric Waveguides by Two-Dimensional Fourier Analysis", *J. Lightwave Tech.*, **7**, 308-313, 1989.
- [24] Mittra, R. , Hou, Y. L. and Jamnejad, V., "Analysis of Open Dielectric Waveguides Using Mode-Matching Technique and Variational Methods", *IEEE Tran. MTT*, **28**, 36-43, 1980.
- [25] Bierwirth, K. Schultz, N., and Arndt, F. , "Finite-Difference Analysis of Rectangular Dielectric Waveguide Structures", *IEEE Tran. MTT*, **34**, 1104-1113, 1986.
- [26] Griffiths, D.V., and Smith, I.M. , "Numerical Methods for Engineers", Blackwell Scientific Publications, Oxford, 1991.

Chapter 3

The Beam Propagation Method

3.1 Introduction

The Beam Propagation Method (BPM) is a numerical technique invented in the late 1970's by Feit and Fleck [1-7], to approximate the solution of the wave equation when applied to optical waveguides. The method is very suitable for both classes of longitudinally varying and non-varying dielectric devices. The basic idea of the BPM is to march any arbitrary transverse field distribution, inside an optical device, from one position to another, in the propagation direction over a small distance, where this process could be repeated several times to cover the entire length of the optical device. In mathematical terms this reads as writing a relationship between two transverse fields spaced by a small longitudinal distance using an approximate numerical formula for the wave equation where one of the field is always known (the input). The BPM is attractive to the designer of optical devices, because it overcomes the difficulties of mode theory when applied to complicated structures and because of its flexibility as a propagational technique. The main advantage of the BPM over the traditional mode approach is that it computes both guided and radiation modes of linear z-variant devices, in the same formalism; no special arrangement is needed to account for radiation modes. In addition, optical devices containing nonlinear effects are very difficult to analyse using the mode theory. On the other hand, it is very simple for the BPM to include the nonlinear part in the formulation using little adjustment to the linear version. Generally, the application of the BPM to model optical devices proves to be very useful in terms of simplicity and accuracy [1-39].

The classical BPM has been used widely, in the past decade or so, to analyse many optical waveguides structures [1-10]. It is mainly an approximation to the scalar wave equation in its parabolic form using the Fast Fourier Transform (thus called FFT-BPM). In the FFT-BPM, the discretised optical field is simulated by a spectrum of plane waves, for

the transverse variations, in the spectral domain and a phase correction due to the inhomogeneity of the media is introduced in the spatial domain. The role of the FFT is to provide a transformation between the spectral domain and the spatial domain, where the optical field has to be transformed several times between the two domains for each propagational step. Although the computation of the transverse variations using the FFT is highly accurate, nevertheless the use of the FFT restricts the overall efficiency of the algorithm. In addition to the poor efficiency of the FFT-BPM, a large variation in the transverse refractive index profile of optical waveguides will force the method to use extremely small propagational steps, resulting in very large computational time. Also, the restriction imposed by the FFT makes the method inflexible of using nonuniform grid spacing and radiation boundary conditions [10].

The original FFT-BPM was based on a split-operator, which is an approximation to the scalar paraxial wave equation for small contrast media and only as a forward-going technique [1-7]. The scalar approximation of the method neglects the difference between different polarisations (see chapter 2), the paraxial approximation restricts the method for propagation with small angles with respect to the axial direction, and the one-way propagation does not account for any backward scattered reflections. However, since the first BPM until now there have been many techniques reported in the literature to improve the efficiency and the accuracy of the classical BPM. Generally, it has been found that using other approaches than the FFT approximation will lead to a much more efficient and flexible method [10-23]. In addition, much work has been done to include other features in the BPM such as wide-angle propagation [24-27], backward reflections [28-31], vectorial formulation [32-39] and boundary conditions [40-41]. The wide-angle BPM uses a Padé approximant technique that overcomes the paraxial limit and accounts for the neglected second derivative with respect to the direction of propagation of the field [24-27]. For large longitudinal index discontinuity, the bi-directional BPM in [28], for example, could be employed to account for backward reflections. In this method, the field is propagated forward and then backward while reflections are computed and stored, and this process is repeated several times, to update the forward and backward reflections, until a stable solution is obtained. The BPM techniques in [32-38] solves the three-dimensional vector wave equation which contains two coupled paraxial equations, to account for the two field components (x and y), where they are solved simultaneously. Also the three-dimensional semi-vectorial BPM based on the paraxial approximation has been reported in [39]. It uses the assumption that the two field components could be decoupled, resulting in two separate paraxial wave equations, one for each polarisation, that could be solved independently. The decoupling of the two equations reduces the computation time substantially compared to

the vector counterpart. In addition, the work in [27] has combined the wide angle feature and the vector feature into one BPM. Finally, a technique for the BPM's boundary conditions to account for fields that reach the edges of the computational window has been implemented using a so-called transparent boundary condition (TBC) [40-41]. This technique is particularly useful in modelling waveguides that contain significant scattered radiation (e.g. Y-junction). If such scattered radiation is not properly absorbed at the edges of the computational window, it tends to reflect back to the region of interest causing unwanted interference.

On the other hand, improvements of the efficiency and flexibility of the classical BPM were mainly concentrated on the replacement of the FFT. An alternative numerical technique to the FFT is to solve the BPM wave equation in the spatial domain using Finite-Difference (FD) approximations to replace the partial derivatives in the wave equation (thus called FD-BPM) [10-22]. The prime advantage of this process is to avoid the excessive time loss in the multiple operation of changing from spatial domain to spectral domain for each propagation step. Finite-Elements (FE) also could be used to approximate the BPM wave equation in the spatial domain, however FD-BPM has the advantage of being simpler to program and more efficient than the FE-BPM [14]. FD-BPM proved to be much more efficient and more flexible than the FFT-BPM [10-22], and some algorithms are very well suited to parallel computers [45-46]. In terms of efficiency, for example, the FFT-BPM requires, per propagation step, $O(M \log M)^1$ operations per transverse dimension whereas the corresponding count for the FD methods is $O(M)$ [17]. In addition, the FD technique allows for nonuniform discretisations of the transverse fields, and radiation boundary conditions can be used at the edges of the computational window [14,22,40,41]. Recently, the FD-BPM approach has received wide attention from many workers [10-22,32-39,42-44] using different techniques. Strictly speaking, FD-BPMs have two classes of expressing the operator formulation, the implicit approach and the explicit approach [47,48]. The implicit methods are based on the split-operator using the idea of the Alternating Direction Implicit (ADI) approximation (sometimes referred to as Crank Nicholson), are the most popular [10-16,19-22], are unconditionally stable, but require inversion of a large system of matrix equations for each propagational step. On the other hand, the explicit approach has mainly two ways of formulating the problem. The first is by using the same operators of the ADI and the FFT and is called the Real Space method (RS-BPM) [42-44] which is based on the idea of splitting the finite-difference matrix operator into two other matrix operators. The splitting is chosen such that each matrix is

¹ M is the total number of discretisations in one of the transverse direction (x or y).

block-diagonal where each block contains a small sub-matrix which may be exponentiated analytically [42]. The RS-BPM is unconditionally stable and proves to be much more efficient than the ADI because it does not involve solving a system of equations but multiplication of small independent matrices [42]. The second approach is truly explicit, and thus called the Explicit Finite-Difference method (EFD-BPM) [17,18] which is based on applying the finite-difference approximation directly to the wave equation. This leads to an algorithm that involves multiplication of the initial optical field with a very sparse matrix. However, the EFD-BPM is only conditionally stable, but proves to be very efficient indeed. It is important to note that a stable algorithm does not necessarily produce correct results. All the three finite-difference methods mentioned before are discussed later in this chapter.

The main emphasis of this chapter is to show the theory of the BPM and to give a full understanding of the differences between the different techniques of the method. Some of the numerical details are not shown here but they will be given in the following chapters when the methods are implemented. In this chapter, the derivation of the parabolic equation from the scalar Helmholtz equation is given directly in the next section where the relationships between the eigenvalues and the eigenvectors of the two equations are shown. These relationships are very useful, since they enable the characteristic solution of the Helmholtz to be determined from the parabolic characteristic solutions. All the formulations in this chapter will be derived for three-dimensional structures where the two-dimensional version (slab waveguides) could be extracted easily, for most cases, from the 3-D equations by setting off one of the transverse dimension. In section 3.3, the formulation and the derivation of the FFT-BPM and the three different FD-BPMs mentioned before are shown. Then in section 3.4, two different techniques to compute the modes' propagation constants and their three-dimensional field distributions from the BPM field for z-invariant waveguides are shown in details.

3.2 The Parabolic Equation

The scalar wave equation in (2.13) contains second derivatives with respect to the propagation direction z which takes care of propagation in both directions of z . It is very convenient to transform the scalar wave equation into an approximate one way equation. Let us expand the Helmholtz equation (2.13) into Cartesian co-ordinates as

$$\frac{\partial^2 E}{\partial x^2} + \frac{\partial^2 E}{\partial y^2} + \frac{\partial^2 E}{\partial z^2} + k_0^2 n^2(x, y, z) E = 0 \quad (3.1)$$

We extract a rapidly changing phase factor in the direction of propagation z and write the field $E(x, y, z)$ in Eq. 3.1 as

$$E(x, y, z) = \Phi(x, y, z) e^{-jk_0 n_0 z}, \quad (3.2)$$

where n_0 is a reference refractive index and k_0 is the vacuum wave number. It is to be remembered for later discussion that $\Phi(x, y, z)$ also represents the solution of the Helmholtz equation without the reference phase. Substituting (3.2) in (3.1) will give the following equation

$$\begin{aligned} & \frac{\partial^2 \Phi}{\partial x^2} e^{-jk_0 n_0 z} + \frac{\partial^2 \Phi}{\partial y^2} e^{-jk_0 n_0 z} + \frac{\partial^2 \Phi}{\partial z^2} e^{-jk_0 n_0 z} \\ & - 2jk_0 n_0 \frac{\partial \Phi}{\partial z} e^{-jk_0 n_0 z} - k_0^2 n_0^2 \Phi e^{-jk_0 n_0 z} + k_0^2 n^2 \Phi e^{-jk_0 n_0 z} = 0 \end{aligned} \quad (3.3)$$

Now we may rewrite Eq. 3.3 under the parabolic approximation (sometimes referred to as the slowly varying envelope approximation) which assumes that the change of field Φ in the direction of propagation is very small over a wavelength. This assumption mathematically reads as

$$\left| \frac{\partial^2 \Phi}{\partial z^2} \right| \ll \left| 2k_0 n_0 \frac{\partial \Phi}{\partial z} \right| \quad (3.4)$$

Choosing the reference refractive index n_0 as a representative index value from the propagation effective indices of the modes minimises the error caused by the parabolic approximation [14]. Then Eq. 3.4 can be simplified into the following parabolic or Fresnel wave equation

$$2jk_0 n_0 \frac{\partial \phi}{\partial z} = \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + k_0^2 (n^2 - n_0^2) \phi \quad (3.5)$$

where ϕ is the parabolic field which is the paraxial approximation to the Helmholtz field Φ . Both the Helmholtz equation and the parabolic equation have the same set of

eigenfunctions and the Helmholtz eigenvalues can be obtained from those of the parabolic equation. Let us assume that the solution of the Helmholtz equation can be written as [4]

$$\Phi(x, y, z) = u_N^h(x, y)e^{-jk_{z,N}^h z} \quad (3.6)$$

and the solution of the parabolic equation can also be written as

$$\phi(x, y, z) = u_N^p(x, y)e^{-jk_{z,N}^p z} \quad (3.7)$$

where h and p notations refer to the Helmholtz and the parabolic solutions respectively, $k_{z,N}^h$ and $k_{z,N}^p$ are the propagation constants of the Nth modes based on the Helmholtz equation and the parabolic equation respectively. The functions $u(x,y)$ are the transverse mode distributions, and the prime on the Helmholtz propagation constant (h') is just a reminder that a carrier phase constant ($k_0 n_0$) has been removed from the true value. Substituting Eqs. 3.6 and 3.7 into Eqs. 3.3 and 3.5 respectively, gives

$$\left[(k_{z,N}^h)^2 + 2k_0 n_0 k_{z,N}^h \right] u_N^h = \nabla_{\perp}^2 u_N^h + k_0^2 (n^2 - n_0^2) u_N^h \quad (3.8a)$$

$$2k_0 n_0 k_{z,N}^p u_N^p = \nabla_{\perp}^2 u_N^p + k_0^2 (n^2 - n_0^2) u_N^p \quad (3.8b)$$

where the transverse Laplacian operator is defined as $\nabla_{\perp}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. Since the operators on the right hand side of Eqs. 3.8a and 3.8b are the same then $u_N^h(x, y) = u_N^p(x, y)$ which means that the eigenfunctions are identical and also from the same equations the relationship between the Helmholtz propagation constant and the parabolic one can be written as [4]

$$(k_{z,N}^h)^2 + 2k_0 n_0 k_{z,N}^h = 2k_0 n_0 k_{z,N}^p \quad (3.9a)$$

or

$$k_{z,N}^p = \frac{(k_{z,N}^h)^2 + 2k_0 n_0 k_{z,N}^h}{2k_0 n_0} \quad (3.9b)$$

then the true Helmholtz propagation constant $k_{z,N}^h$ (after adding the carrier phase constant) can be computed from the parabolic propagation constant using [4]

$$k_{z,N}^h = k_0 n_0 \sqrt{1 + \frac{2k_{z,N}^p}{k_0 n_0}} \quad (3.10)$$

The above relationships are very useful since they enable the eigenvalues and the eigenvectors of the Helmholtz equation to be computed from the parabolic counterparts.

3.3 The BPM Solutions

As mentioned in the introduction, there are many ways to solve the parabolic equation in (3.5) numerically. In this section the formulation of the FFT-BPM and the FD-BPMs based on the solution of the parabolic equation are shown. Let us start with the famous split-operator which is the basic equation for the FFT-BPM, ADI-BPM and RS-BPM. If we write the parabolic equation in (3.5) as

$$j \frac{\partial \phi(x, y, z)}{\partial z} = G \phi(x, y, z) \quad (3.11)$$

where

$$G = \frac{1}{2k_0 n_0} \left[\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + k_0^2 (n^2 - n_0^2) \right],$$

then the solution of (3.11) can be written formally as [47,49,50]

$$\phi(x, y, z + \Delta z) = e^{(-j\Delta z G)} \phi(x, y, z) \quad (3.12)$$

where Δz is the propagation interval along the z -axis which should be small. In order to put Eq. 3.12 into a computational form, the equation can be rewritten as the following well-known symmetrised split operator which is accurate to the second order in Δz as [1-7]

$$\phi(x, y, z + \Delta z) = e^{\left(-j\frac{\Delta z}{2a_0} \nabla_x^2\right)} e^{\left(-j\frac{\Delta z}{2a_0} \nabla_y^2\right)} e^{\left(-j\frac{\Delta z}{a_0} d_0\right)} e^{\left(-j\frac{\Delta z}{2a_0} \nabla_x^2\right)} e^{\left(-j\frac{\Delta z}{2a_0} \nabla_y^2\right)} \phi(x, y, z) + O((\Delta z)^3) \quad (3.13)$$

where

$$\begin{aligned} a_0 &= 2k_0 n_0 \\ d_0(x, y, z) &= k_0^2 [n^2(x, y, z) - n_0^2] \\ \nabla_\alpha^2 &= \frac{\partial^2}{\partial \alpha^2}, \quad (\alpha = x, y) \end{aligned}$$

Eq. 3.13 is now suitable for generating a numerical solution. It must be noted that the operators in the above equation are unitary for real refractive index and thus the numerical solution is unconditionally stable.

3.3.1 The Fast Fourier Transform (FFT-BPM)

Eq. 3.13 could be solved numerically using Fourier transforms to compute some operators in this equation [1-7]. The transverse space for both the field $\phi(x,y,z)$ and the refractive index $n(x,y,z)$ is discretised into 2-D mesh points with equal mesh spacing for each direction. In the Fourier space the partial derivatives of the transverse directions (∇_x^2, ∇_y^2) are equal to $(-k_x^2, -k_y^2)$ where k_x and k_y are the transverse wave numbers equal to [1-7]

$$k_x = \frac{2\pi i}{L_x}, \quad \left(i = 0, \pm 1, \pm 2, \dots, \pm \frac{M_x}{2} \right)$$

$$k_y = \frac{2\pi m}{L_y}, \quad \left(m = 0, \pm 1, \pm 2, \dots, \pm \frac{M_y}{2} \right)$$

where L_x and L_y are the total length of the computational windows in the x and y direction respectively and M_x and M_y are the total number of mesh points in the x and y directions respectively. Now Eq. 3.13 can be executed numerically using the following steps; first, the discretised field is transformed into Fourier space $\phi(k_x, k_y, z)$ and multiplied by the last two operators on the right hand side of Eq. 3.13 in Fourier space. The result of these two multiplications is inverted back into spatial space and multiplied with the operator that contains the refractive index. The result is then transformed again into Fourier space for the multiplication with the first two operators on the right hand side of Eq. 3.13 and finally inverted into spatial domain to find the field $\phi(x,y,z+\Delta z)$ which is a small distance spaced from the field $\phi(x,y,z)$. The operation of transforming from Fourier space and its inverse can be executed numerically using the well known algorithm Fast Fourier Transform (FFT). Then the field can be propagated to any distance by repeating the same process above [1-7].

3.3.2 The Finite Difference BPM (FD-BPM)

Another way to solve Eq. 3.13 numerically, other than the FFT approximation, is to use finite-differences (FD). FD methods are well known for their simplicity and flexibility

where it is desirable to transform complicated mathematical problems into simpler forms. This section discusses three different FD methods to solve the parabolic equation numerically.

3.3.2.1 The Alternating Direction Implicit method (ADI-BPM)

We start again with the split operator equation in (3.13) where this time the operators that contain the partial derivatives can be approximated using finite-difference techniques rather than the Fourier transform. We use the following Cayley form for the representation of these operators, which is second order accurate [49,50]

$$e^{\left(-j\frac{\Delta z}{2a_0}\nabla_\alpha^2\right)} = \frac{\left[1 - j\frac{\Delta z}{4a_0}\frac{\partial^2}{\partial\alpha^2}\right]}{\left[1 + j\frac{\Delta z}{4a_0}\frac{\partial^2}{\partial\alpha^2}\right]}, \quad (\alpha = x, y) \quad (3.14)$$

The form of (3.14) could operate on any field F at any stage of the split operator equation (e.g. at s) as [49]

$$\left[1 + j\frac{\Delta z}{4a_0}\frac{\partial^2}{\partial\alpha^2}\right]F^{s+\Delta}(\alpha) = \left[1 - j\frac{\Delta z}{4a_0}\frac{\partial^2}{\partial\alpha^2}\right]F^s(\alpha) \quad (3.15)$$

where $s+\Delta$ is an intermediate stage for F in the split operator. The partial derivatives in (3.15) can be replaced by their finite-difference approximations; if the second order central difference approximation is used, then the right hand side of (3.15) involves a direct multiplication of the discretised field with a tridiagonal matrix, and the operator on the left side involves an inversion of a tridiagonal matrix. This method is stable because the operator in (3.14) is unitary. On the other hand, if the operator is approximated as [50]

$$e^{\left(-j\frac{\Delta z}{2a_0}\nabla_\alpha^2\right)} = \left[1 - j\frac{\Delta z}{4a_0}\frac{\partial^2}{\partial\alpha^2}\right] \quad (3.16a)$$

or

$$e^{\left(-j\frac{\Delta z}{2a_0}\nabla_\alpha^2\right)} = \left[1 + j\frac{\Delta z}{4a_0}\frac{\partial^2}{\partial\alpha^2}\right]^{-1} \quad (3.16b)$$

The block diagonal structure of S_x^e and S_x^o simplifies the calculation of the operator tremendously, because the problem has been reduced to an exponentiation of 2 by 2 matrices of the form [42]

$$A_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (3.22)$$

The eigenvalues of A_x are $\zeta_1 = 1$ and $\zeta_2 = -1$ and the corresponding eigenvectors are $v_1 = (1,1)^T$ and $v_2 = (1,-1)^T$ respectively. The exponentiation of (3.21) can be carried out analytically as

$$e^{(-j\tau_x A_x)} = C e^{(-j\tau_x J)} C^{-1}, \quad \tau_x = \frac{\Delta z}{2a_0 \Delta x^2} \quad (3.23)$$

where

$$C = [v_1, v_2] = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$C^{-1} = \frac{1}{|C|} \begin{bmatrix} c_{22} & -c_{12} \\ -c_{21} & c_{11} \end{bmatrix} = -\frac{1}{2} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$e^{(-j\tau_x J)} = \begin{bmatrix} e^{-j\tau_x} & 0 \\ 0 & e^{j\tau_x} \end{bmatrix}$$

Then (3.23) is equal to

$$e^{(-j\tau_x A_x)} = -\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} e^{-j\tau_x} & 0 \\ 0 & e^{j\tau_x} \end{bmatrix} \begin{bmatrix} -1 & -1 \\ -1 & 1 \end{bmatrix}$$

Combining all terms we reach [42]

$$e^{(-j\tau_x A_x)} = \begin{bmatrix} \cos \tau_x & -j \sin \tau_x \\ -j \sin \tau_x & \cos \tau_x \end{bmatrix} = \begin{bmatrix} \alpha_x & \beta_x \\ \beta_x & \alpha_x \end{bmatrix} \quad (3.24)$$

Thus we may now show how the operator in Eq. 3.21 can be implemented when it is acting on a discretised field F :

$$e^{\left(-j\frac{\Delta z}{2a_0}v_x^2\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^s = e^{\left(-j\frac{\Delta z}{2a_0}s_x^o\right)} e^{\left(-j\frac{\Delta z}{2a_0}s_x^e\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^s, \quad (3.25a)$$

which may be written as

$$e^{\left(-j\frac{\Delta z}{2a_0}v_x^2\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^s = e^{\left(-j\frac{\Delta z}{2a_0}s_x^o\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s*} = \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s+\Delta} \quad (3.25b)$$

where

$$\begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s*} = e^{\left(-j\frac{\Delta z}{2a_0}s_x^e\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^s = \begin{bmatrix} \alpha_x & \beta_x \\ \beta_x & \alpha_x \end{bmatrix} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^s$$

with $(i = 1, 3, 5, \dots, M_x - 1)$

$$\begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s+\Delta} = e^{\left(-j\frac{\Delta z}{2a_0}s_x^o\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s*} = \begin{bmatrix} \alpha_x & \beta_x \\ \beta_x & \alpha_x \end{bmatrix} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s*}$$

with $(i = 0, 2, 4, \dots, M_x)$

Periodic boundary conditions are used for the boundary field points at F_0 and F_{M_x+1} to ensure the unitarity of the operators. This means that the value of F_0 is set equal to the value of F_{M_x} and then on the opposite side F_{M_x+1} is set equal to the value of F_1 . We must note that all the computational operations in Eq. 3.25 are simple summations of two local points which makes the method very efficient and highly parallel [45-46]. Now we can write the full equation of the RS-BPM which is accurate to the second order in z [10,42]

$$\begin{aligned} \phi(x, y, z + \Delta z) = & e^{(ps_y^o)} e^{(qs_y^e)} e^{(ps_y^o)} e^{(ps_x^o)} e^{(qs_x^e)} e^{(ps_x^o)} e^{\left(-j\frac{\Delta z}{a_0}U\right)} \\ & \cdot e^{(ps_y^e)} e^{(qs_y^o)} e^{(ps_y^e)} e^{(ps_x^e)} e^{(qs_x^o)} e^{(ps_x^e)} \phi(x, y, z) + O((\Delta z)^3) \end{aligned} \quad (3.26)$$

where

$$U(x, y, z) = d_o(x, y, z) - \frac{2}{\Delta x^2} - \frac{2}{\Delta y^2}$$

$$q = -j \frac{\Delta z}{2a_o} = 2p$$

It is to be noticed from Eq. 3.26 that to regain second order accurate formula with respect to z , the operator splitting, for each dimension, must be written as three operators rather than two operators of Eq. 3.21 [42].

3.3.2.3 The Explicit Finite-Difference method (EFD-BPM)

The Explicit Finite-Difference (EFD) Method is the simplest and the most attractive BPM method. It involves a direct application of the central finite-difference approximation to the parabolic equation [17-18]. Let us consider the parabolic equation in (3.5) where we use the following central difference approximation to replace the partial derivatives which are second order accurate formulas

$$\left. \begin{aligned} \frac{\partial^2 F(\alpha)}{\partial \alpha^2} &= \frac{F(\alpha - \Delta\alpha) - 2F(\alpha) + F(\alpha + \Delta\alpha)}{\Delta\alpha^2}, & (\alpha = x, y) \\ \frac{\partial F(z)}{\partial z} &= \frac{F(z + \Delta z) - F(z - \Delta z)}{\Delta z} \end{aligned} \right\} \quad (3.27)$$

Then the Explicit Finite Difference BPM (EFD-BPM) can be expressed simply by the following numerical equation [17-18]

$$\begin{aligned} \phi_{i,m}(z + \Delta z) &= \phi_{i,m}(z - \Delta z) - \frac{2j\Delta z}{a_o \Delta x^2} [\phi_{i-1,m}(z) + \phi_{i+1,m}(z)] \\ &\quad - \frac{2j\Delta z}{a_o \Delta y^2} [\phi_{i,m-1}(z) + \phi_{i,m+1}(z)] - \frac{2j\Delta z}{a_o} U_{i,m} \phi_{i,m}(z) \end{aligned} \quad (3.28)$$

where i and m represent the discretisation of the transverse co-ordinates x and y respectively. The propagation of the optical field in Eq. 3.28 is straight forward since it involves a multiplication of the input field with a very sparse matrix with only five elements in each row, which makes the method very efficient and highly parallel [45,46]. This algorithm is stable under the following condition [18]

$$\Delta z < 2k_o n_o \left[\frac{4}{\Delta x^2} + \frac{4}{\Delta y^2} + k_o^2 |n_{i,m}^2 - n_o^2|_{\max} \right]^{-1} \quad (3.29)$$

3.4 The Spectral Method

In chapter 2, we have reviewed some numerical methods to compute propagation constants and field distributions of guided modes for three-dimensional z-invariant waveguides. The BPM also could be employed to compute guided-mode propagation constants and field distributions. There exist two methods that can compute the characteristics of z-invariant waveguides from the evolution of the BPM field. The first is called the imaginary distance propagation technique [9,14] and the second is called the spectral technique [3,6]. The imaginary distance propagation method first computes the field distribution and then calculates the propagation constant from the computed field distribution. Opposite to this, the spectral method first calculates the propagation constant, and then the field distribution is computed using the computed propagation constant. In the imaginary distance propagation method, the parabolic equation is transformed into a diffusion equation by replacing the $-j$ in the numerical solution of the equation by unity. Then the new numerical imaginary distance solution is used to propagate an arbitrary input field to a certain distance while constantly normalising the field to avoid numerical instability (overflow). From guided-wave theory, it is known that the lowest order mode has the largest propagation constant therefore during propagation it has the highest amplification compared to the other modes, and in the limit of large distance the lowest order mode dominates. Once the lowest order mode distribution is found, the parabolic propagation constant can be obtain using the following relation [9]

$$k_z^p = \frac{\iint [-(\nabla_{\perp} \phi)^2 + k_o^2 \phi^2 (n^2(x, y) - n_o^2)] dx dy}{\iint 2n_o k_o \phi^2 dx dy} \quad (3.30)$$

The above imaginary distance propagation method can be modified to generate a number of lowest order modes [9,14]. On the other hand, the spectral method simply uses the spectral correlation function of the BPM field and the initial field to locate the parabolic propagation constants in the spectral domain. During propagation of the BPM field (using any of the BPM solutions in section 3.3), we may write the correlation function between the initial and the BPM fields as [3,6]

$$P_c(z) = \iint \phi^*(x, y, 0) \phi(x, y, z) dx dy \quad (3.31)$$

If the BPM field $\phi(x,y,z)$ is expressed as the superposition of orthogonal eigenfunctions of the z-invariant waveguide

$$\phi(x,y,z) = \sum_N a_N u_N(x,y) e^{-jk_{z,N}^p z} \quad (3.32)$$

where $u_N(x,y)$ and $k_{z,N}^p$ are the eigenfunctions and the parabolic propagation constant of the Nth mode of the z-invariant structure. Substituting (3.32) into (3.31) leads to

$$P_c(z) = \sum_N |a_N|^2 e^{-jk_{z,N}^p z} \quad (3.33)$$

The Fourier transform of (3.33) is

$$P_c(k_z) = \sum_N |a_N|^2 \delta(k_z - k_{z,N}^p) \quad (3.34)$$

The delta functions in Eq. 3.34 indicate that in the spectral representation of the correlation function $P_c(z)$, there is a series of resonance peaks at $k_z = k_{z,N}^p$, and their height values are proportional to the mode weight coefficients $|a_N|^2$. Thus, we may obtain any parabolic propagation constant $k_{z,N}^p$, of a z-invariant structure, by performing the numerical calculation of the Fourier transform of the correlation function in Eq. (3.31) and then locating the position of the peaks in the Fourier spectrum. In the ideal case of infinite propagation distance, the Fourier transform of (3.33) gives (3.34), but in practice the field can only be propagated to a finite distance. To account for the finite length of z, the correlation function in (3.33) is multiplied with a window function $w(z)$ before being Fourier transformed [3,50]. A typical Hanning window function can be used which is given as [3,50]

$$w(z) = \begin{cases} 1 - \cos\left(\frac{2\pi z}{Z}\right), & 0 \leq z \leq Z \\ 0, & z > Z \end{cases} \quad (3.35)$$

where Z is the total propagation distance. Then the Fourier transform of the correlation function becomes [3]

$$P_c^w = \sum_N |a_N|^2 \mathcal{L}(k_z - k_{z,N}^p) \quad (3.36)$$

where the line shape function is given by

$$\mathcal{L}(k_z - k_{z,N}^p) = \frac{1}{Z} \int_0^Z e^{j(k_z - k_{z,N}^p)z} w(z) dz \quad (3.37)$$

After computing this line function numerically, then the parabolic propagation constant can be extracted from the spectrum of P_c^m ; increasing the propagation distance Z will increase the accuracy of the computed propagation constant (more sampled points). But it is far more efficient to use a line shape fitting technique that computes accurately the value of the propagation constant from the peak value and its two adjacent neighbouring points (see the appendix of [3]). Once the parabolic propagation constant is found, then the Helmholtz propagation constant can be computed using the relation in Eq. 3.10. Now, the field distribution of any mode can be found using the calculated parabolic propagation constant. If both sides of Eq. 3.32 are multiplied by $Z^{-1}w(z)e^{jk_z z}$ and integrated from 0 to Z , we get [4]

$$\phi(x, y, k_z) = \frac{1}{Z} \int_0^Z \phi(x, y, z) w(z) e^{jk_z z} dz \quad (3.38a)$$

$$= \sum_N a_N u_N(x, y) \mathcal{L}(k_z - k_{z,N}^p) \quad (3.38b)$$

The function in $\phi(x, y, k_z)$ in the above equation will show a maximum if $k_z = k_{z,n}^p$ where we can write $\phi(x, y, k_{z,n}^p)$ as [4]

$$\phi(x, y, k_{z,n}^p) = a_n u_n(x, y) \mathcal{L}(0) + \sum_{N \neq n} a_N u_N(x, y) \mathcal{L}(k_z - k_{z,N}^p) \quad (3.39)$$

The second summation is negligible if most of the excited power belongs to the mode $k_z = k_{z,n}^p$, which can be achieved by exciting the input field with a distribution that resembles the desired mode. Then the numerical computation of a particular mode field distribution from the BPM field can be obtained using Eq. 3.38a with $k_z = k_{z,n}^p$ where the integral is performed numerically.

To summarise, any desired mode of a z -non-varying dielectric waveguide can be computed from any BPM field solutions, of section 3.3, using the spectral method, by propagating an input field which is very close to the shape of the desired mode field, to a certain distance. To find the propagation constant of the mode, first the correlation function

in (3.31) is evaluated numerically during the course of propagation, then the result is multiplied with the Hanning window function in (3.35) and Fourier transformed. The parabolic propagation constant is computed by locating the peak of the mode in the spectral domain. To find the field distribution of the mode, the same input field is re-propagated using the BPM while performing the integral in Eq. 3.38a numerically after replacing the propagation constant in that integral with the already computed one. The main advantage of this method over the imaginary distance technique is that it is not restricted to computing the lowest order modes, but can be applied to cases where all modes, including radiation modes, can be present in the structure [3,4,6,7].

REFERENCES

- [1] Feit, M. D. and Fleck, J. A., "Light Propagation in Graded-Index Optical Fibers", *App. Opt.*, **17**, 3990-3998, 1978.
- [2] Feit, M. D. and Fleck, J. A., "Calculation of Dispersion in Graded-Index Multimode Fibers by a Propagating-Beam Method", *App. Opt.*, **18**, 2843-2851, 1979.
- [3] Feit, M. D. and Fleck, J. A., "Computation of Mode Properties in Optical Fiber Waveguides by a Propagating Beam Method", *App. Opt.*, **19**, 1154-1164, 1980.
- [4] Feit, M. D. and Fleck, J. A., "Computation of Mode Eigenfunctions in Graded-Index Optical Fibers by the Propagating-Beam Method", *App. Opt.*, **19**, 2240-2246, 1980.
- [5] Feit, M. D. and Fleck, J. A., "Mode Properties and Dispersion for Two Optical Fiber-Index Profiles by the Propagating Beam Method", *App. Opt.*, **19**, 3140-3150, 1980.
- [6] Feit, M. D. and Fleck, J. A., "Mode Properties of Optical Fibers With Lossy Components by the Propagating Beam Method", *App. Opt.*, **20**, 848-856, 1981.
- [7] Feit, M. D. and Fleck, J. A., "Analysis of Rib Waveguides and Couplers by the Propagating Beam Method", *J. Opt. Soc. Am. A*, **7**, 73-79, 1990.
- [8] Roey, J. V., Donk, J. V., Lagasse, P. E., "Beam Propagation Method: Analysis and assessment", *J. Opt. Soc. Am.*, **71**, 803-810, 1981.
- [9] Yevick, D. and Hermansson, B., "New Formulations of the Matrix Beam Propagation Method: Application to Rib Waveguides", *IEEE J. Quant. Elect.*, **25**, 221-229, 1989.
- [10] Yevick, D. and Hermansson, B., "Efficient Beam Propagation Techniques", *IEEE J. Quant. Elect.*, **26**, 109-112, 1990.
- [11] Yevick, D. and Hermansson, B., "Split-Step Finite-Difference Analysis of Rib Waveguides", *Elect. Lett.*, **25**, 461-462, 1989.
- [12] Hermansson, B., Yevick, D., Bardyszewski, W. and Glasner, M., "The Unitarity of Split-Operator Finite Difference and Finite-Element Methods: Applications to

- Longitudinally Varying Semiconductor Rib Waveguides", *J. Lightwave Tech.*, **8**, 1866-1873, 1990.
- [13] Yevick, D., Glasner, M. and W., Hermansson, "Numerical Studies of Split-Operator Finite-Difference Alternating-Direction Implicit Propagation Techniques Based on Generalized Pade Approximants", *Opt. Lett.*, **17**, 725-727, 1992.
- [14] Yevick, D., "A Guide to Electric Field Propagation Techniques for Guided-Wave Optics", *Opt. Quant. Elect.*, **26**, 185-197, 1994.
- [15] Hendow, S. T. and Shaker, S. A., "Recursive Numerical Solution for Nonlinear wave Propagation in Fibers and Cylindrically Symmetric System", *Appl. Opt.*, 1759-1764, **25**, 1986.
- [16] Chung, Y. and Dagli, N., "An Assessment of Finite Difference Beam Propagation Method", *IEEE J. Quant. Elect.*, **26**, 1335-1339, 1990.
- [17] Chung, Y. and Dagli, N., "Explicit Finite Difference Beam Propagation Method: Application to Semiconductor Rib Waveguide Y-Junction Analysis", *Elect. Lett.*, **26**, 711-713, 1990.
- [18] Chung, Y. and Dagli, N., "Analysis of Z-Invariant and Z-Variant Semiconductor Rib Waveguides by Explicit Finite Difference Beam Propagation Method with Nonuniform Mesh Configuration", *IEEE J. Quant. Elect.*, **27**, 2296-2305, 1991.
- [19] Yamauchi, J., Ando, T. and Nakano, H., "Beam-Propagation Analysis of Optical Fibers by Alternating Direction Implicit Method", *Elect. Lett.*, **27**, 1663-1665, 1991.
- [20] Scarmozzino, R. and Osgood, R. M., "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.
- [21] Accornero, R., Artiglia, M., Coopa, G., Di Vita, P. Lapenta, G., Potenza, M., and Ravetto, P., "Finite Difference Methods for the Analysis of Integrated Optical Waveguides", *Elect. Lett.*, **26**, 1959-1960, 1990.
- [22] Artiglia, M., Vita, P. D., Potenza, M., Coopa, G., Lapenta, G., and Ravetto, P., "Variable Grid Finite Difference Methods for Study of longitudinally Varying Planar Waveguides", *Elect. Lett.*, **27**, 474-475, 1991.
- [23] Pregla, R. and Gerdes, J., "Beam-Propagation Algorithm Based on the Method of Lines", *J. Opt. Soc. Am. B*, **8**, 389-394, 1991.
- [24] Yevick, D. and Glasner, M., "Analysis of Forward Wide-Angle Light Propagation in Semiconductor Rib Waveguides and Integrated-Optic Structures", *Elect. Lett.*, **25**, 1611-1613, 1989.
- [25] Hadley, G. R., "Wide-Angle Beam Propagation Using Pade' Approximation Operators", *Opt. Lett.*, **17**, 1426-1428, 1992.

- [26] Hadley, G. R., "Multistep Method For Wide-Angle Beam Propagation", *Opt. Lett.*, **17**, 1743-1745, 1992.
- [27] Huang, W. P. and Xu, C. L., "A Wide-Angle Vector Beam Propagation Method", *IEEE Phot. Tech. Lett.*, **4**, 1118-1120, 1992.
- [28] Kaczmarek, P. and Lagasse, P. E., "Bi-directional Beam Propagation Method", *Elect. Lett.*, **24**, 675-676, 1988.
- [29] Yevick, D., Bardyszewski, W., Hermansson, B., and Glasner, M., "Split-Operator Electric Field Reflection Techniques", *IEEE Phot. Tech. Lett.*, **3**, 527-529, 1991.
- [30] Yevick, D., Xu, J. and Bardyszewski, W., "Split-Operator Calculations of Reflected Electric Field Profiles", *IEEE Phot. Tech. Lett.*, **4**, 1383-1386, 1992.
- [31] Lin, H. H. and Korpel, A., "Heuristic Scalar Paraxial Beam Propagation Method Taking into Account Continuous Reflections", *J. Opt. Soc. Am. B*, **8**, 849-857, 1991.
- [32] Clauser, R. and Von-Allmen, P., "Vectorial Beam Propagation Method for Integrated Optics", *Elect. Lett.*, **27**, 654-655, 1991.
- [33] Huang, W. P., Xu, C. L. and Chaudhari, S. K., "A Vector Beam Propagation Method for Guided-Wave Optics", *IEEE Phot. Tech. Lett.*, **3**, 910-913, 1991.
- [34] Huang, W. P., Xu, C. L. and Chaudhari, S. K., "A Vector Beam Propagation Method on H Fields", *IEEE Phot. Tech. Lett.*, **3**, 1117-1120, 1991.
- [35] Huang, W. P., Xu, C. L. and Chaudhari, S. K., "A Finite-Difference Vector Beam Propagation Method for Three-Dimensional Waveguide Structures", *IEEE Phot. Tech. Lett.*, **4**, 148-151, 1992.
- [36] Huang, W. P., Xu, C. L., Chu, S. T., C. L. and Chaudhari, S. K., "The Finite-Difference Vector Beam Propagation Method: Analysis and Assessment", *J. Lightwave Tech.*, **10**, 295-305, 1992.
- [37] Huang, W. P., Xu, C. L. and Chaudhari, S. K., "Application of the Finite-Difference Vector Beam Propagation Method to Directional Coupler Devices", *IEEE J. Quant. Elect.*, **28**, 1527-1532, 1992.
- [38] Huang, W. P. and Xu, C. L., "Simulation of Three-Dimensional Optical Waveguides by a Full-Vector Beam Propagation Method", *IEEE J. Quant. Elect.*, **29**, 2639-2649, 1993.
- [39] Liu, P. L. and Li, B. J., "Semivectorial Beam-Propagation Method for Analyzing Polarized Modes of Rib Waveguides", *IEEE J. Quant. Elect.*, **28**, 778-782, 1992.
- [40] Hadley, G. R., "Transparent Boundary Condition for Beam Propagation", *Opt. Lett.*, **16**, 624-626, 1991.
- [41] Hadley, G. R., "Transparent Boundary Condition for the Beam Propagation Method", *IEEE J. Quant. Elect.*, **28**, 363-370, 1992.

- [42] DeRaedt, H., "Product Formula Algorithms for Solving the Time Dependent Schrodinger Equation", *Comp. Phys. Rep.*, **7**, 1-72, 1987
- [43] Richardson, J., "Visualizing Quantum Scattering on the CM-2 Supercomputer", *Comp. Phys. Comm.*, **63**, 84-94, 1991.
- [44] Richardson, J., "Computational Physics on the CM-2 Supercomputer", *Phys. Rep.*, **207**, 305-320, 1991.
- [45] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Methods", *IEEE Phot. Tech. Lett.*, **6**, 848-850, July 1994.
- [46] Masoudi, H. M. and Arnold, J. M., "Parallel Three-Dimensional Finite-Difference Beam Propagation Methods", *Int. J. of Num. Mod.*, **8**, 95-107, March-April 1995.
- [47] Mitchell, A. R. and Griffiths, D. F. , "The Finite Difference Method in Partial Differential Equations" , John Wiley & Sons, 1980.
- [48] Strikwerda, J. C. , "Finite Difference Schemes and Partial Differential Equations", Wadsworth & Books, California, 1989.
- [49] Goldberg, A., Schey, H. M. and Schwarz, J. L., "Computer Generated Motion Pictures of One-Dimensional Quantum-Mechanical Transmission and Reflection Phenomena", *Am. J. Phys.* , **35**, 177-186, 1967.
- [50] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. " The Art of scientific computing: numerical recipes in FORTRAN", Cambridge Uni. Press, 1992.

Chapter 4

Parallel Computing

4.1 Introduction

Parallel computers have evolved from many stages of conventional computers through the last three decades: the main-frame stage in the 1960s; the mini-frame stage in the 1970s; and the personal computers in the 1980s. They are designed specifically to speed up the execution of large numerical algorithms using the idea of breaking the problem into small pieces and arranging to solve for all pieces simultaneously [1-7]. They may contain a few or thousands of processors connected via communication links for exchanging data. The original idea of producing a fast computational tool was mainly driven by the need of solving large computational problems (e.g. problems containing partial differential equations). However, scientific problems are not always the same, which means that a multi-purpose parallel machine is required that can solve all kinds of large computational problems efficiently (which is not feasible), or that a more realistic approach must be adopted, consisting of different flexible designs [1-7]. There are two principal parallel architectures that can solve most scientific applications efficiently; the first is the fine-grain architecture known as Single Instruction Multiple Data (SIMD) machines, and the second is the medium-grain architecture known as Multiple Instruction Multiple Data (MIMD) machines [2]. Each of these machines has its advantages and disadvantages compared to the other, and each has its own application. In addition, some applications are well suited to both of them.

On the other hand, the existence of large scale processors created a new challenge for scientists in terms of the difficulty of dealing with such large machines, and in terms of reformulating many existing algorithms which were originally derived for serial computers. It has to be noticed, however, that the variety and the flexibility of parallel computers made it difficult for researchers to program algorithms in the new environment, but this has eased the difficulty of finding a new parallel algorithm because it gives the

freedom of parallelising applications efficiently using different, or combinations of, approaches.

In this chapter, we are concerned mainly with two issues, namely the differences between the two parallel machines used in this thesis, and how applications can be implemented on these machines. Section 4.2 shows the configurations of the transputer array (MIMD computer) and the Connection Machine (SIMD computer) where the definitions and the differences between the two parallel mechanisms are illustrated using a simple example. The issues of programming MIMD and SIMD machines are discussed in section 4.3. These issues deal with new parameters that must be considered in parallel programming that were never encountered in serial one. Due to the variety of parallel machines and techniques of parallelising applications, it is very difficult to derive strict rules for parallel programming as for the serial counterpart. For this reason, section 4.3 gives important parameters that affect the performance of a certain parallel application. Also in the same section, two advanced compilers for developing parallel algorithms on the two parallel machines are summarised. Both of these compilers use ordinary FORTRAN coding system, which is one of the most popular scientific computer languages.

4.2 The Parallel Machines¹

Throughout this thesis, we will be concerned with the following two kinds of parallel machines:

- 1- The *transputer array* is a Multiple Instructions Multiple Data (MIMD) which is a PARSYTEC super-cluster consisting of 64 IMS-T800 processors each with 4 Mbytes RAM and the front end computer to the machine is a SUN-SPARC station 2 [8-9].
- 2- The Single Instructions Multiple Data (SIMD) Machine is a Connection Machine CM-200 consisting of 16k (16,384) bit serial processors plus 512 floating-point accelerator co-processors and a total of 0.5 G bytes, where the front end computer is a SUN (4/370) [10-13].

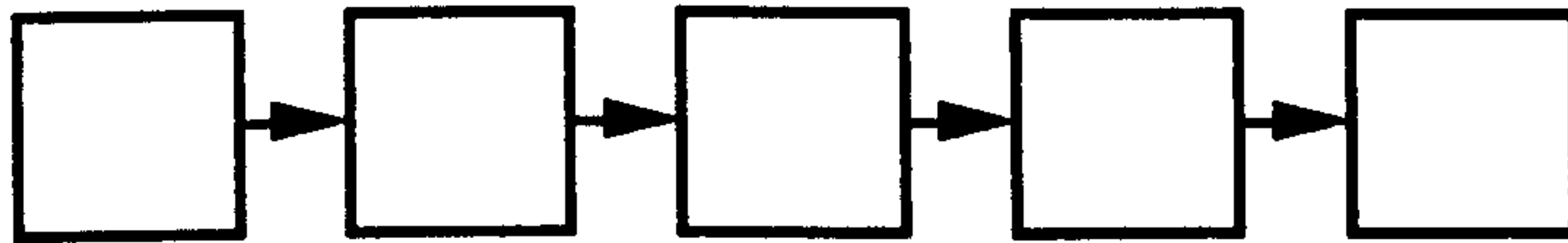
Both of the above machines represent the most common types used frequently in the literature [2,7,14,15]. MIMD computing means that many processors can execute simultaneously different instructions on different data. We can imagine these processors as

¹These machines were available for use during the course of my PhD program. The PARSYTEC was in the Department of Electronics and Electrical Engineering at the University of Glasgow (Glasgow, UK). The Connection Machine was in the Edinburgh Parallel Computing Centre at the University of Edinburgh (Edinburgh, UK) and was operated on remotely from the University of Glasgow.

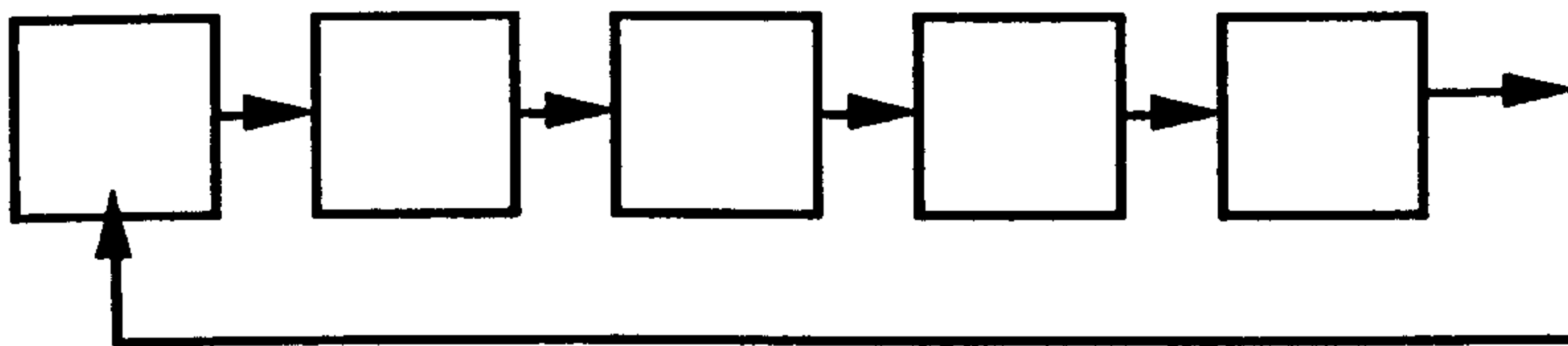
separate computers without a central control unit. The processors are configured and connected explicitly according to the application. This flexibility of the MIMD computer means that the programming is more difficult to control in terms of operational synchronisation and efficient distribution of computations between processors. However, generally this computer is very useful when the algorithm contains several heterogeneous tasks to be executed concurrently. Fig. 4.1 shows different common topologies for configuring these processors. In the figure, the boxes mean processors and the arrows mean communication links between the processors for exchanging data. With these flexible arrangements, different algorithms or subroutines could be executed separately and simultaneously on different processors [8,9].

On the other hand, the SIMD Machine is the most powerful computer available, and its application is very useful for massively parallel scientific computing. SIMD means that the central control unit of the machine broadcasts a stream of identical instructions to all processors, then these instructions are executed simultaneously by all processors on all local data that exist in the private memory of each processor. In other words the synchronisation mechanism is achieved globally, rather than locally as in the MIMD machine. The Connection Machine (CM), in Fig. 4.2, has a SUN computer as a central control unit where the code of instructions resides and is executed serially [12].

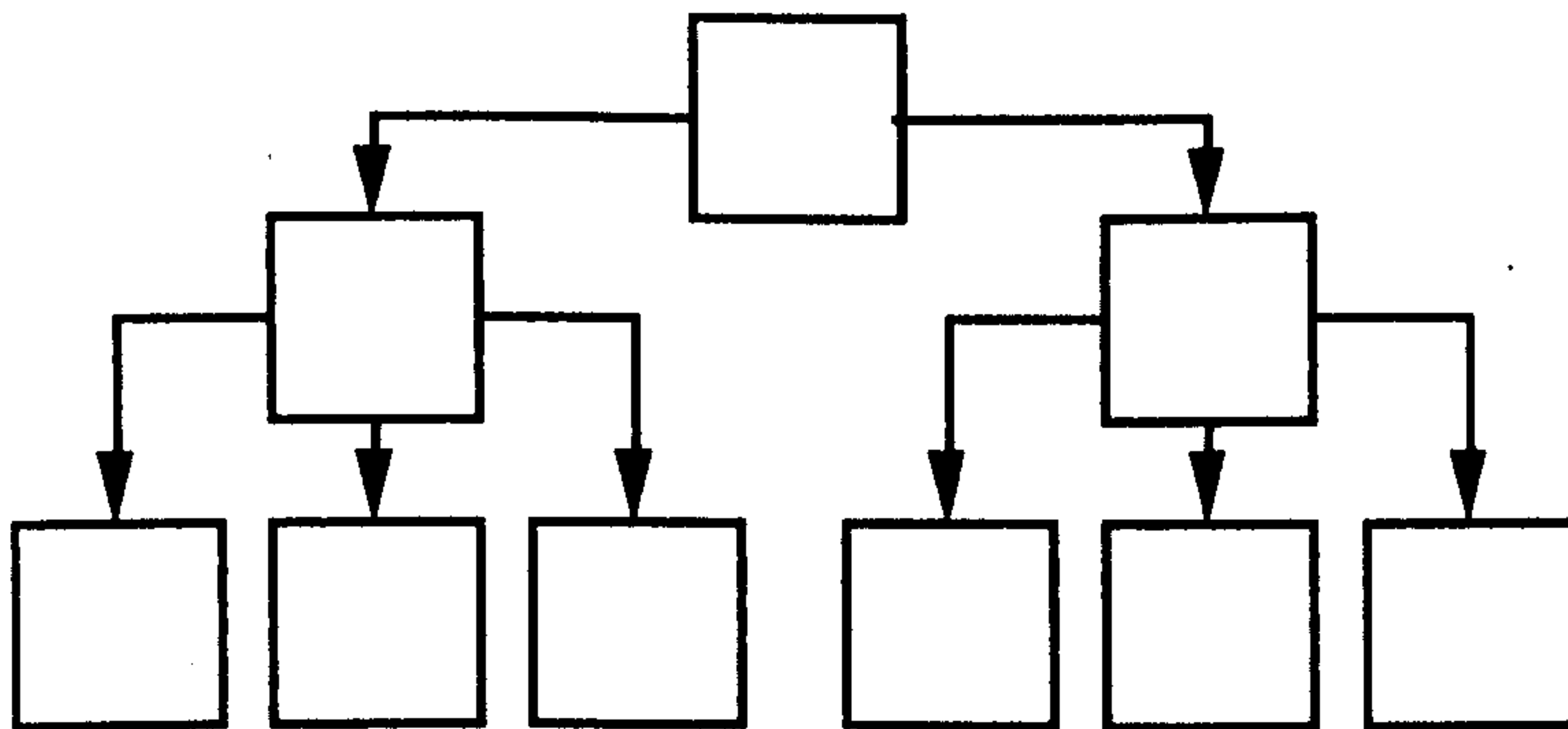
To understand the practical mechanisms of these two large computers, the following simple example shows how to sum two numerical arrays using two different implementations (MIMD and SIMD) for both of the architectures described before. Although the following example is very simple, it is intended to focus on the principal difference between the two machines, and also to help in understanding parallel programming languages in the following section.



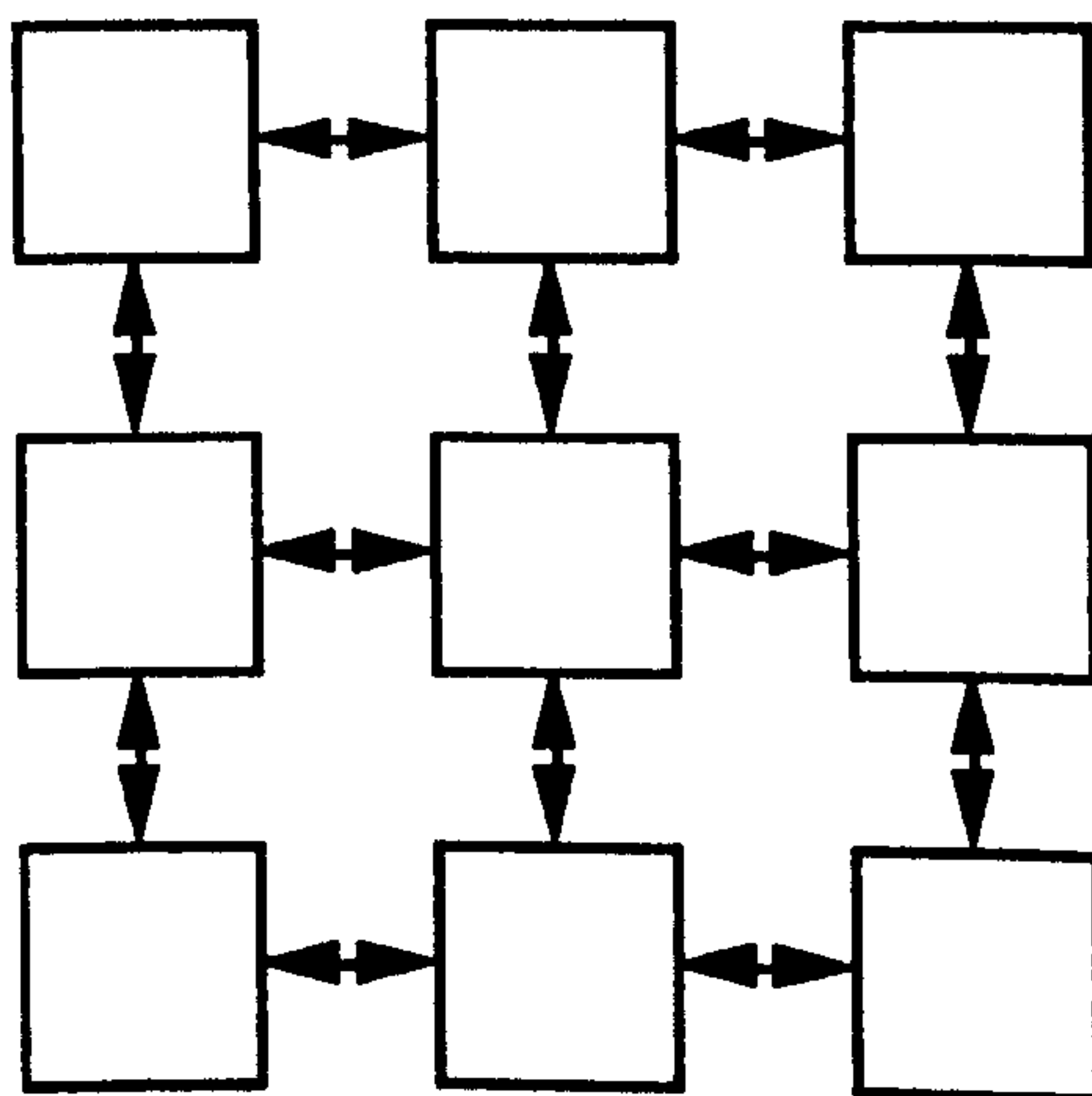
a-Pipeline Topology



b-Ring Topology



c-Farm Topology



d-2-D Grid Topology

Fig.4.1 Common transputer topologies.

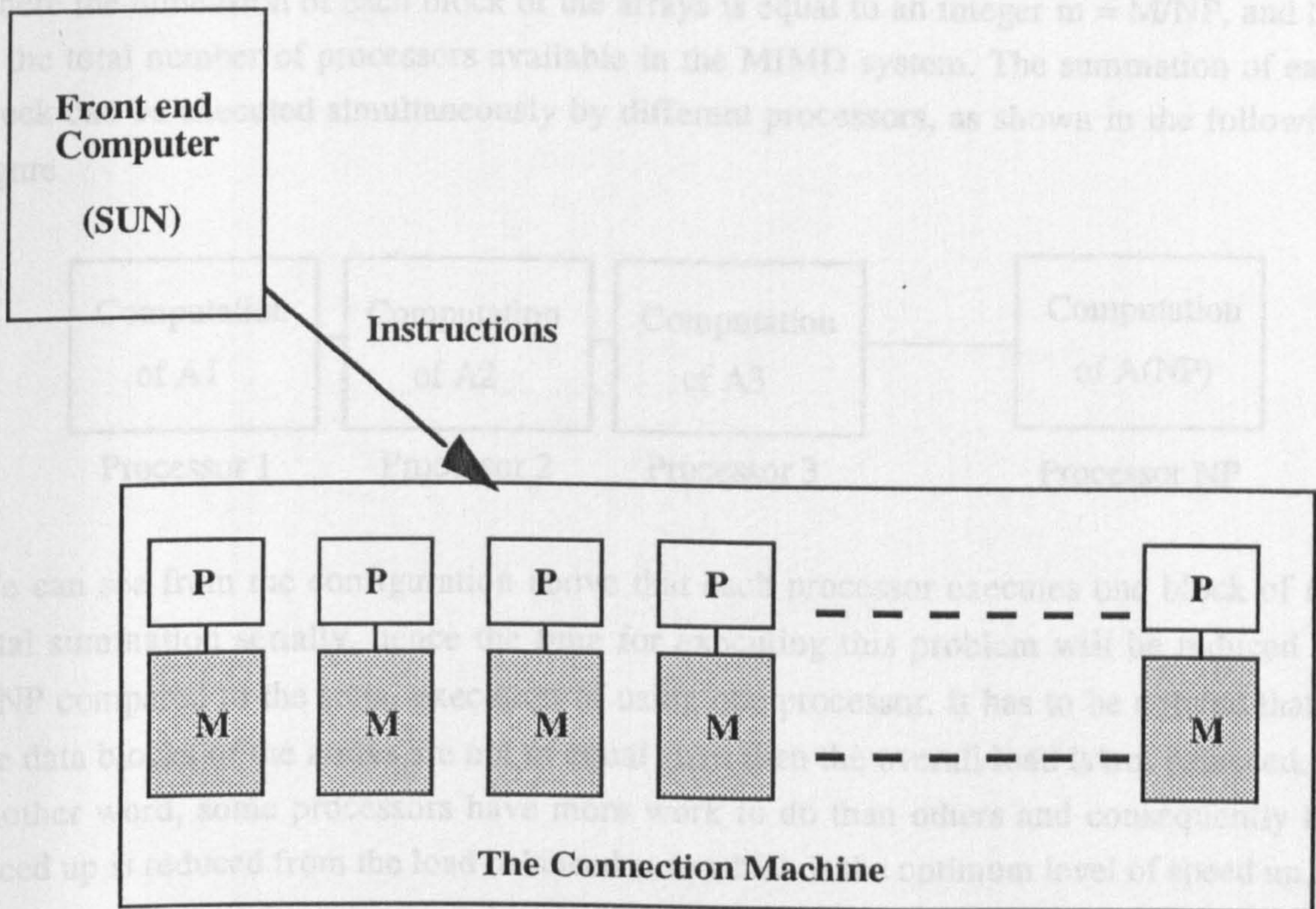


Fig 4.2 The SIMD Machine

Example 4.1 :

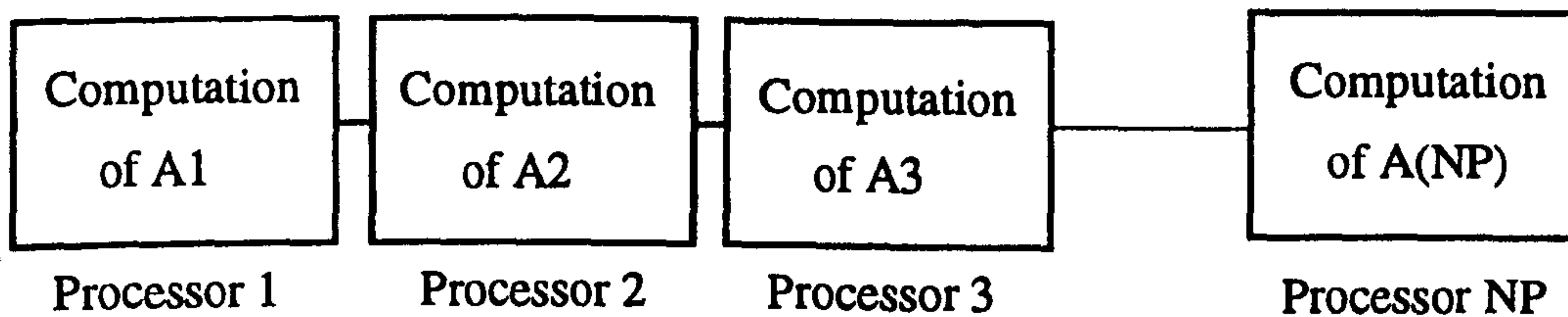
Suppose that it is required to sum two large one-dimensional arrays **B** and **C** with **M** elements each.

1-MIMD Implementation:

Let the summation of these arrays can be divided into equal blocks of data as

$$\mathbf{A} = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ \vdots \\ A_{NP} \end{bmatrix} = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ \vdots \\ B_{NP} \end{bmatrix} + \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ \vdots \\ C_{NP} \end{bmatrix} = \begin{bmatrix} B_1 + C_1 \\ B_2 + C_2 \\ \vdots \\ \vdots \\ B_{NP} + C_{NP} \end{bmatrix}$$

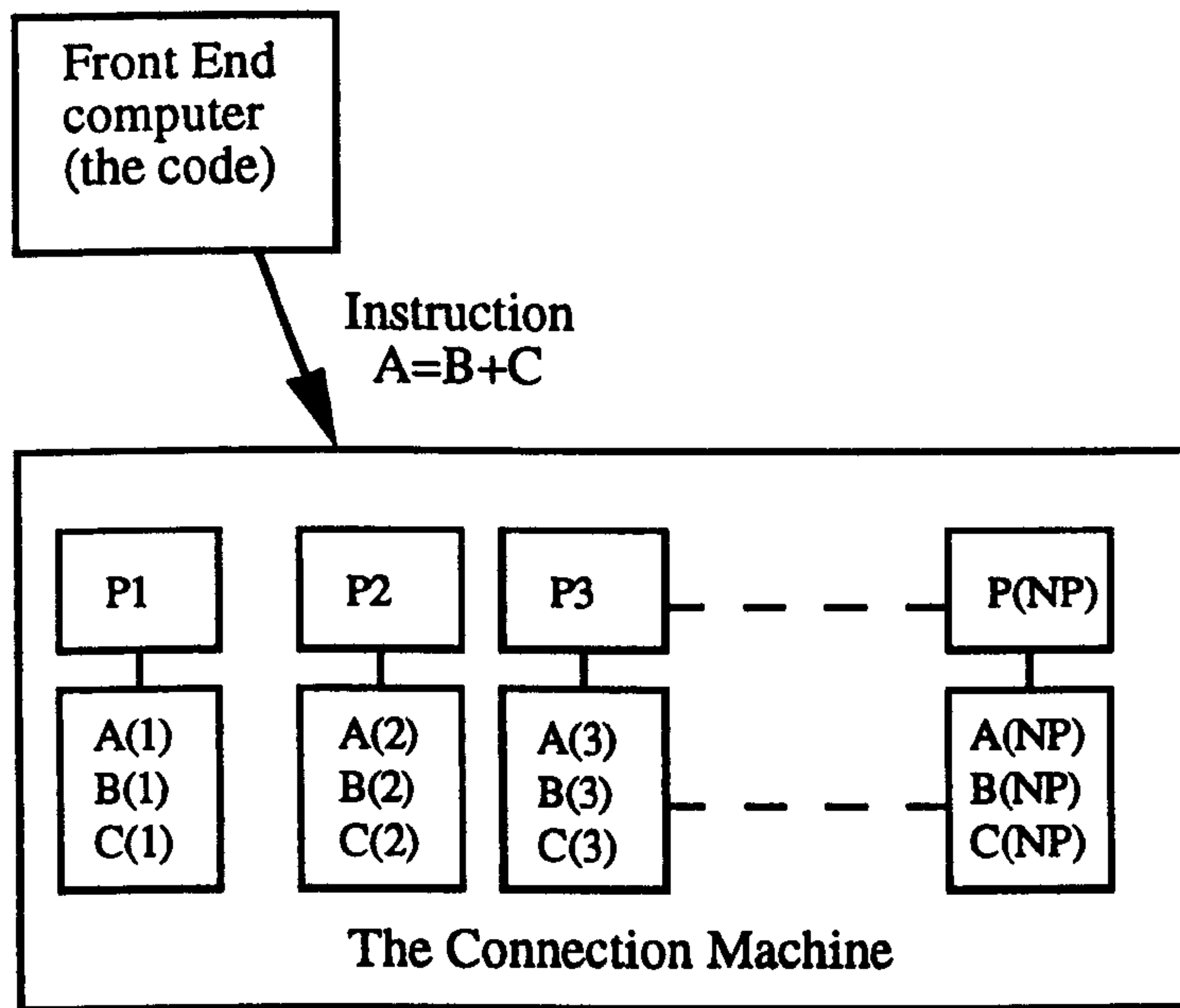
where the dimension of each block of the arrays is equal to an integer $m = M/NP$, and NP is the total number of processors available in the MIMD system. The summation of each block can be executed simultaneously by different processors, as shown in the following figure



We can see from the configuration above that each processor executes one block of the total summation serially, hence the time for executing this problem will be reduced by $1/NP$ compared to the serial execution of using one processor. It has to be noticed that if the data blocks of the arrays are not of equal sizes then the overall load is not balanced. In another word, some processors have more work to do than others and consequently the speed up is reduced from the load balanced case which is the optimum level of speed up.

2-SIMD implementation:

Let A also be the summation of the two arrays as $(A=B+C)$ where each element of A and its corresponding values from B and C are stored in the same memory of a single processor of the CM ($M = NP$). The mechanism of executing this operation is simply as follows. The front end computer sends the summation instruction $(A=B+C)$ to the CM. Then all the CM's processors receive and execute the summation instruction simultaneously (see the Figure below).



It can be observed from the previous example that there are some similarities between the two implementations on both machines. In the MIMD implementation the arrays were divided into blocks of data, where as in the SIMD implementation the arrays were divided into scalar elements that could be solved concurrently. However, there is a feature in MIMD application that has not been revealed in this example, which is the ability to execute multiple different tasks (e.g. subroutines or algorithms) at the same time. On the other hand, the synchronisation in the MIMD implementation has to be forced by the application, but in the SIMD implementation the synchronisation is automatically there.

4.3 Parallel Programming

Due to the diversity of applications, techniques and parallel machines, it is difficult to generalise rules for parallel programming. However, we may write some guidelines that are applicable for most situations. As mentioned before, the goal of parallel techniques is to speed up the execution of large applications efficiently; thus it is very important to note that parallel programming is not a simple extension of serial programs. Although in some situations, blind extension of serial programs to parallel may lead to a speed up of the execution of the algorithm, a very low efficiency may result. In the parallel environment, we must think in parallel terms in order to exploit the maximum efficiency out of the

parallel machine. The starting point to design a parallel algorithm would follow one of the three following methods [4]:

- 1- A modification of an existing serial algorithm.
- 2- An invention of a new parallel scheme.
- 3- Adapting another parallel method that solves a similar problem.

The most important element that decides which method to adopt in developing a parallel algorithm is the type of the machine involved. In other words, some algorithms are better suited to a MIMD machine than to a SIMD machine, and the opposite is also true. From the three methods above, the first method may be well suited to the SIMD environment since the instruction codes of this computer are executed serially. The second method is usually difficult and time consuming; however, for some applications it is the only solution. The third method is the most convenient because there is a reasonable amount of literature for parallel techniques of scientific problems from different areas of research [7]. In most circumstances, the solution of scientific problems of partial differential equations is very similar. In addition, there is a great interest in research for parallelising generic matrix manipulation methods for both types of machines MIMD and SIMD (e.g. multiplication, summation and inversion of different kinds of matrices) [7].

The existence of a multi-processor environment, for MIMD computers in particular, means that there are several important issues in parallel implementations that are not usually addressed in conventional serial computations; namely: *task allocation*, *communication* and *synchronisation* [1]. Task allocation is the first step of MIMD parallel implementation, which means breaking the total algorithm into smaller sub-tasks to be assigned to different processors. For some applications, this issue could be the bottle neck in which there exist sub-tasks that can not be computed concurrently with the others. On the other hand, other applications are naturally highly parallel, and the task allocation could be a straight forward job. It is also important from the efficiency point of view to spread the total workload uniformly across working processors. This is known as *load balancing*, which means that no processor is left idle while others are computing. The communication which take place between processors is another very important factor in the parallel implementation which reflects seriously on the total performance of parallel algorithms. Ideally (see example 4.1), NP processors solving a single problem will be NP times faster than a single processor, if zero communication between processors is considered. However, practically this is very difficult to achieve because most applications require movement of information between different tasks of the problem. In terms of parallel computing, very efficient parallel computing requires minimum communications overhead. The third issue to be considered is the synchronisation of the computations of

sub-tasks, where at some point in parallel algorithms processors must wait for the arrival of data or the completion of a certain computation. Consequently, this also could have an effect on the total performance of the parallel algorithm. Finally, it has to be said that combining all of the above issues for MIMD implementation is not trivial. On the other hand, for the SIMD environment the task allocation and the synchronisation issues are enforced by the design of the machine, although the communication is also enforced by the machine for which processors send and receive data from each other automatically (it does not have to be written explicitly). But attention should be paid, for example, to allocating corresponding elements of different numerical arrays (when the two arrays have different sizes) to the same processor avoiding movement of data inside the machine.

We may conclude that writing programs for parallel computers is as complex as the serial ones, but the debugging could be much more difficult². Due the difficulties of efficient parallel programming and error debugging, here follows some guidelines that we use to write a parallel program [2,4,8,9,12]:

- 1- Serial:** The first version of the application should be tested on a serial computer. The test should be in the framework of typical validation of numerical methods : accuracy, convergence, stability, and parameters affecting the overall algorithm. It could be a big mistake and a waste of time if one were to start to write a parallel program without writing a serial one first. In addition, the results of the serial program can be used as references for the parallel ones.
- 2- Reliability:** The first version of the parallel program should not require all the features to be fast and complex. It should be reliable, and produce identical results to the serial program. Trying to include all the features from the start is almost an impossible task especially for complicated applications. For MIMD applications, it is always recommended that the first program should be written to run on a few processors (e.g. two or three).
- 3- Efficiency:** If the program is reliable enough then it should be made as efficient as possible by balancing the load and minimising the communication overhead between processors (some attention should be paid to blind spots especially in MIMD applications).
- 4- Functionality and Flexibility:** We may now increase the range of the problem and add other parameters to make the program flexible.

² A famous dictum: perfect software never exists.

Throughout this work , the PARIX 1.2 operating environment software [8,9] has been used to develop parallel algorithms for the transputer array (MIMD machine) and CM-FORTRAN 2.1.1 [10-13] has been used to develop parallel algorithm that run on the Connection Machine. In the following sub-sections we will discuss these two compilers briefly. Some of the operations, and the overall understanding of parallel programming, may be clarified during practical implementations of applications on these machines.

4.3.1 Programming under PARIX [8,9]

This subsection shows some of the essential elements to develop a parallel algorithm, using PARIX and FORTRAN 77, for the implementation on the super-cluster. PARIX is a powerful operating software built for PARSYTEC systems, which runs under UNIX commands where ordinary FORTRAN 77 could be used for local computations and parallel processing commands (e.g. communication) are realised through built-in library subroutines. Under PARIX, one main program is needed to run all processors from the start to the end, and a copy of this program is loaded, at run time, to all processors. The system of processors is arranged into a 2-D grid (see Fig.4.3).

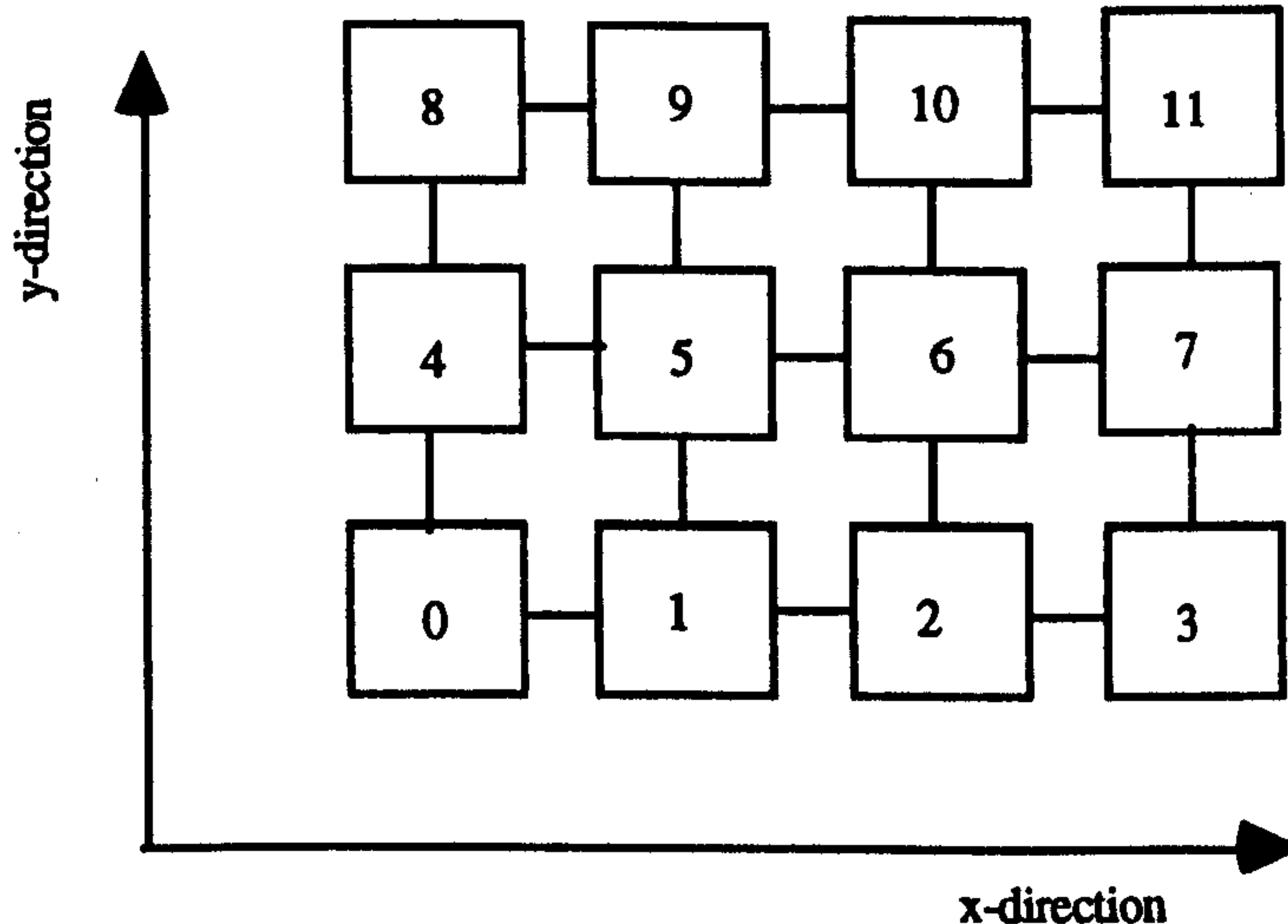


Fig. 4.3 The physical network structure of a 2-D configuration for a 4 by 3 network. The numbers shown are the identification position of processors in term of the 2-D grid.

Through this arrangement, and using an identification parameter for each processor with respect to position in the 2-D grid, different or similar instruction mechanisms by all processors can be achieved, where conditional FORTRAN statements are used to execute different instructions. To get information about the processors' network topology, Table 4.1 shows some of the subroutines that could be called, in the main program, to obtain full

information about the system such as the number of processors used, the co-ordinates in terms of the 2-D grid and the total dimension in the x- and y-directions.

Instructions	Definitions
ID=MYPROCID()	Determines the processor identification position in terms of the 2-D grid.
NP=NPROCS()	Determines the total number of processors used.
CALL GETMYPOS(MYPX, MYPY, MYPZ)	Determines the co-ordinates of the processor in terms of 2-D grid (MYPZ=1, not used for the super-cluster).
CALL GETDIM(DIMX, DIMY, DIMZ)	Determines the physical dimensions of the 2-D grid (DIMZ=1, not used for the super-cluster).

Table 4.1 Functions under PARIX to determine information about the system in use.

One of the main features of PARIX is that the number of processors can be decided at run time. This feature gives the freedom to write only a single computer code where the number of processors and the dimensions in both x and y-directions are variables in the main program. Consequently, the same program can run on different number of processors without the need to alter the computer code. This feature is very important and desirable in MIMD environments from two points of view. The first is that a multi-user environment for the transputer array does not guarantee the availability of the required number of processors. The second is that it gives a flexible parameter to the application to be tested and applied using different configurations and arrangements without any new programming effort. The second step in programming under PARIX is to decide the kind of topology that connects the processors by which information is transferred from one processor to another. This step depends entirely on the application. A specific virtual link from one processor to another could be made explicitly and/or a whole topology (e.g. ring, farm, 2-D grid, pipe) could be built using a library call to one of the subroutines in Table 4.2. Finally, after establishing a topology or links between processors then exchange of data from one processor to another is possible through *send* and *receive* commands (see Table 4.3).

Instructions	Definitions
IDTOP=NEWTOP (NL)	Makes a topology with NL links, where IDTOP is a reference integer to this topology.
LN=ADDNEWLINK (IDTOP, IDL, IREQ)	Add a new explicit link between the processor executing the command to the processor of ID=IDL, IREQ is any integer and LN is the link reference integer.
IDTOP=MAKERING (IREQ, NP, XMIN, XMAX, YMIN, YMAX, ZMIN, ZMAX, ID, LINK())	Makes a ring between NP processors using processors between XMIN to XMAX and YMIN to YMAX and LINK contains the link reference number. ZMIN and ZMAX not used and should be made -1.
IDTOP=MAKE2DGRID (IREQ, IDIMX, IDIMY, XMIN, XMAX, YMIN, YMAX, ZMIN, ZMAX, MYPX, MYPY, LINK())	Makes a 2-D grid topology between processors (other details see above)

Table 4.2 Some intrinsic functions under PARIX to establish virtual links and/or whole topologies for processors. (There are similar functions for making FARM and PIPE topologies).

Instructions	Definitions
CALL SEND (IDTOP, IDR, V, L)	Sends a variable V with length L to a processor with ID=IDR.
CALL RECV (IDTOP, IDS, V, L)	Receives a variable V with length L from a processor with ID=IDS.

Table 4.3 Functions under PARIX for sending and receiving information between processors.

4.3.2 Programming under CM-FORTRAN [10-13]

CM-FORTRAN is an extension of FORTRAN 77 with the addition of array processing operations from FORTRAN 90. It combines the familiarity of FORTRAN 77, being one of the best language for scientific computing, the power of FORTRAN 90 which is a rich language for manipulating arrays, and the power of the CM as a fast computational tool that can handle large arrays using one instruction mechanism. In FORTRAN 77, operations on arrays require the usual explicit stepping through all elements to execute scalar operations. For example, if A is a three-dimensional array, then updating A by a scalar requires the following

```

DO 10 I1=1,N1
DO 20 I2=1,N2
DO 30 I3=1,N3
      A(I1, I2, I3)=A(I1, I2, I3)+10
30 CONTINUE
20 CONTINUE
10 CONTINUE

```

On the other hand, FORTRAN 90 treats arrays as first class objects where it is not necessary to reference array elements separately but the whole array may be referenced in the operation which means referring to all elements as one block. In FORTRAN 90, updating the above A array would simply be

$$A=A+10$$

From the above, FORTRAN 77 evaluates the updating of A in the order specified by the DO loops, but FORTRAN 90 allows the operations on A to be in any order including concurrent operations. Therefore in CM-FORTRAN, the (A=A+10) operation is treated as a parallel operation where all the elements of A are operated on by different processors. Unlike the transputer array (as mentioned before), synchronisation and load balancing in the CM do not need any special arrangements because they are enforced by the design of the machine. In CM-FORTRAN, there are many intrinsic functions (e.g. FORALL, WHERE, etc.) that enable operations, on arrays or particular sections of arrays, to be executed or manipulated concurrently. These functions use the mechanism of one instruction operating on multiple data, where the instruction is fully synchronised at all processors. One feature of the CM-FORTRAN is that serial and parallel computations can be handled in the same computer code, and the declaration of arrays can take either of these two modes (serial or parallel). In other words, arrays can have their *home* in the front end computer memory or the CM machine distributed memory in which computations in the front end computer are executed serially and computations in the CM machine are executed in parallel (see Fig. 4.4).

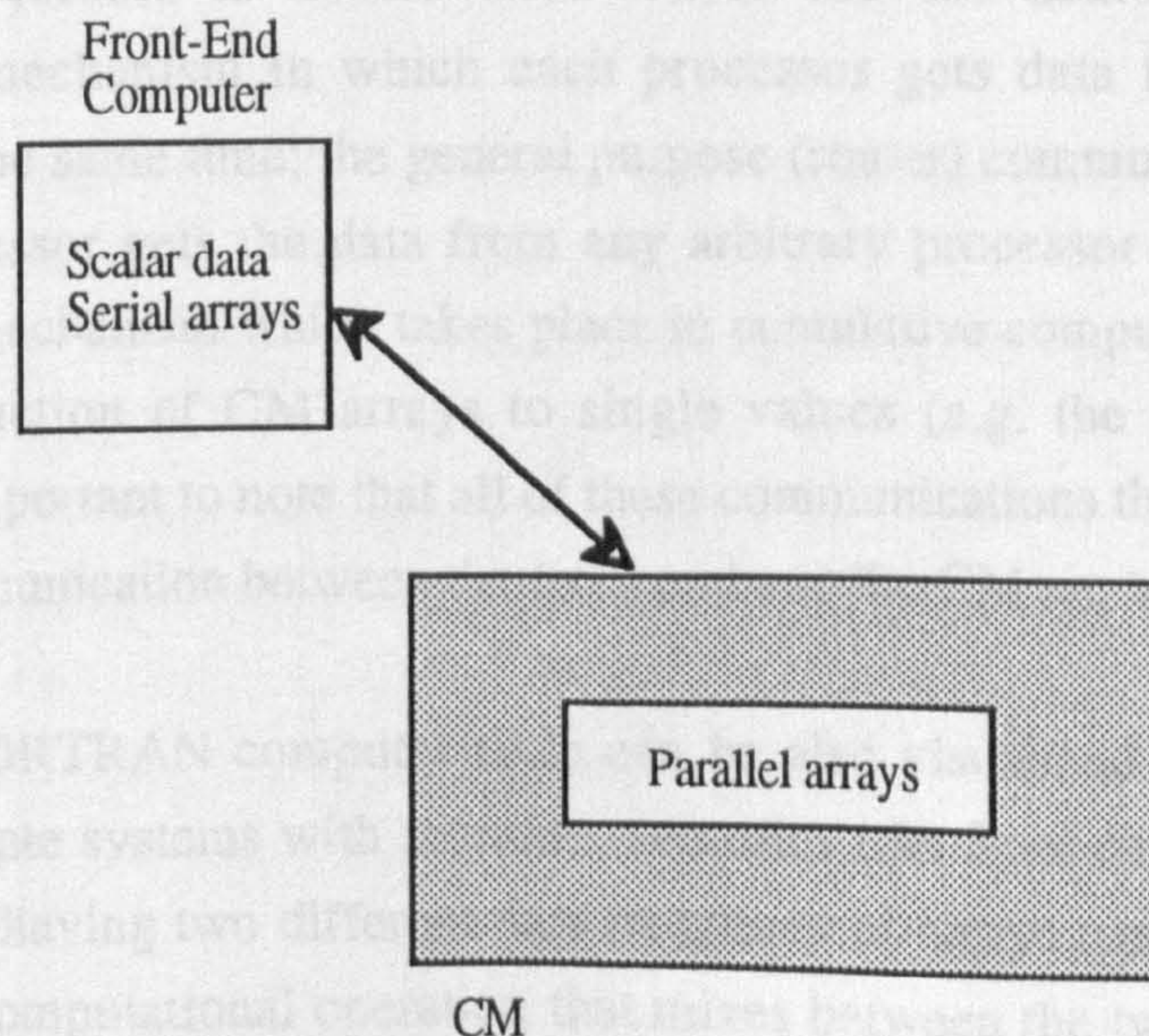


Fig. 4.4 The distribution of data between the front end and the CM

Both the front end computer and the CM machine work as one unit, but the front end runs all the commands of the CM-FORTRAN code (serial and parallel), thereby executing all serial or scalar operations locally and sending all parallel operations to the CM for execution. All the data of a parallel array are distributed to all CM processors' memory, one element per processor. However, large problems may have data sets that are larger than the number of processors. In such a situation, the system will create a *virtual processing* mechanism which means that each physical processor simulates a number of virtual processors by dividing its local memory in order to accommodate all assigned data. The CM receives the instruction from the front end, and distributes it to all processors. Then the instruction is executed simultaneously by all processors on their local data. In the case that a processor has many virtual processors, the instruction is looped serially around all local data for execution. This means that the more virtual processors exist the slower the code runs. Due to the synchronisation of instructions, no processor is allowed to execute the next instruction until all processors have finished the current one. In the situation where processors are excluded from a certain computation, due to a conditional instruction, they remain idle until they are reactivated by another instruction. A conditional instruction means that the operation is intended for part of the CM array (e.g. summing a part of an array).

Since the memory of the CM is distributed between processors, there are some flexible mechanisms that enable processors to access each others memories, where each access mechanism depends on the function used to get the data. There are three methods for the CM processors to access data. These are the nearest neighbour (news) communication mechanism in which each processor gets data from its neighbouring processor, all at the same time; the general purpose (router) communication mechanism in which each processor gets the data from any arbitrary processor and finally the global communication mechanism which takes place in cumulative computations along grid axis of arrays or reduction of CM arrays to single values (e.g. the total sum of an array $SUM(A)$). It is important to note that all of these communications that take place inside the CM and the communication between the front end and the CM are transparent to the user.

The CM-FORTRAN computer code can be also visualised as a program that can control two separate systems with separate memories (the front end and the CM). Due to this flexibility of having two different data structures of arrays (serial and parallel) in the same code, any computational operation that mixes between the two array structures will cause the operation to be executed serially in the front end. Consequently, the CM array

will be moved element by element to the front end computer to perform the operation, resulting in a large reduction in the efficiency of the program. To show this home mixing explicitly, suppose that **A** is a CM array; then the following statement

A=A+10

will be executed in the CM and the following statement

```
DO 10 I1=1,N1
DO 20 I2=1,N2
DO 30 I3=1,N3
      A(I1, I2, I3)=A(I1, I2, I3)+10
30 CONTINUE
20 CONTINUE
10 CONTINUE
```

will be executed in the front end despite the fact that the array **A** is declared as a CM array and its elements are distributed among the processors of the CM. Such an operation must be avoided when writing a CM-FORTRAN computer code; however, in the event that the algorithm demands a mixed home operation, then it is recommended to use CM-FORTRAN utility routines to transfer the whole array from one machine to the other. On the other hand, the above operation of mixing scalar with CM arrays (**A=A+10**) results in parallel operation where the scalar element, which is residing in the front end, is broadcast to all processors for executing the operation .

REFERENCES

- [1] Bertsekas, D. and Tsitsiklis, J. "Parallel and Distributed Computation, Numerical Methods", Prentice Hall, New Jersey, 1989.
- [2] Lewis, T. and El-Rewini, H., "Introduction to Parallel Computing", Printce Hall Int., New Jersey, 1992.
- [3] Fleming, P. , "Parallel Processing in Control : The Transputer and Other Architecture's", Peter Peregrinus Ltd. UK, 1988.
- [4] Carriero, N. Gelernter, D., "How to Write Parallel Programs: A first Course", The MIT Press, Cambrige, Mass., 1990.
- [5] Lipovski, G. J. and Malek, M., "Parallel Computing", John Wiley & Sons, New York, 1987.
- [6] Quinn, M. J., "Designing Efficient Algorithms for Parallel Computers", McGraw-Hill Inc., New York, 1987.

- [7] Modi, J. J., "Parallel Algorithms and Matrix Computation", Oxford Series Book, 1988.
- [8] PARIX User Guide, Release 1.2, Parsytec Computer GmbH, Germany, 1992.
- [9] PARIX Reference Manual, Release 1.2, Parsytec Computer GmbH, Germany, 1992.
- [10] Getting Started in CM FORTRAN , Thinking Machines Corporation, Cambridge, MA, 1991.
- [11] CM-FORTRAN User's Guide, Thinking Machines Corporation, Cambridge, MA, 1991.
- [12] CM-FORTRAN Programming Guide, Thinking Machines Corporation, Cambridge, MA, 1991.
- [13] CM-FORTRAN Reference Manual, Thinking Machines Corporation, Cambridge, MA, 1991.
- [14] Richardson, J., "Visualizing Quantum Scattering on the CM-2 Supercomputer", *Comp. Phys. Comm.*, **63**, 84-94, 1991.
- [15] Richardson, J., "Computational Physics on the CM-2 Supercomputer", *Phys. Rep.*, **207**, 305-320, 1991.

Part II

**Implementations
(Analysis and assessments)**

Chapter 5

Two-Dimensional Beam Propagation Methods

5.1 Introduction

As a first step toward the implementation of the 3-D finite-difference Beam Propagation Methods discussed in chapter 3, this chapter shows the implementations of the 2-D FD-BPMs. Although the aim of this thesis is to model three-dimensional rectangular optical waveguides, the 2-D analysis of the BPM is very useful because it is much simpler to implement than the 3-D BPM, and also the results of the 2-D BPM can be compared, in terms of accuracy, with closed form analytical solutions of z-invariant 2-D waveguide structures. In addition, this chapter can be used as a guide for simulation of 2-D waveguides and the results of the accuracy analysis can be used as a reference or as an indicator for the 3-D implementations. On the other hand, a close look at the literature of the BPM theory shows that the analysis and the assessment of the CN-BPM (ADI) are well established for 2-D [1-9] and 3-D [10-12] models with comparisons with the FFT-BPM. The RS-BPM and EFD-BPM, however, are only implemented and tested using the 3-D models [12-17]. In addition, there is no comparison between the performance of all the three FD-BPMs in 2-D or 3-D in terms of accuracy and efficiency.

This chapter consists mainly of three parts in addition to a conclusion at the end of the chapter. The first part, section 5.2, contains the 2-D formulation of the three FD-BPMs (CN, RS and EFD) of chapter 3. In this section, also the details of implementing the FD-BPMs numerically are given. For the 2-D algorithm of the CN-BPM which requires the inversion of a tridiagonal matrix for each propagation step, a more efficient recursive numerical technique that solves a tridiagonal system of equations avoiding the process of storing and inverting the matrix has been discussed numerically. The second part, section 5.3, contains the results of validating the three 2-D FD-BPM computer codes. As an initial experiment to test the mechanism of the FD-BPMs, the propagation of different input

fields (Gaussian and guided modes) in homogenous and 2-D structure has been considered. The second step involves detailed tests on the influence of the discretisation parameters of the FD-BPMs, and on the accuracy of the fields and the propagation constants of guided modes inside a z-invariant slab waveguide. The spectral method of chapter 3 is used to compute the propagation constants of the guided modes from the BPM fields after testing the spectral method itself using a simple example. These results are then compared with the analytical solutions of the slab waveguide. In addition, comparisons between the performance of all the three FD-BPMs are discussed at the end of the section. Section 5.4 contains a discussion on a modification made to the EFD-BPM to improve the stability constraints on the propagational step size. Although the stability of the EFD-BPM is improved by this modification it adds a spurious field to the true solution. The third part, in section 5.5, contains the application of the CN-BPM to devices that have some practical value, like directional couplers and Y-junctions.

5.2 Two-Dimensional BPM Implementations

In this section, the three-dimensional finite-difference BPMs of chapter 3 are reduced to 2-D and numerically prepared to analyse 2-D waveguides.

5.2.1 Formulation

The treatment of two dimensional waveguides, using the BPM, assumes that one of the transverse variations of the three-dimensional is zero. Let us assume the TE polarisation case (see the analysis of slab waveguides of chapter 2) where we may set the y-variation to zero and rewrite the parabolic equation in (3.5) as

$$2jk_0n_0 \frac{\partial \phi}{\partial z} = \frac{\partial^2 \phi}{\partial x^2} + k_0^2(n^2 - n_0^2)\phi \quad (5.1)$$

In the above equation, it has been assumed that the refractive index is a function of x and z ($n(x,z)$). We may now write the three 2-D FD-BPM solutions of Eq. 5.1 based on the derivation of chapter 3.

5.2.1.1 The Crank Nicholson method (CN-BPM)

Crank-Nicholson is always referred to as the two dimensional version of the ADI [18]. Combining Eqs. 3.11, 3.12 and 3.14 for two-dimensions leads to

$$\phi(x, z + \Delta z) = e^{(-j\Delta z G)} \phi(x, z) = \frac{\left[1 - j\frac{\Delta z}{2} G\right]}{\left[1 + j\frac{\Delta z}{2} G\right]} \phi(x, z) \quad (5.2)$$

where

$$G = \frac{1}{2k_0 n_0} \left[\frac{\partial^2}{\partial x^2} + k_0^2 (n^2 - n_0^2) \right]$$

Rewriting (5.2) as

$$\left\{ 1 + j\frac{\Delta z}{2a_0} \left[\frac{\partial^2}{\partial x^2} + k_0^2 (n^2 - n_0^2) \right] \right\} \phi(x, z + \Delta z) = \left\{ 1 - j\frac{\Delta z}{2a_0} \left[\frac{\partial^2}{\partial x^2} + k_0^2 (n^2 - n_0^2) \right] \right\} \phi(x, z) \quad (5.3)$$

Note that in Eq. 5.2 the refractive index term is combined with the spatial variation. Discretising the electric field ϕ and the refractive index n in the x -direction, with equal mesh spacing Δx , where we apply the second order central finite-difference approximation to the partial second derivative with respect to x (see Eq. 3.27), then Eq. 5.3 can be written to second order accuracy with respect to z and x as

$$\begin{aligned} \phi_i^{z+\Delta z} + j\frac{\Delta z}{2a_0} \left[\frac{\phi_{i+1}^{z+\Delta z} - 2\phi_i^{z+\Delta z} + \phi_{i-1}^{z+\Delta z}}{\Delta x^2} + k_0^2 (n_i^2 - n_0^2) \phi_i^{z+\Delta z} \right] = \\ \phi_i^z - j\frac{\Delta z}{2a_0} \left[\frac{\phi_{i+1}^z - 2\phi_i^z + \phi_{i-1}^z}{\Delta x^2} + k_0^2 (n_i^2 - n_0^2) \phi_i^z \right] \end{aligned} \quad (5.4)$$

where i represent the discretisation of the field and the refractive index ($i = 1, 2, 3, \dots, M_x$). Simplifying Eq. 5.4 into a more convenient form of

$$\phi_{i+1}^{z+\Delta z} + h_i \phi_i^{z+\Delta z} + \phi_{i-1}^{z+\Delta z} = -\phi_{i+1}^z + g_i \phi_i^z - \phi_{i-1}^z \quad (5.5a)$$

or

$$\phi_{i+1}^{z+\Delta z} + h_i \phi_i^{z+\Delta z} + \phi_{i-1}^{z+\Delta z} = \Omega_i^z \quad (5.5b)$$

where

$$h_i = \left[\frac{2a_o \Delta x^2}{j\Delta z} - 2 + \Delta x^2 k_o^2 (n_i^2 - n_o^2) \right]$$

$$g_i = \left[\frac{2a_o \Delta x^2}{j\Delta z} + 2 - \Delta x^2 k_o^2 (n_i^2 - n_o^2) \right]$$

$$\Omega_i^z = -\phi_{i+1}^z + g_i \phi_i^z - \phi_{i-1}^z$$

We may notice from Eq. 5.5 that the right hand side Ω_i^z is a known quantity containing the initial field at z and the left hand side of the equation contains three unknown fields (at $i+1$, i and $i-1$). Actually, Eq. 5.5 is a tridiagonal matrix equation that can be expressed as [18]

$$\begin{bmatrix} h_1 & 1 & & & & & \\ 1 & h_2 & 1 & & & & \\ & 1 & \ddots & \ddots & & & \\ & & \ddots & \ddots & 1 & & \\ & & & 1 & h_{M_x-1} & 1 & \\ & & & & 1 & h_{M_x} & \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \vdots \\ \phi_{M_x-1} \\ \phi_{M_x} \end{bmatrix}^{z+\Delta z} = \begin{bmatrix} g_1 & -1 & & & & & \\ -1 & g_2 & -1 & & & & \\ & -1 & \ddots & \ddots & & & \\ & & \ddots & \ddots & -1 & & \\ & & & -1 & g_{M_x-1} & -1 & \\ & & & & -1 & g_{M_x} & \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \vdots \\ \phi_{M_x-1} \\ \phi_{M_x} \end{bmatrix}^z \quad (5.6)$$

In writing (5.6), zero boundary conditions for the field points at both edges have been assumed. The implementation of the CN method in (5.6) involves two steps; first the initial field ϕ_i^z has to be multiplied with the tridiagonal matrix on the right hand side of Eq. 5.6, and the second step is inverting the matrix on the left hand side of Eq. 5.6. This process is repeated several times to cover the entire length of the device. Instead of storing and inverting the left hand side matrix of equation 5.6 for each propagational step, a more efficient recursive technique that deals with tridiagonal system of equations can be used [3,18,19]. Let us consider Eq. 5.5b and assume that one of the unknown fields in (5.5b) can be written in terms of one of the other two unknowns [3,18,19]

$$\phi_i^{z+\Delta z} = R_{i+1} \phi_{i+1}^{z+\Delta z} + Q_{i+1} \quad (5.7)$$

where R_{i+1} and Q_{i+1} are unknown auxiliary functions. If Eq. 5.7 is written in terms of $i=i-1$

$$\phi_{i-1}^{z+\Delta z} = R_i \phi_i^{z+\Delta z} + Q_i \quad (5.8)$$

and the substitution of (5.8) into Eq. 5.5 will lead to

$$\phi_{i+1}^{z+\Delta z} + (h_i + R_i)\phi_i^{z+\Delta z} + Q_i = \Omega_i^z \quad (5.9)$$

We rearrange (5.9) as

$$\phi_i^{z+\Delta z} = -\frac{1}{(h_i + R_i)}\phi_{i+1}^{z+\Delta z} + \frac{(\Omega_i^z - Q_i)}{(h_i + R_i)} \quad (5.10)$$

and observe the similarity between Eq. 5.10 and Eq. 5.7; the comparison between these two equations leads to the relationships between the coefficients by equating their corresponding auxiliary functions

$$R_{i+1} = -\frac{1}{(h_i + R_i)} \Rightarrow R_i = -\frac{1}{R_{i+1}} - h_i \quad (5.11a)$$

$$Q_{i+1} = \frac{(\Omega_i^z - Q_i)}{(h_i + R_i)} \Rightarrow Q_i = \Omega_i^z + \frac{Q_{i+1}}{R_{i+1}} \quad (5.11b)$$

From the above equations, the i values of R and Q depend on the $i+1$ values of these coefficients where the coefficients h_i and Ω_i^z are known quantities. The above recursive relations for R_i and Q_i can be solved using the initial values for these two coefficients. These initial values can be obtained from the boundary conditions at the extreme ends of the computational window. For simplicity we may assume that the fields at the two boundaries are zero ($\phi_0 = \phi_{M_x+1} = 0$). Using $\phi_{M_x+1} = 0$ in Eq. 5.5b when $i = M_x$ gives

$$h_{M_x}\phi_{M_x}^{z+\Delta z} + \phi_{M_x-1}^{z+\Delta z} = \Omega_{M_x}^z \quad (5.12)$$

and comparing this with Eq. 5.7 for $i = M_x-1$

$$\phi_{M_x-1}^{z+\Delta z} = R_{M_x}\phi_{M_x}^{z+\Delta z} + Q_{M_x} \Rightarrow -R_{M_x}\phi_{M_x}^{z+\Delta z} + \phi_{M_x-1}^{z+\Delta z} = Q_{M_x} \quad (5.13)$$

leads to the initial values for the auxiliary coefficients which are

$$\left. \begin{aligned} R_{M_x} &= -h_{M_x} \\ Q_{M_x} &= \Omega_{M_x}^z \end{aligned} \right\} \quad (5.14)$$

The other values of R_i and Q_i can be computed from Eq. 5.11. Knowing the auxiliary coefficients, the field at $z+\Delta z$ can be computed using the recursive relation in (5.7) which can be written as

$$\phi_{i+1}^{z+\Delta z} = \frac{\phi_i^{z+\Delta z} - Q_{i+1}}{R_{i+1}} \quad (5.15)$$

The other boundary condition ($\phi_0 = 0$) can be used to start the computation in (5.15) where this equation is equal to (for $i = 0$)

$$\phi_1^{z+\Delta z} = \frac{-Q_1}{R_1} \quad (5.16)$$

Here is a summary for the use of the recursive method discussed before. Given an optical field which needs to be marched from z to $z+\Delta z$ through a given refractive index. The first step is to compute the Ω_i^z function in (5.5b) by multiplying the input field (ϕ_i^z) at z with the right hand side tridiagonal matrix in (5.6). Then the auxiliary functions R_i and Q_i can be computed using the recursive relations in (5.11a) and (5.11b) and starting with the initial values in (5.14). Note that the auxiliary functions' recursive equations start the computations in the direction of decreasing i , from the top end of the computational window, at $i = M_x$, to the bottom end at $i = 1$. Finally the field ($\phi_i^{z+\Delta z}$) at $z+\Delta z$ can be computed using the recursive relation in (5.15) and starting with the initial values in (5.16).

The main advantage of the above recursive method over the usual inversion of a tridiagonal matrix is that it requires far less storage memory of the computer [3,18,19]. From the discussion above we may notice for z -invariant refractive index the coefficients h_i , g_i and R_i can be computed and stored once during the whole calculation. This means that two numerical arrays Q_i and the field at $z+\Delta z$ need to be updated.

5.2.1.2 The Real Space method (RS-BPM)

From the three-dimensional formula (Eq. 3.26) of chapter 3 we may write the two-dimensional Real Space BPM representation as [20]

$$\phi(x, z + \Delta z) = e^{(ps_x^*)} e^{(qs_x^*)} e^{(ps_x^*)} e^{\left(-j \frac{\Delta z}{a} U\right)} e^{(ps_x^*)} e^{(qs_x^*)} e^{(ps_x^*)} \phi(x, z) + O((\Delta z)^3) \quad (5.17)$$

where

$$U(x, z) = k_0^2 [n^2(x, z) - n_0^2] - \frac{2}{\Delta x^2}$$

$$q = -j \frac{\Delta z}{2a_0} = 2p$$

The numerical application of Eq. 5.17 follows a similar pattern to the one discussed in chapter 3. Let us expand that discussion to show the practical implementation of Eq. 5.17 numerically. We may notice that the first three operators and the last three operators of Eq. 5.17 are exactly the same and obviously their implementations are also the same. The implementation of the middle operator ($\exp(-j\Delta z U/a_0)$) of Eq. 5.17 is straightforward since it involves a multiplication of the discretised component of this operator with its corresponding field mesh point (one to one multiplication of mesh points). On the other hand the three split operators that represent the transverse variation in Eq. 5.17 could be implemented on any discretised field F_i as [20]

$$e^{\left(-j\frac{\Delta z}{2a_0}\nabla_x^2\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix} = e^{\left(-j\frac{\Delta z}{4a_0}s_x^2\right)} e^{\left(-j\frac{\Delta z}{2a_0}s_x^2\right)} e^{\left(-j\frac{\Delta z}{4a_0}s_x^2\right)} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix} \quad (5.18)$$

The execution of the right-most operator on the right hand side of (5.18) could be done by replacing the operator with the analytical solution in (3.24). Then the computation of this operation reduces to the multiplication of every two neighbouring discrete field points with a 2 by 2 matrix as (see the derivation in chapter 3)

$$\begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{*} = \begin{bmatrix} \alpha_x & \beta_x \\ \beta_x & \alpha_x \end{bmatrix} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix} \quad (5.19)$$

with $(i = 0, 2, 4, \dots, M_x)$

where

$$\alpha_x = \cos \tau_x, \quad \beta_x = -j \sin \tau_x \quad \text{and} \quad \tau_x = \frac{\Delta z}{4a_0 \Delta x^2}$$

and M_x should be even. The same could be done for the middle operator of Eq. 5.18 where the operation can be written as

$$\begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s**} = \begin{bmatrix} \alpha_x & \beta_x \\ \beta_x & \alpha_x \end{bmatrix} \begin{bmatrix} F_i \\ F_{i+1} \end{bmatrix}^{s*} \quad (5.20)$$

with $(i = 1, 3, 5, \dots, M_x - 1)$

where

$$\alpha_x = \cos \tau_x, \quad \beta_x = -j \sin \tau_x \quad \text{and} \quad \tau_x = \frac{\Delta z}{2a_0 \Delta x^2}$$

Obviously the application of the left-most operator of Eq. 5.18 is exactly the same as the right-most one in (5.19). To summarize the implementations of the three previous steps, Fig. 5.1 shows the application of these operators on an even number of mesh points.

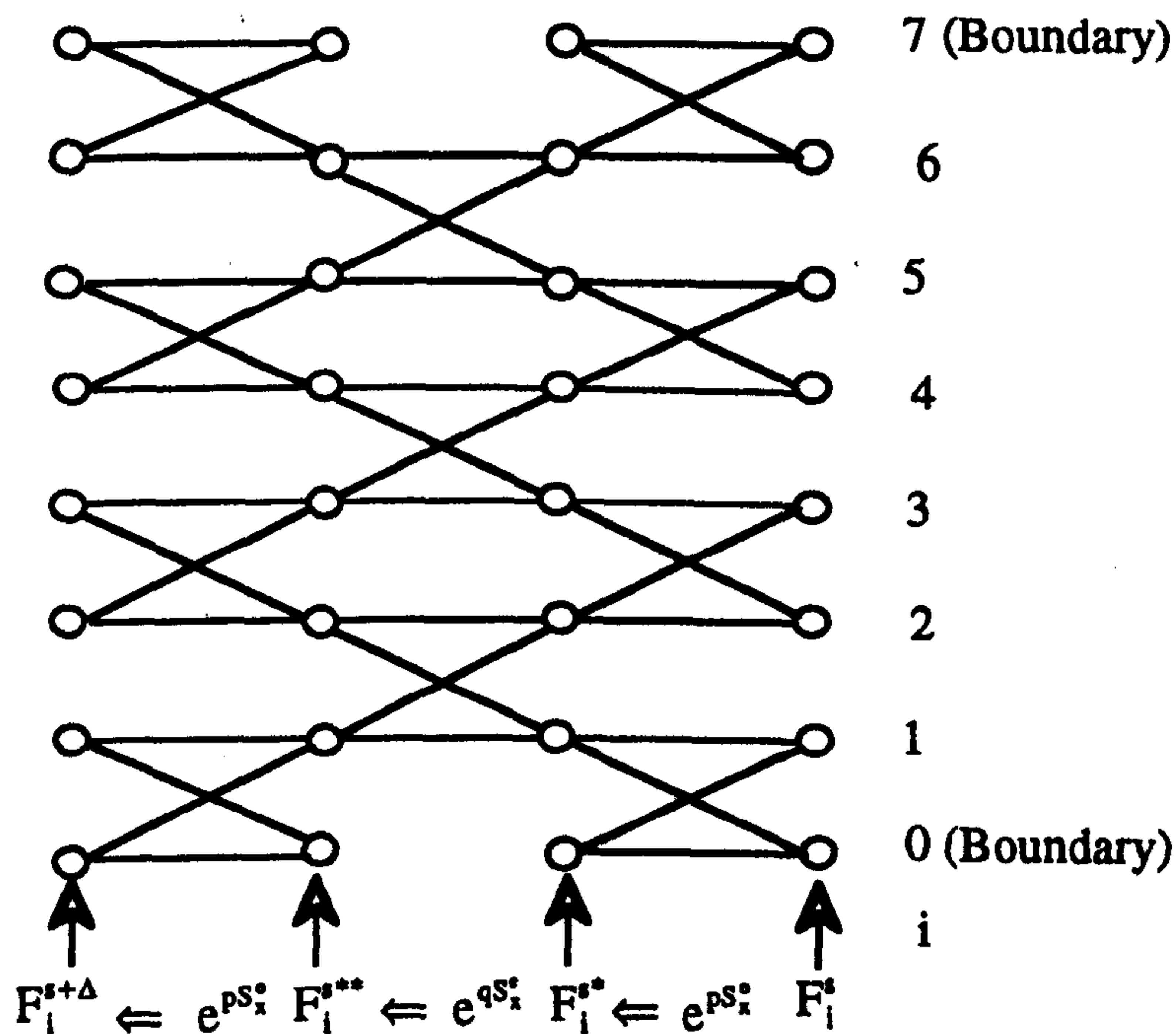


Fig. 5.1 A diagram showing the operation of the RS-BPM split operator of Eq. 5.18 on an even number of points (in this case the number of mesh points is equal to 6 and two boundary points). The circles represent field mesh points and the lines refer to the dependence of the computation of that field mesh point on the previous step field mesh points by multiplying with the coefficients in (5.19) and (5.20).

5.2.1.3 The Explicit Finite-Difference method (EFD-BPM)

From chapter 3 we may re-write the three-dimensional equation in (3.28) as a two-dimensional Explicit Finite-Difference BPM (EFD-BPM) [13-17]

$$\phi_i(z + \Delta z) = \phi_i(z - \Delta z) - \frac{2j\Delta z}{a_0 \Delta x^2} [\phi_{i-1}(z) + \phi_{i+1}(z)] - \frac{2j\Delta z}{a_0} U_i \phi_i(z) \quad (5.21)$$

Also the stability condition for the above equation can be written, from (3.29), as

$$\Delta z < 2k_0 n_0 \left[\frac{4}{\Delta x^2} + k_0^2 |n_{i,m}^2 - n_0^2|_{\max} \right]^{-1} \quad (5.22)$$

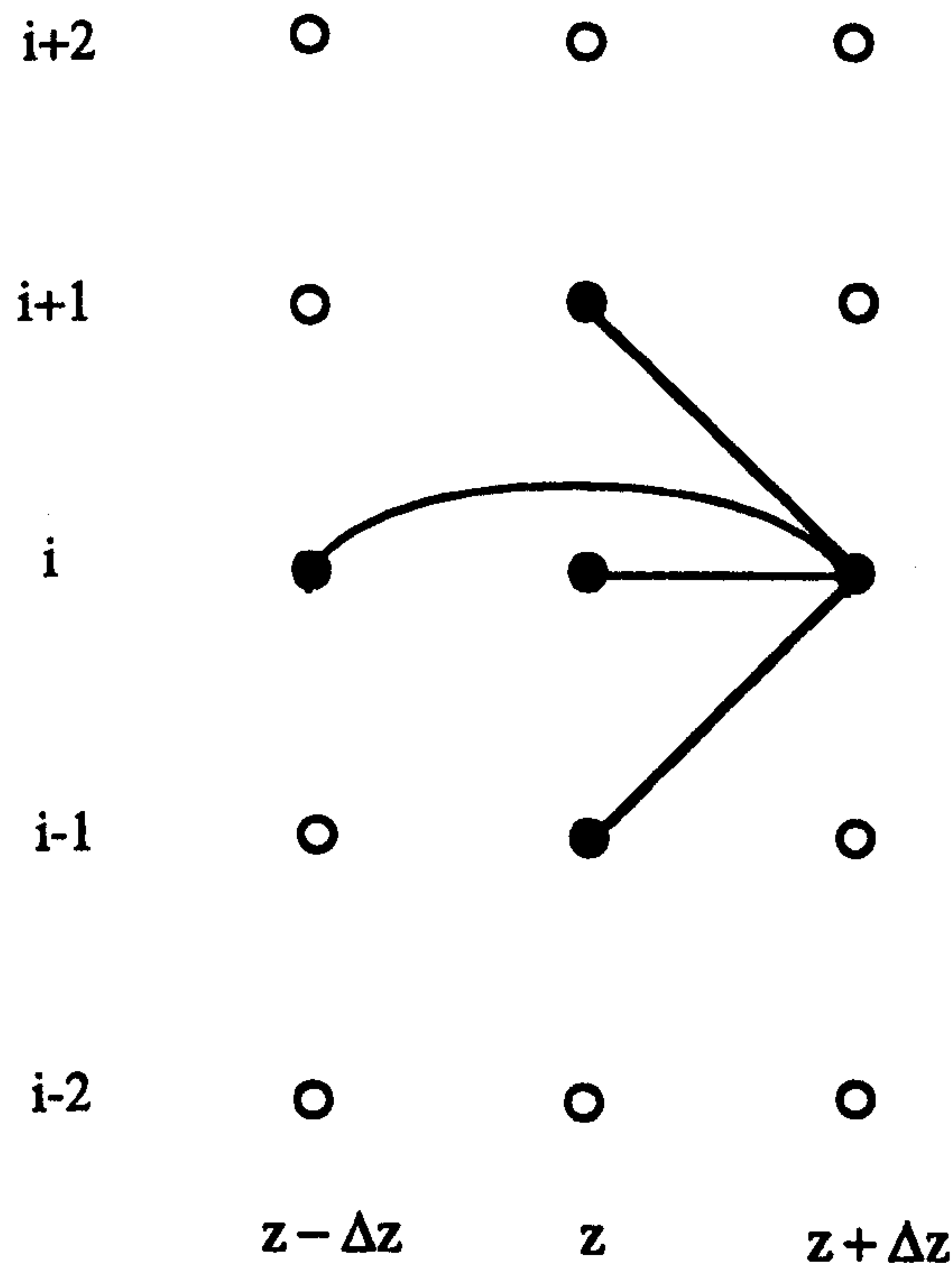


Fig. 5.2 A diagram showing the operation of the two dimensional EFD-BPM of Eq. 5.21. The circles represent field mesh points and the lines refer to the dependence of the computation of that field mesh point on the previous two steps field mesh points by multiplying with the coefficients in (5.21).

Fig 5.2 shows a simple diagram that describes the dependence of the *i*th mesh points on the previous two steps. The application of two-dimensional EFD-BPM numerically is very simple compared to the CN-BPM and the RS-BPM. During the propagation of the EFD-BPM field, three 1-D numerical arrays are needed; two to store the field at *z* and *z-Δz* and the third to compute the field at *z+Δz*. We may notice from the EFD-BPM that two fields (at *z-Δz* and *z*) are required to start the algorithm. One may use any other BPM (FFT, CN or RS) to find the field at *z* given an initial field at *z-Δz* and then continue the rest of the propagation using the EFD-BPM. A second approach is to calculate the first step of the EFD-BPM using a forward difference approximation instead of the central difference of the EFD-BPM of Eq. 3.27, which results in an algorithm similar to the EFD-BPM [13,14]. On the other hand, we may assume that the field does not change over the first step where the refractive index profile can be assumed to be independent of *z* and the propagational

step distance is very small; then the fields at z and $z - \Delta z$ may be assumed to be equal to the initial field. For simplicity, the last approach has been adopted in the implementations of the EFD-BPM.

5.3 Analysis and Assessments

All the three Finite-Difference BPMs (CN, RS and EFD) discussed in the last section have been programmed using FORTRAN 77 with double precision accuracy. This section shows the results of applying and testing the FD-BPMs using 2-D waveguides. In addition, comparisons, in terms of accuracy and efficiency, between all these methods are discussed. The most important parameters that affect the overall performance of the BPM are the mesh spacing Δx , the propagational step Δz , the reference refractive index n_0 and the boundary conditions at the edges of the computational window. The choice of the reference refractive index n_0 should not break down the parabolic approximation made in (3.4) where it was assumed that the envelope is varying slowly. In the case of a single mode analysis, n_0 could be chosen as the effective index of the mode (or very close to it if the exact value is not known). For multi-mode waveguides, n_0 could be chosen as a representative (an average) value of the values of effective indices of the modes [5,21]. On the other hand, the choice of the boundary conditions is also important for implementing the BPM, particularly for waveguides that produce many scattering fields (e.g. Y-junction). If such scattered radiation is not properly absorbed at the edges of the computational window, they tend to reflect back to the region of interest causing an unwanted interference with guided modes. One way to absorb some of the radiation at the edges of the computational window is by placing an absorbing material at the boundary. The practical implementation of this absorbing material is to add an imaginary part to the refractive index profile near the boundary where the choice of this imaginary value and the thickness of the absorbing material need to be investigated properly for every problem [2]. Another method is to use what is called transparent boundary conditions (TBC) [22, 23]. The TBC depends on the idea of allowing the travelling waves toward the boundary to pass through. For example, the application of the TBC at one of the boundary (e.g. at $x = M_x \Delta x$) requires that the field should satisfy the following [22,23]

$$\frac{\partial \phi}{\partial x} = -jk_x \phi \quad (5.23)$$

or could be written in a discretised form as [22, 23]

$$\phi_{M_x}^{z+\Delta z} = \phi_{M_x-1}^{z+\Delta z} e^{-jk_x \Delta x} \quad (5.24)$$

where the complex wave vector k_x is computed from the previous step by evaluating the ratio $(\phi_{M_x-1}^z / \phi_{M_x-2}^z)$. In order to ensure that only radiation outflow occurs from the computational window, the real part of k_x in Eq. 5.24 must be restricted to positive values. The TBC has been applied successfully using the CN-BPM [22, 23]. However in some circumstances it produces unwanted numerical noises [23]. On the other hand, tests of applying the TBC to the RS-BPM showed that the TBC breaks the unitarity of the split operators of the RS-BPM after a few propagation steps. For simplicity and to focus on the influence of the numerical discretisation parameters in the following analysis, zero boundary conditions at the edges of the computational window have been used for both the CN-BPM and EFD-BPM methods and periodic boundary conditions have been used for the RS-BPM. The periodic BC for the RS-BPM is to ensure that the split operator of the method remains unitary. These BC choices are good enough if the boundaries are not influencing the propagated field. This can be avoided by setting the boundaries far away from the region of interest (e.g. the waveguides regions). The following assessment for the FD-BPMs consists of three parts; first the mechanism of propagating different initial fields (Gaussian and guided modes) in homogenous and 2-D structure is tested. The second part involves detailed tests on the influence of the discretisation parameters (Δx and Δz). The third part (in section 5.5) is the application of the FD-BPM to some practical devices (directional coupler and Y-junction). In addition, the performance of each of the three FD-BPMs is also compared.

The first numerical test involves the propagation of a Gaussian field distribution in a homogenous medium. The CN-BPM has been used to propagate a Gaussian field with an input half beam width of $w_0 = 2.0 \mu\text{m}$ and a wavelength of $\lambda = 1.0 \mu\text{m}$ inside a homogenous medium that has a refractive index of $n = 1.0$ (air). Fig. 5.3 shows the evolution of the 2-D Gaussian beam, using the CN-BPM, in the air to a distance of $50 \mu\text{m}$ with $\Delta x = 0.05 \mu\text{m}$, $\Delta z = 0.1 \mu\text{m}$ and a total window size of $X = 50 \mu\text{m}$. The same results of Fig. 5.3 have been observed using the RS-BPM and the EFD-BPM¹. We may observe from the figure the spread of the beam during the propagation where the half beam width at $z = 50 \mu\text{m}$ is equal to $8.2 (w(50))$ which agrees with the analytical prediction when using the following exact expression for the computation of half beam width of a Gaussian spreading at any distance z [24]

¹ As a matter of convenience, all data/figures of similar nature will not be shown.

$$w(z) = w_0 \sqrt{1 + \left[z^2 / (\pi w_0^2 / \lambda)^2 \right]} \quad (5.25)$$

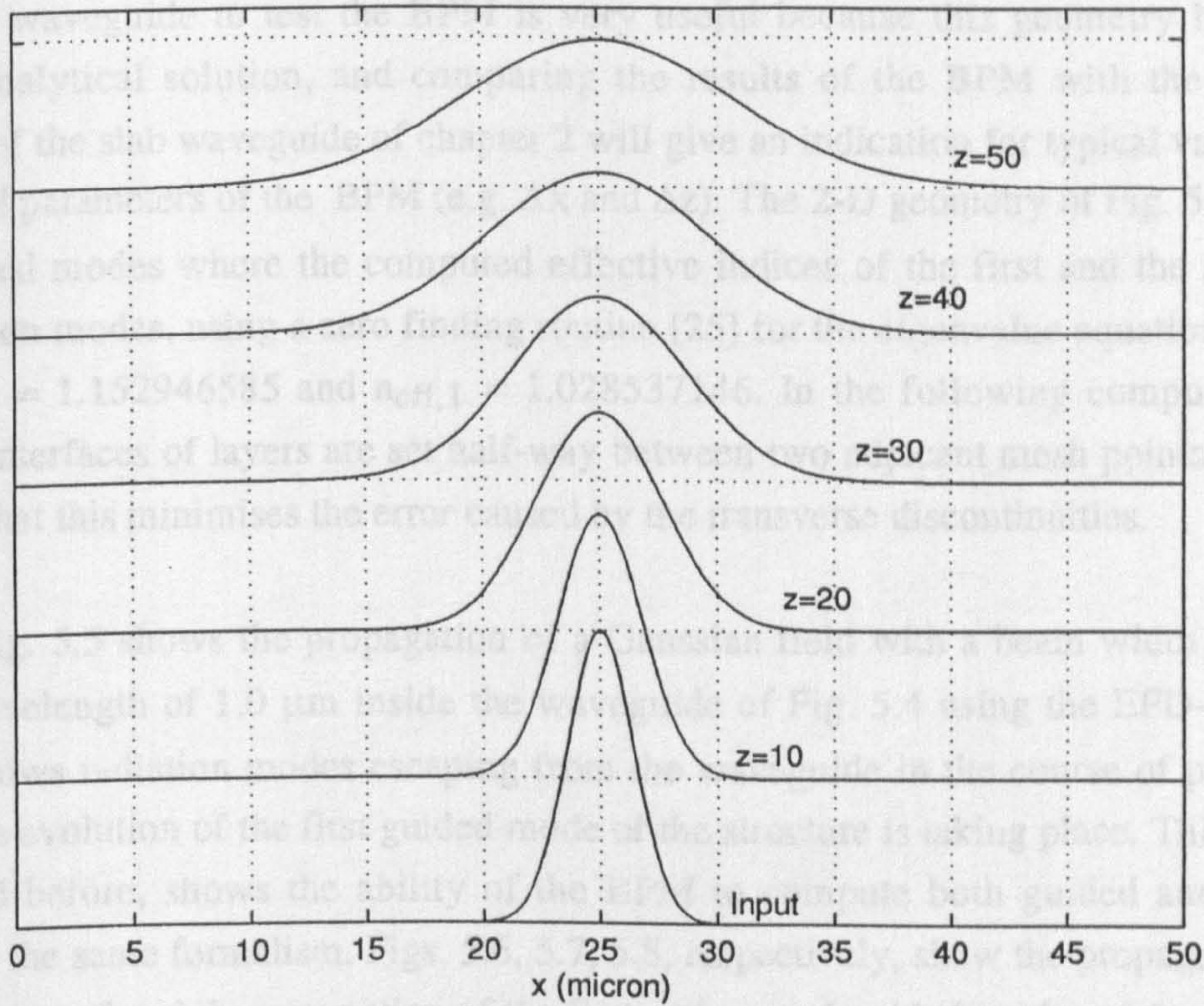


Fig. 5.3 The evolution of the amplitude of a 2-D Gaussian field in free space at several propagational distances using the CN-BPM with an input beam width of 4, $\Delta x = 0.05$, $\Delta z = 0.1$ and $\lambda = 1.0$ (all dimensions are in μm).

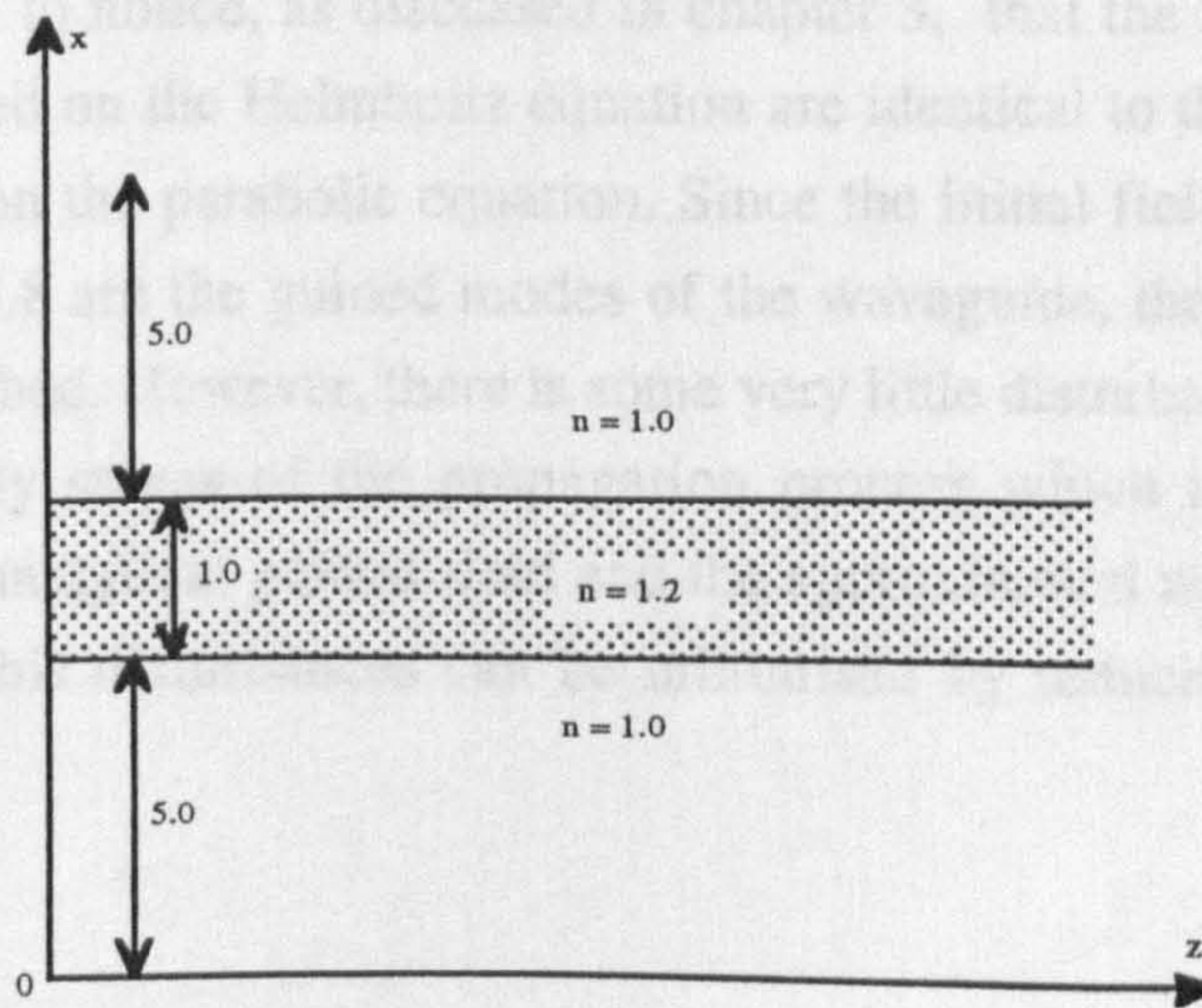


Fig 5.4 The symmetric slab waveguide used to validate the accuracy of the FD-BPMs with a wavelength $\lambda = 1.0$ (all dimensions are in μm).

The second numerical test for all FD-BPMs involves the propagation inside a symmetric slab waveguide of Fig. 5.4 with a wavelength of $\lambda = 1.0 \mu\text{m}$. The choice of z -invariant waveguide to test the BPM is very useful because this geometry has a well-known analytical solution, and comparing the results of the BPM with the analytical solution of the slab waveguide of chapter 2 will give an indication for typical values of the numerical parameters of the BPM (e.g. Δx and Δz). The 2-D geometry of Fig. 5.4 supports two guided modes where the computed effective indices of the first and the second TE polarisation modes, using a zero finding routine [25] for the eigenvalue equation in (2.24), are $n_{\text{eff},0} = 1.152946585$ and $n_{\text{eff},1} = 1.028537146$. In the following computations the internal interfaces of layers are set half-way between two adjacent mesh points [26]; tests showed that this minimises the error caused by the transverse discontinuities.

Fig. 5.5 shows the propagation of a Gaussian field with a beam width of $1.0 \mu\text{m}$ and a wavelength of $1.0 \mu\text{m}$ inside the waveguide of Fig. 5.4 using the EFD-BPM. The figure shows radiation modes escaping from the waveguide in the course of propagation where the evolution of the first guided mode of the structure is taking place. This result, as discussed before, shows the ability of the BPM to compute both guided and radiation modes in the same formalism. Figs. 5.6, 5.7, 5.8, respectively, show the propagation of the first, the second and the summation of the first and second guided modes of the waveguide of Fig. 5.4, using the CN-BPM, the RS-BPM and the EFD-BPM respectively (see the figures captions for details). The guided modes used as the initial input for these simulations are computed numerically from the slab waveguide theory of chapter 2 (see Eq. 2.22). We have to notice, as discussed in chapter 3, that the field distribution of Eq. 2.22 which are based on the Helmholtz equation are identical to their corresponding field distribution based on the parabolic equation. Since the initial fields of the simulations in Figs. 5.6, 5.7 and 5.8 are the guided modes of the waveguide, the fields propagate in the waveguide undisturbed. However, there is some very little disturbance (unnoticeable in the figures) at the early stages of the propagation process which is due to the mismatch between the exact analytical guided field and the approximated numerical guided field of the initial input. This disturbances can be minimised by reducing the transverse mesh spacing [5].

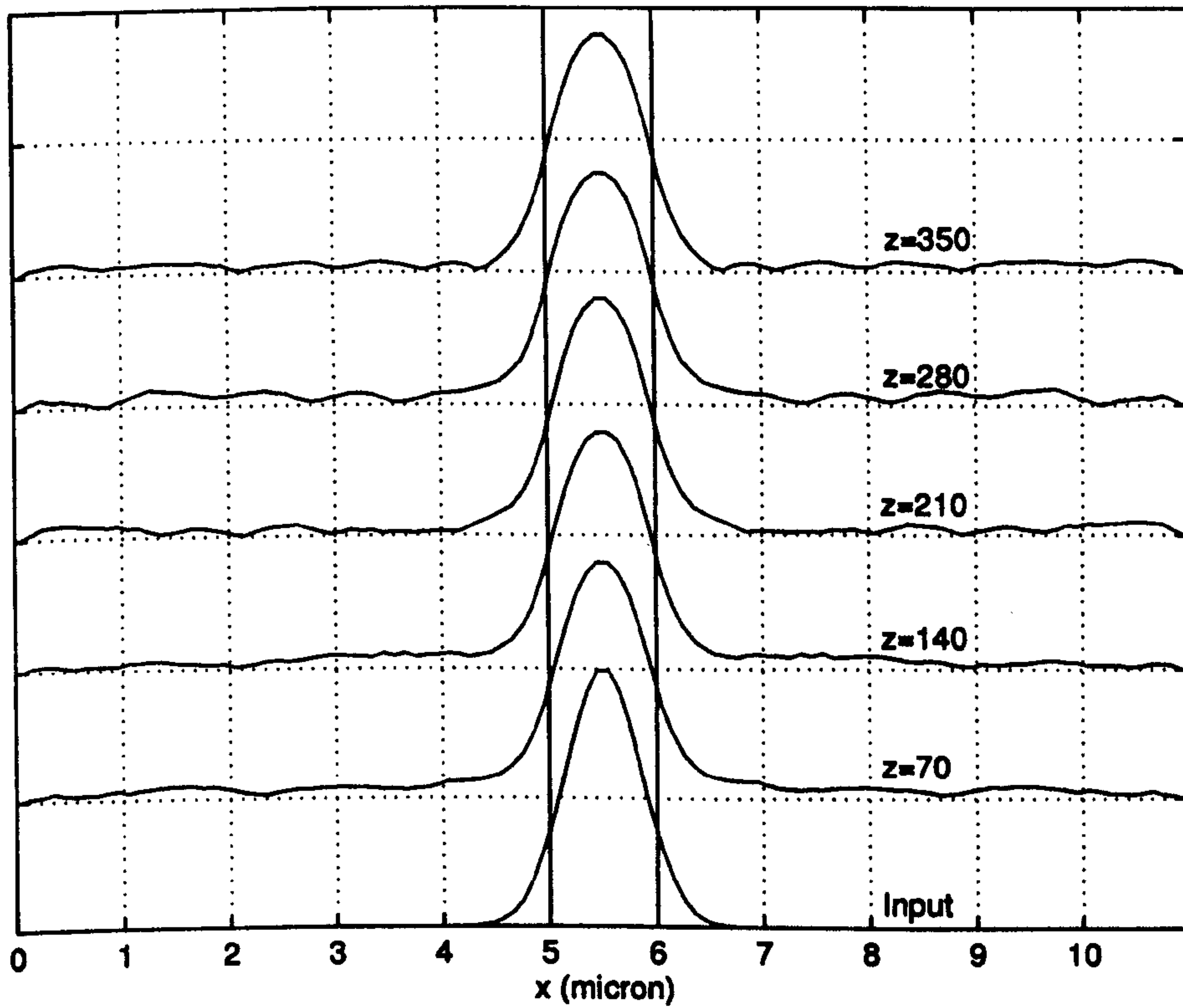


Fig. 5.5 The evolution of the amplitude of a 2-D Gaussian field inside the waveguide of Fig. 5.4 at several propagational distances using the EFD-BPM with an input beam width of 1.0, $\Delta x = 0.1$, $\Delta z = 0.025$ and $\lambda = 1.0$ (all dimensions are in μm).

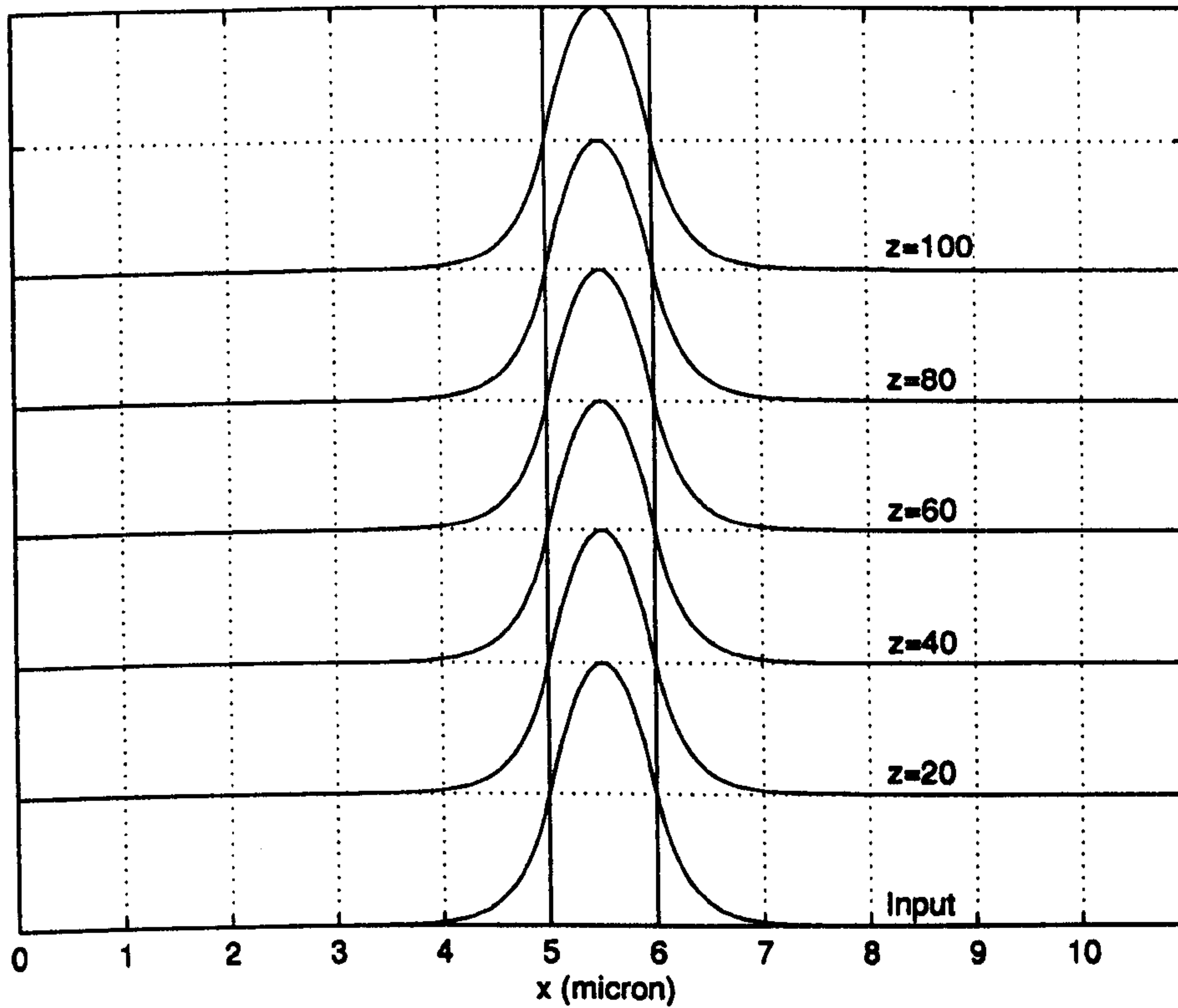


Fig. 5.6 The propagation of the first guided mode (amplitude) inside the waveguide of Fig. 5.4 at several propagational distances using the CN-BPM with $\Delta x = 0.05$, $\Delta z = 0.1$ and $\lambda = 1.0$ (all dimensions are in μm).

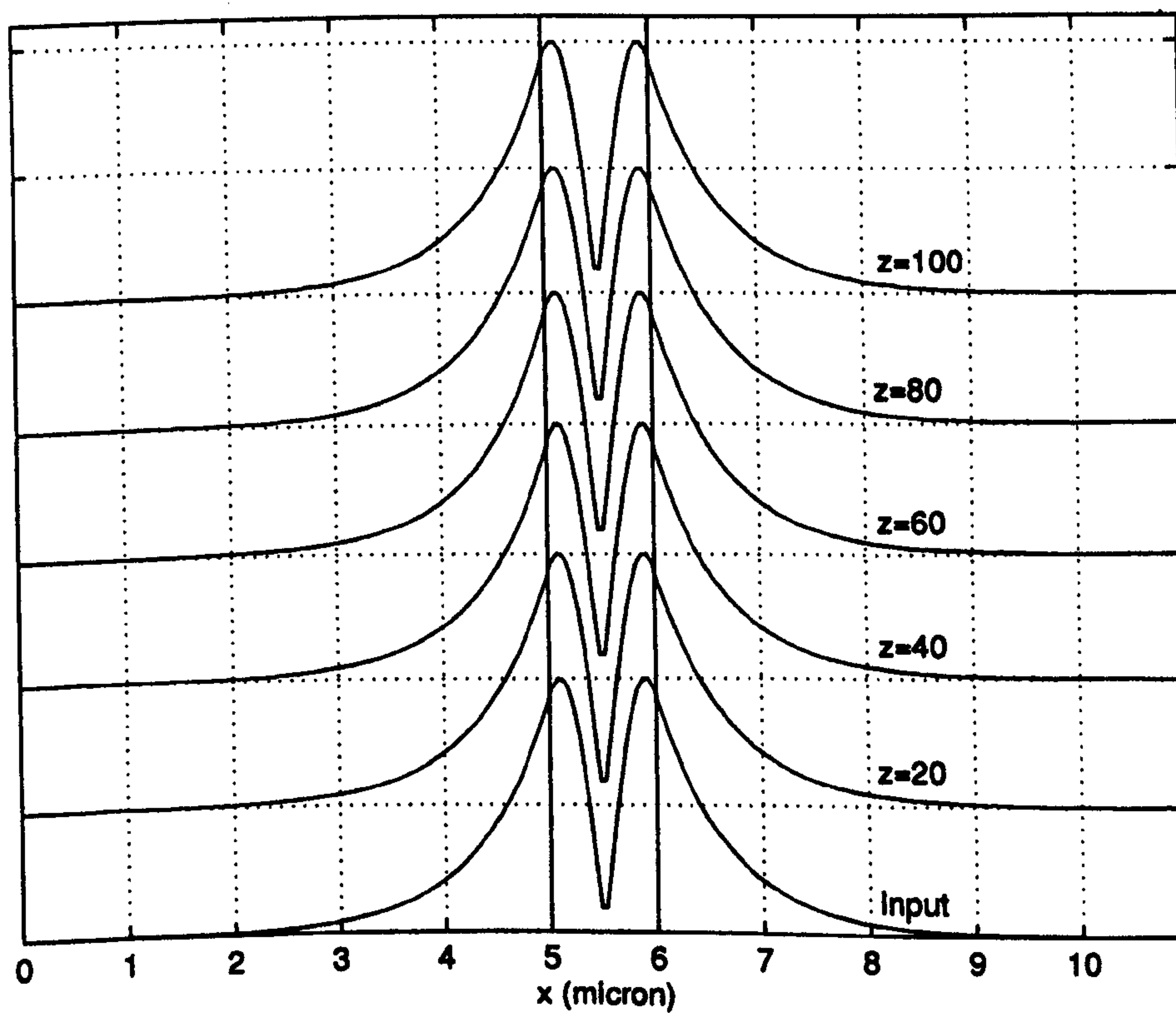


Fig. 5.7 The propagation of the second guided mode (amplitude) inside the waveguide of Fig. 5.4 at several propagational distances using the RS-BPM with $\Delta x = 0.05$, $\Delta z = 0.00625$ and $\lambda = 1.0$ (all dimensions are in μm).

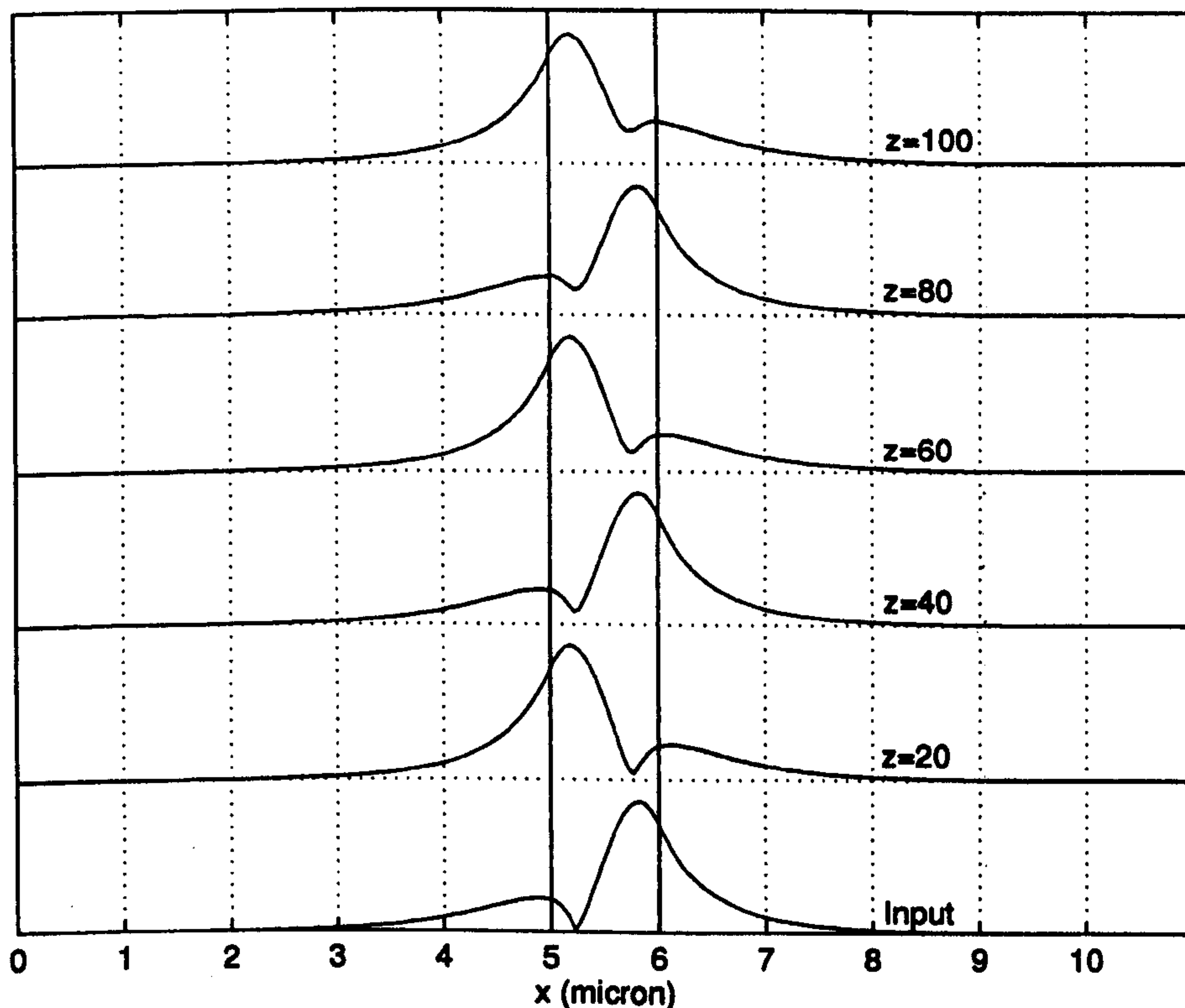


Fig. 5.8 The propagation of the sum of the first and the second guided modes (amplitude) inside the waveguide of Fig. 5.4 at several propagational distances using the EFD-BPM with $\Delta x = 0.05, \Delta z = 0.00625$ and $\lambda = 1.0$ (all dimensions are in μm).

Generally, it has been noticed from testing all FD-BPMs, and will be shown later in this section, that the accuracy increases by reducing the mesh spacing Δx . This agrees with the prediction of the numerical theory where the error reduces by increasing the mesh points (more sampled points). On the other hand, the three algorithms of the FD-BPMs converge differently by changing the longitudinal step size Δz . The accuracy of the CN-BPM is almost unaffected by the size of Δz , the accuracy of the EFD-BPM is also unaffected by the size of the Δz , provided that Δz is lower than the stability limit in (5.22) and the accuracy of the RS-BPM converges by reducing Δz . Fig. 5.9 shows the output RS-BPM fields after propagating the first guided mode of the waveguide of Fig. 5.4 to a distance of $z = 100 \mu\text{m}$ using different longitudinal step size Δz for each simulation. The figure shows that as Δz is reduced the field of the RS-BPM converges to the right shape. In order to show the influence of Δx on the accuracy of the FD-BPMs, the first guided mode of the waveguide of Fig. 5.4 was propagated, using all three FD-BPMs, to a distance of $z = 100 \mu\text{m}$ with a small fixed $\Delta z = 0.00625 \mu\text{m}$ and computing the absolute field difference between the analytical field and the output field which is defined as

$$\text{Error} = \left| |\phi|_{\text{analytical}} - |\phi|_{\text{BPM}} \right| \quad (5.26)$$

and for several values of Δx . Figs. 5.10, 5.11 and 5.12 show, respectively, the absolute field difference for all three FD-BPMs using different Δx of values equal to 0.25, 0.1 and 0.05. From the three figures, generally the field error decreases as the mesh spacing Δx is reduced where the largest error occurs near the interface between layers and this is due to the large difference between the guided refractive index and the substrate refractive index.

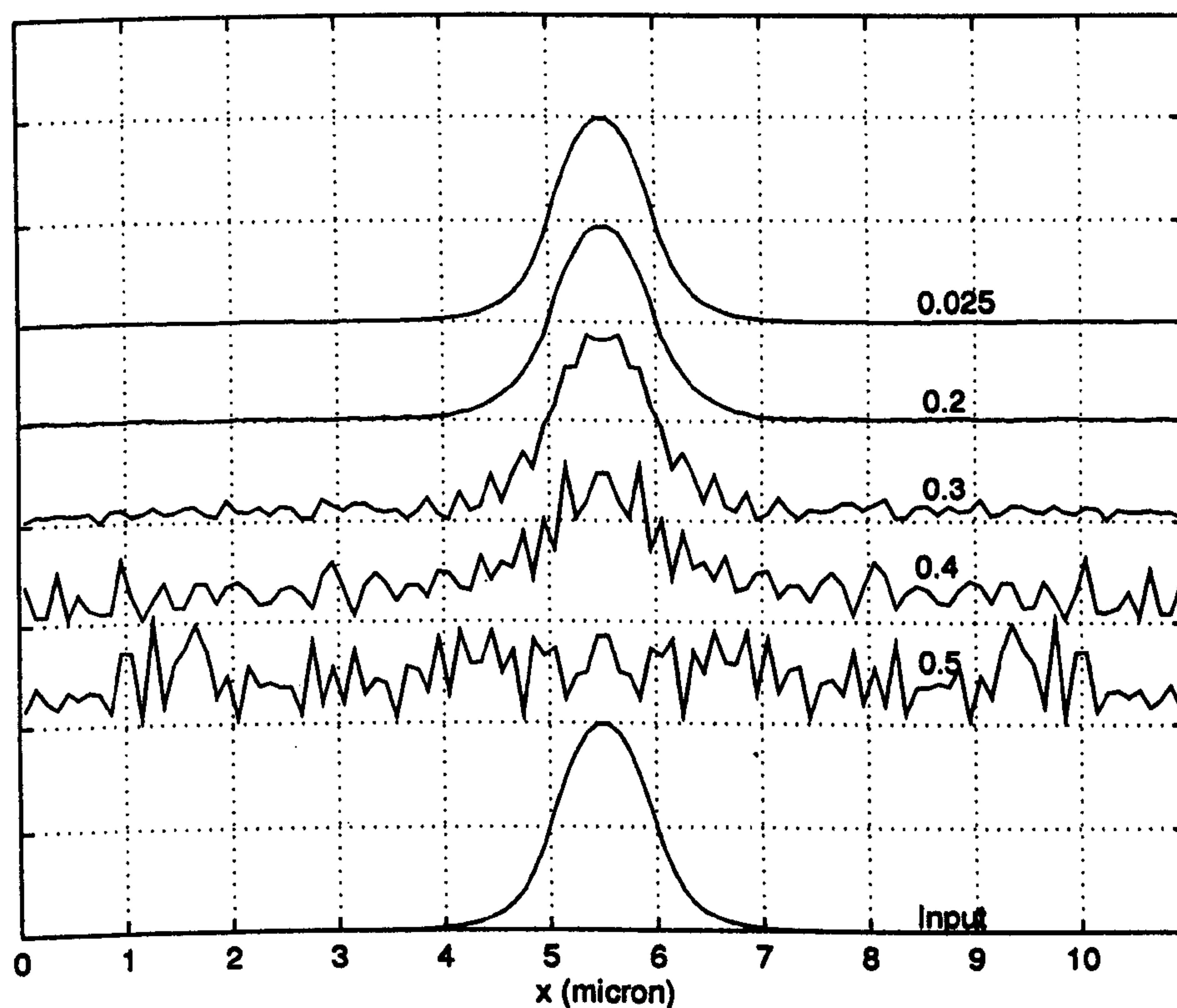


Fig. 5.9 The convergence of the RS-BPM field for the propagation of the first guided mode of the waveguide of Fig. 5.4 to a distance of 100 μm with $\Delta x = 0.1$ and using different longitudinal step size Δz for each simulation. The numbers on plots refer to the size of Δz used in the computation (all dimensions are in μm).

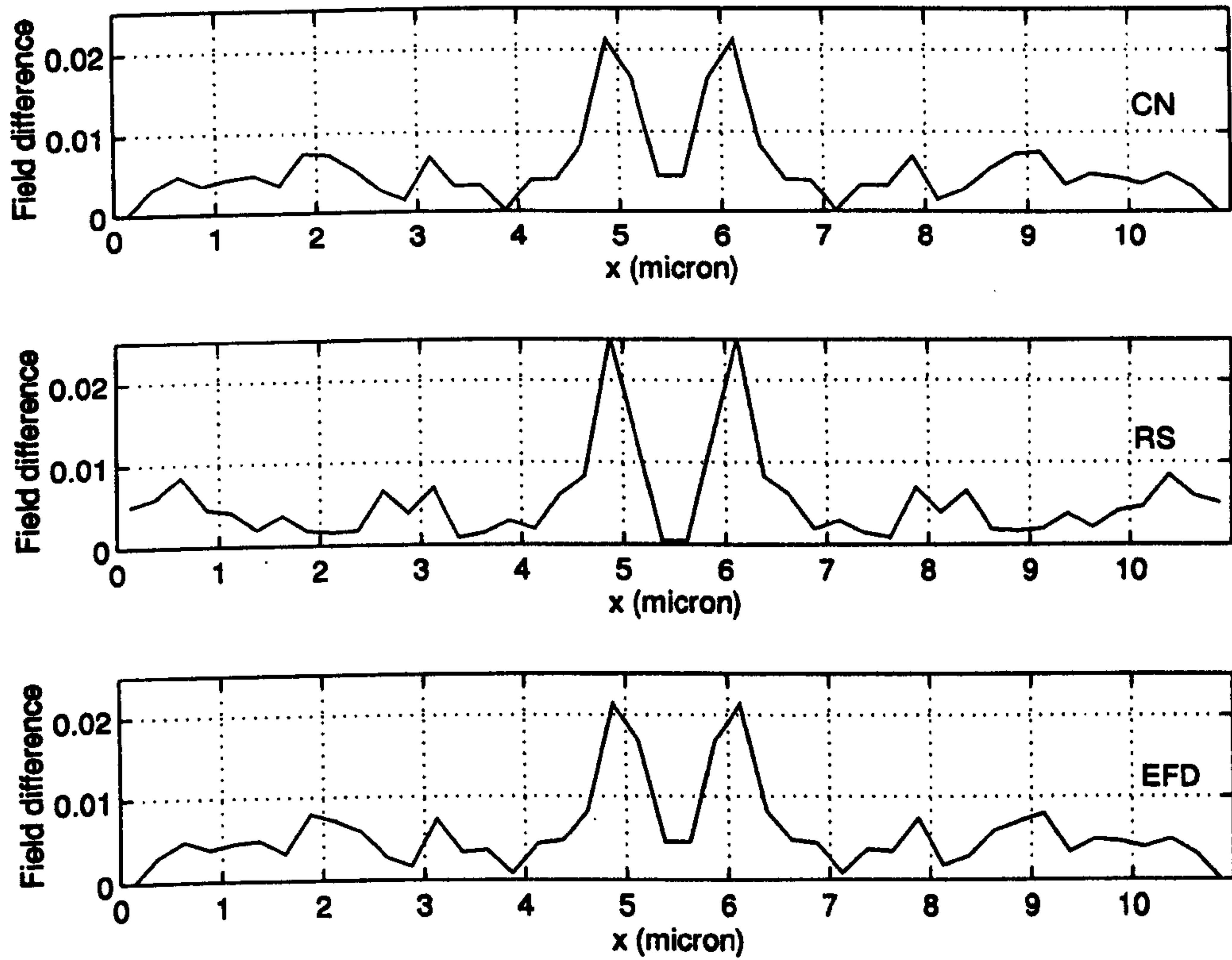


Fig. 5.10 The absolute field difference (Eq. 5.24) produced by the CN-BPM, the RS-BPM and the EFD-BPM after propagating the first guided mode of the waveguide of Fig. 5.4 to a distance of 100 using $\Delta z = 0.00625$ and $\Delta x = 0.25$ (all dimensions in μm).

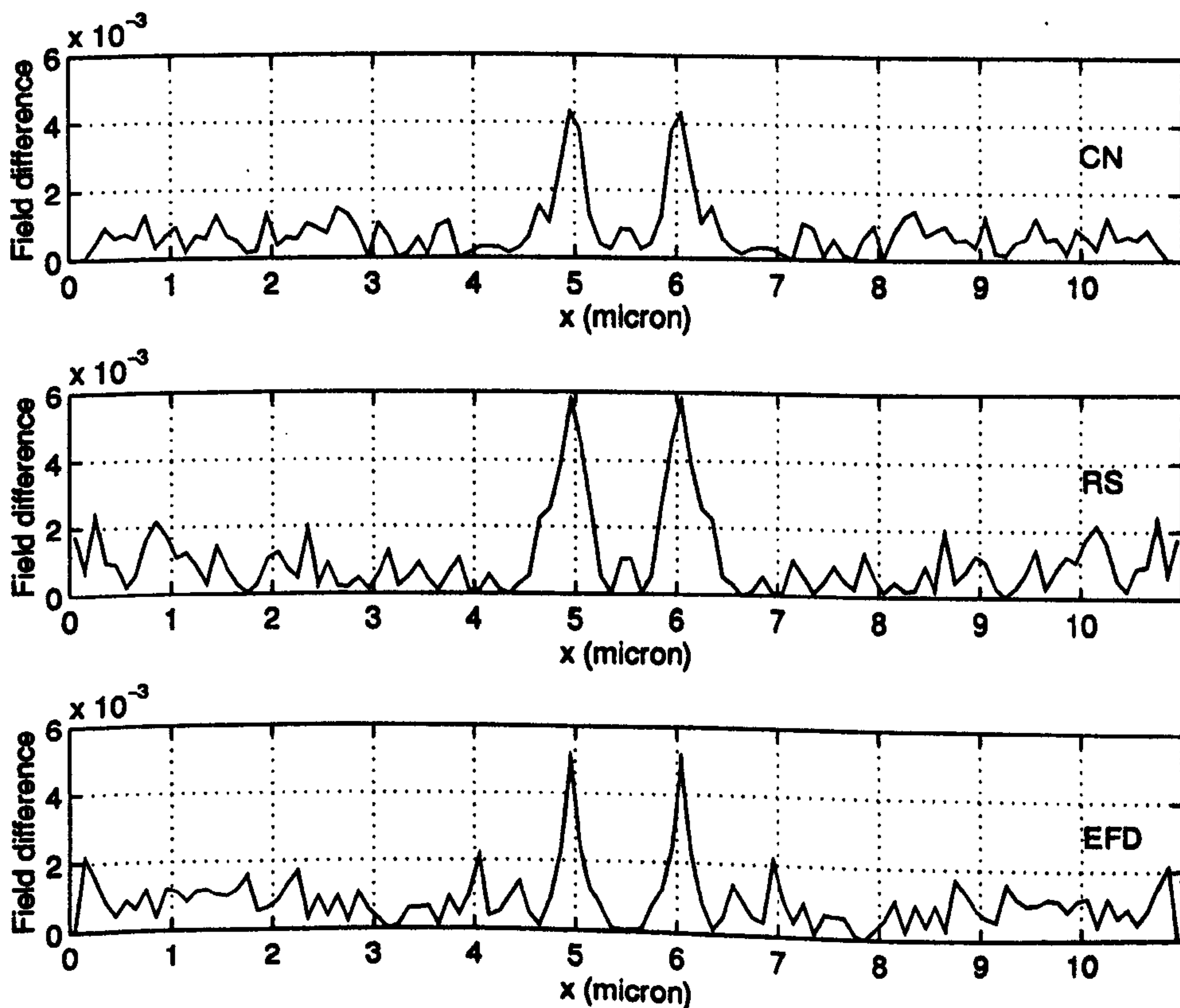


Fig. 5.11 The same as Fig. 5.10 except $\Delta x = 0.1 \mu\text{m}$.

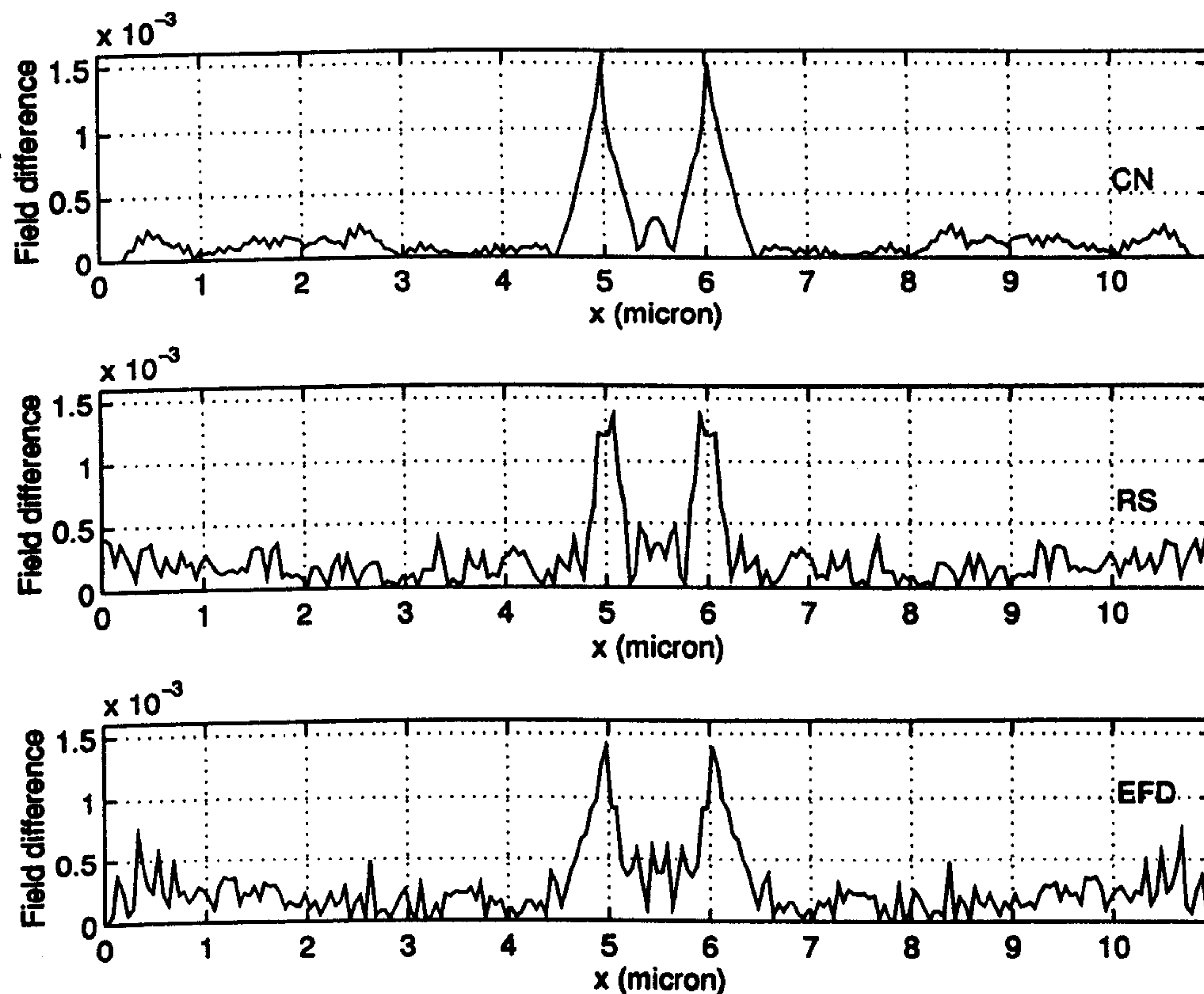


Fig. 5.12 The same as Fig. 5.10 except $\Delta x = 0.05 \mu\text{m}$.

In the following assessment, the influence of Δx and Δz on the value of the effective index of the mode is studied by computing the BPM effective index and comparing it with the analytical value. The computation of the BPM effective index of the mode was carried out by the spectral method which is discussed in chapter 3. First the accuracy of the spectral method was examined by assuming that the function $P_c(z)$, in (3.31), consists of a sum of [27]

$$P_c(z) = \sum_{N=1}^{10} e^{-jk_{z,N}z} \quad (5.27)$$

where the assumed values of $k_{z,N}$ are ranging from 400 to 800 cm^{-1} in an increment of 40 cm^{-1} . The weight of all the ten functions are taken to be unity. The function in (5.27) was computed numerically using a total of $I_z = 8192$ steps with an increment of $10 \mu\text{m}$. Then Eq. 5.27 was multiplied with the Hanning window function (see Eq. 3.35) and the result was Fourier transformed to obtain $P_c^*(k_{z,N})$. In the Fourier domain the function $P_c^*(k_{z,N})$ was then analysed to re-compute the numerical values of the $k_{z,N}$. Fig. 5.13 shows the

modal spectrum of $P_c^w(k_{z,N})$ versus k_z where $k_z = (2\pi i)/Z$, ($i = 0, \pm 1, \pm 2, \pm 3, \dots, \pm L/2$) and Z is the total z distance. The figure shows ten peaks of equal amplitude near the assumed eigenvalues. From the data of Fig. 5.13, the line-shape fitting technique [28] was then employed to compute the exact location of the eigenvalues. Table 5.1 shows a comparison between the true and the calculated $k_{z,N}$. From the table, we can observe the high accuracy of the spectral method with a maximum error of the order of 4×10^{-6} . Figs. 5.14 and 5.15 show, respectively, the modal power spectrum for the 2-D slab waveguide of Fig. 5.4 excited with the first guided mode and a Gaussian field with a half beam width of $0.5 \mu\text{m}$. The only peak of Fig. 5.14, which is in the positive portion of the plot, belongs to the TE_0 mode. On the other hand, Fig. 5.15 shows the a large peak in the positive section of the plot which belongs to the TE_0 mode and many peaks in the negative section of the plot that are associated to radiation modes (see Fig. 5.5).

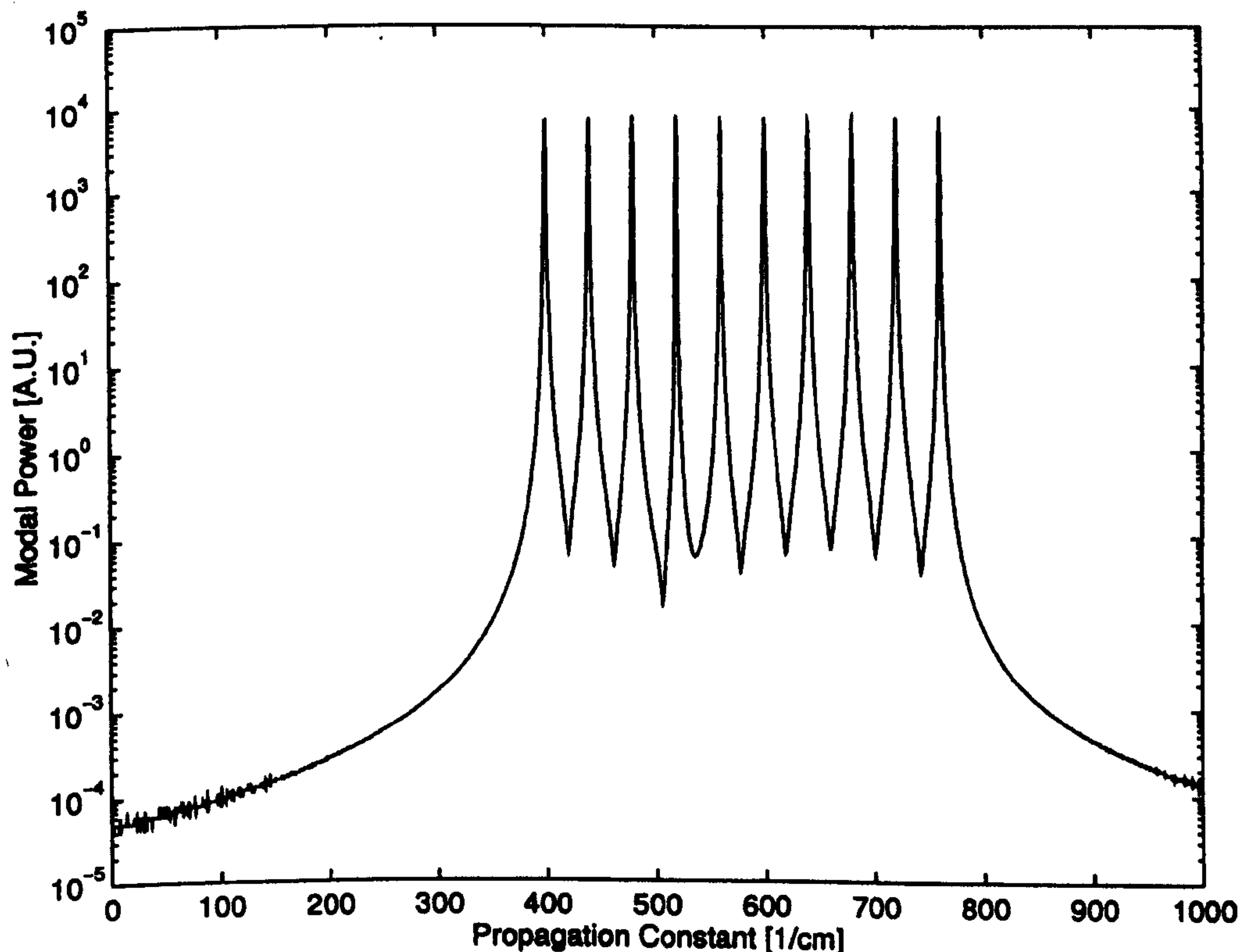


Fig. 5.13 The modal spectrum of the assumed ten harmonics of Eq. 5.27 (see text for details).

N	True $k_{z,N}$	Computed $k_{z,N}$
1	400.000000	400.000002
2	440.000000	440.000004
3	480.000000	480.000004
4	520.000000	520.000003
5	560.000000	560.000002
6	600.000000	600.000002
7	640.000000	640.000001
8	680.000000	680.000001
9	720.000000	720.000001
10	760.000000	759.999999

Table 5.1 Comparison between the true and the computed eigenvalues of the assumed ten harmonics of Eq. 5.27 using the spectral method.

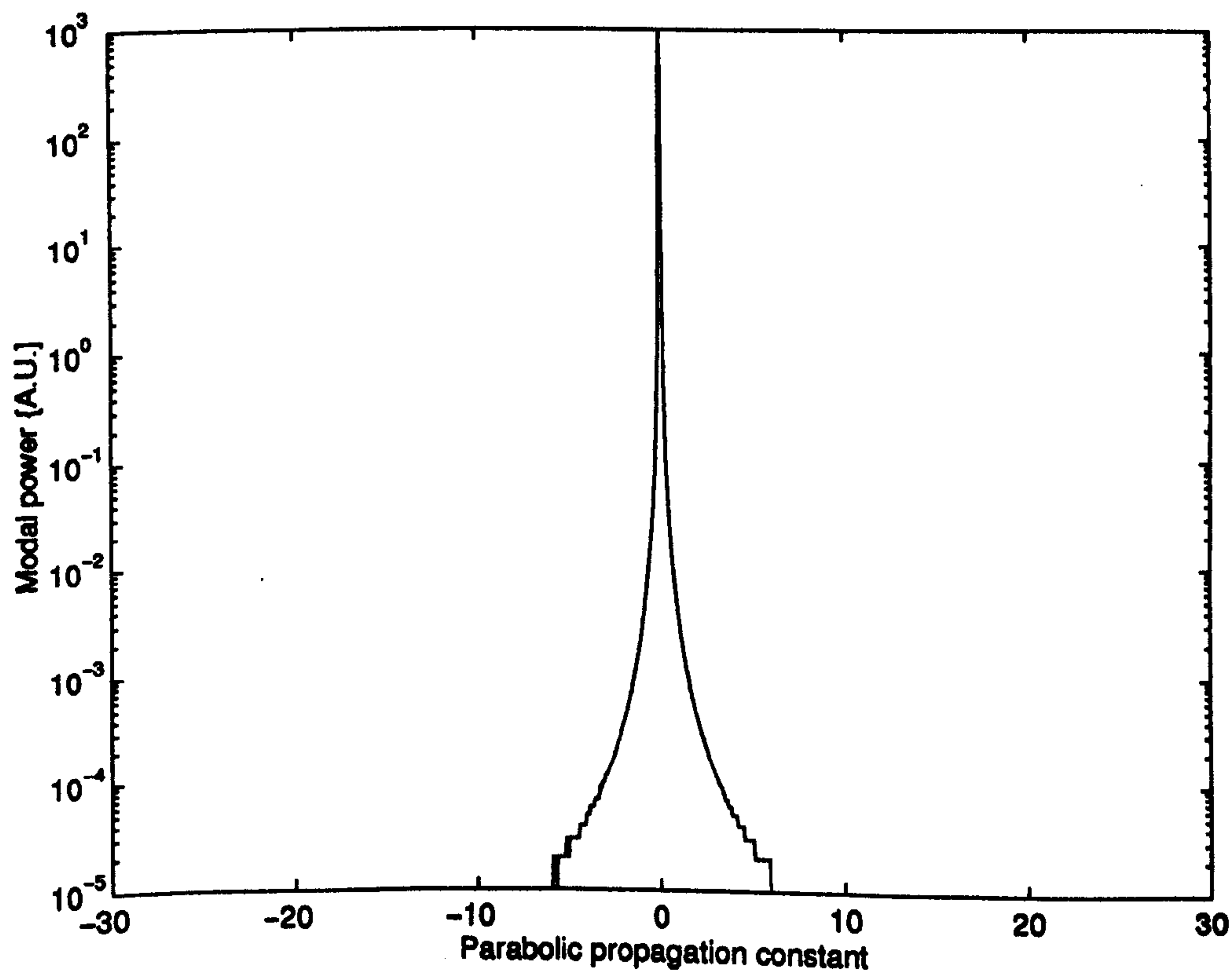


Fig. 5.14 The modal power spectrum for the slab waveguide of Fig. 5.4 excited with the TE_0 guided mode using the CN-BPM with $\Delta x = 0.05$, $\Delta z = 0.1$ and a total distance $Z = 102.4$ (all dimensions are in μm).

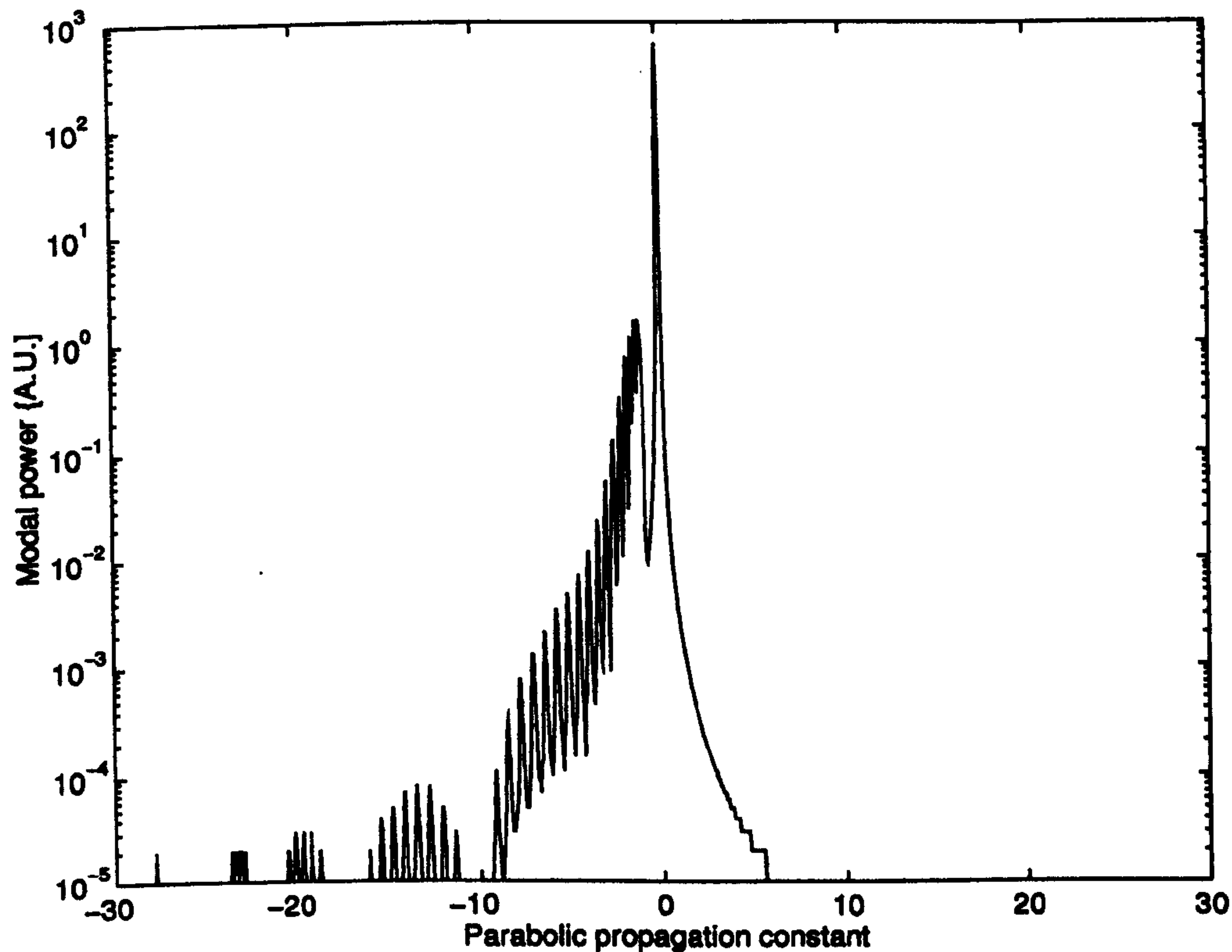


Fig. 5.15 The same as Fig. 5.14 except the initial input field is a Gaussian beam with a half beam width of $0.5 \mu\text{m}$ and $\lambda = 1.0$.

To study the influence of Δx and Δz on the effective indices of the three FD-BPMs, the TE_0 of the symmetric slab waveguide of Fig. 5.4 was used as an initial input field where it has been propagated to a fixed distance of $z = 102.4 \mu\text{m}$ using all FD-BPMs. The spectral method is used to compute the effective index of the TE_0 ($n_{\text{eff},0}^{\text{BPM}}$) from the BPM field. Fig. 5.16 shows the percentage error difference which is defined as

$$\text{Percentage error difference} = \left| n_{\text{eff},0}^{\text{analytical}} - n_{\text{eff},0}^{\text{BPM}} \right| \times 100 \quad (5.28)$$

as a function of Δz for different Δx using the CN-BPM, the RS-BPM and the EFD-BPM. From the figure we can observe that the error generally decreases as Δx is decreased for the three algorithms. The figure also shows that the accuracy of both the CN-BPM and the EFD-BPM are not very sensitive with the change of Δz but Δz for the EFD-BPM must be less than the limit of the stability condition of (5.22). We must note from this stability condition of the EFD-BPM that as Δx is decreased, Δz must be also decreased. On the other hand, Fig. 5.16 shows the convergence of the RS-BPM as a function of Δz and also noticing, as for the EFD-BPM, that Δz must decrease as Δx is reduced.

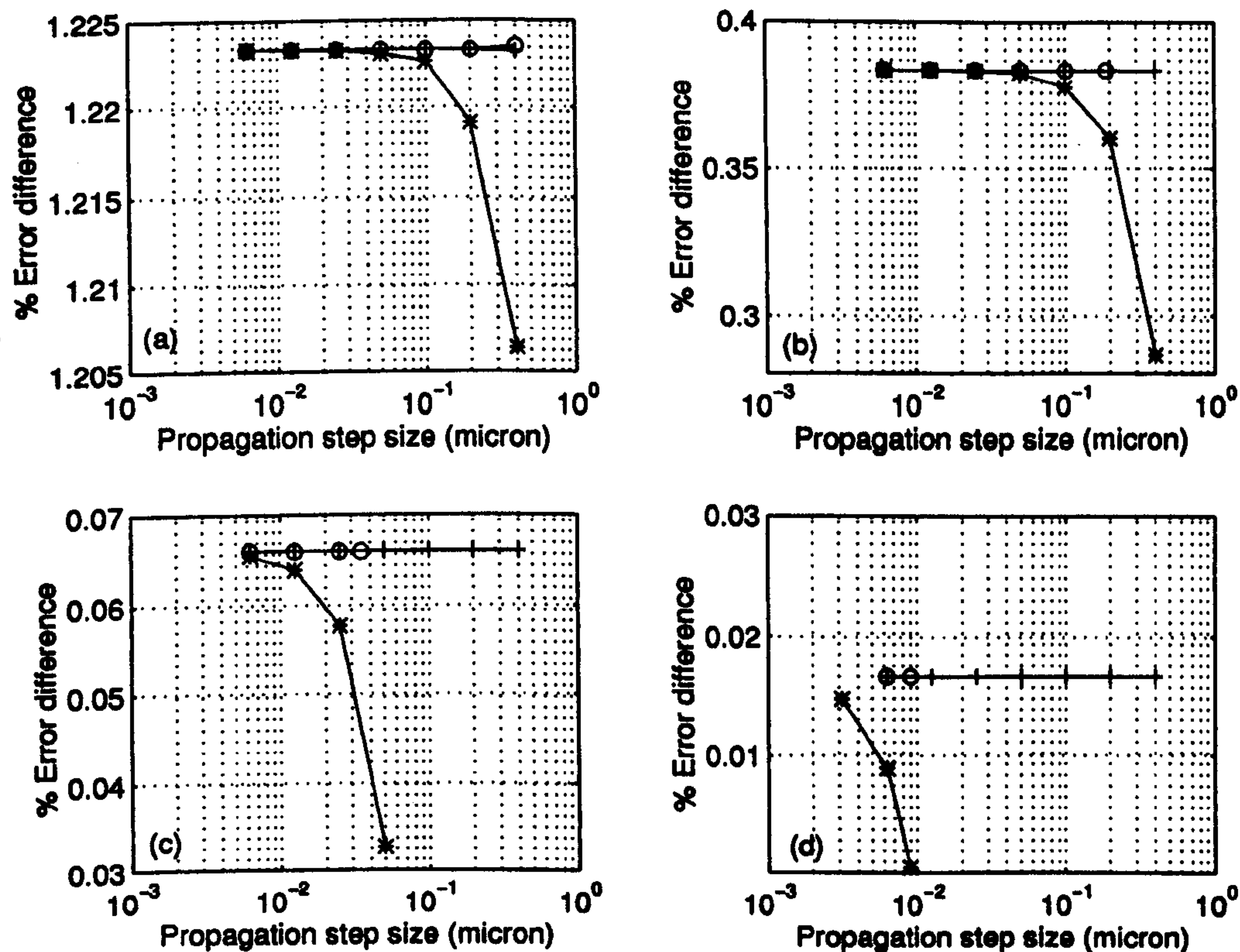


Fig. 5.16 The percentage error difference of the first guided mode FD-BPMs' effective index using the CN-BPM (+), the RS-BPM (*) and the EFD-BPM (o) as a function of the longitudinal step size Δz for a) $\Delta x = 0.5$, b) $\Delta x = 0.25$, c) $\Delta x = 0.1$ and d) $\Delta x = 0.05$ (all dimensions are in μm).

In order to compare between the efficiency of all the implemented FD-BPM methods, Fig. 5.17 shows the CPU time per propagational step as a function of the total transverse mesh points M_x for the three algorithms discussed before. The speed of the 2-D FFT-BPM has been included in the figure for comparison. It can be observed from the figure that the speeds of the three FD-BPMs are much faster than the FFT-BPM. On the other hand, comparisons between the three FD-BPMs show that the EFD-BPM is 2.4 faster than the CN-BPM and 3.2 faster than the RS-BPM when $M_x = 1100$. A 486 personal computer with a speed of 33 MHz has been used for the speed computations of Fig. 5.17. Finally, using the results of Figs. 5.16 and 5.17, we may conclude the following. Both the EFD and the CN are more efficient than the RS since the latter converges using smaller Δz than the first two and also the speed of the RS is slower, per propagation step, than the speed of EFD and the CN. For large Δx , both the EFD and the CN produce similar results near the same Δz , and that shows the EFD is more efficient since it is faster per propagational step. As Δx decreases, the EFD requires smaller Δz for stability constraint and that implies that the CN is more efficient for very small Δx .

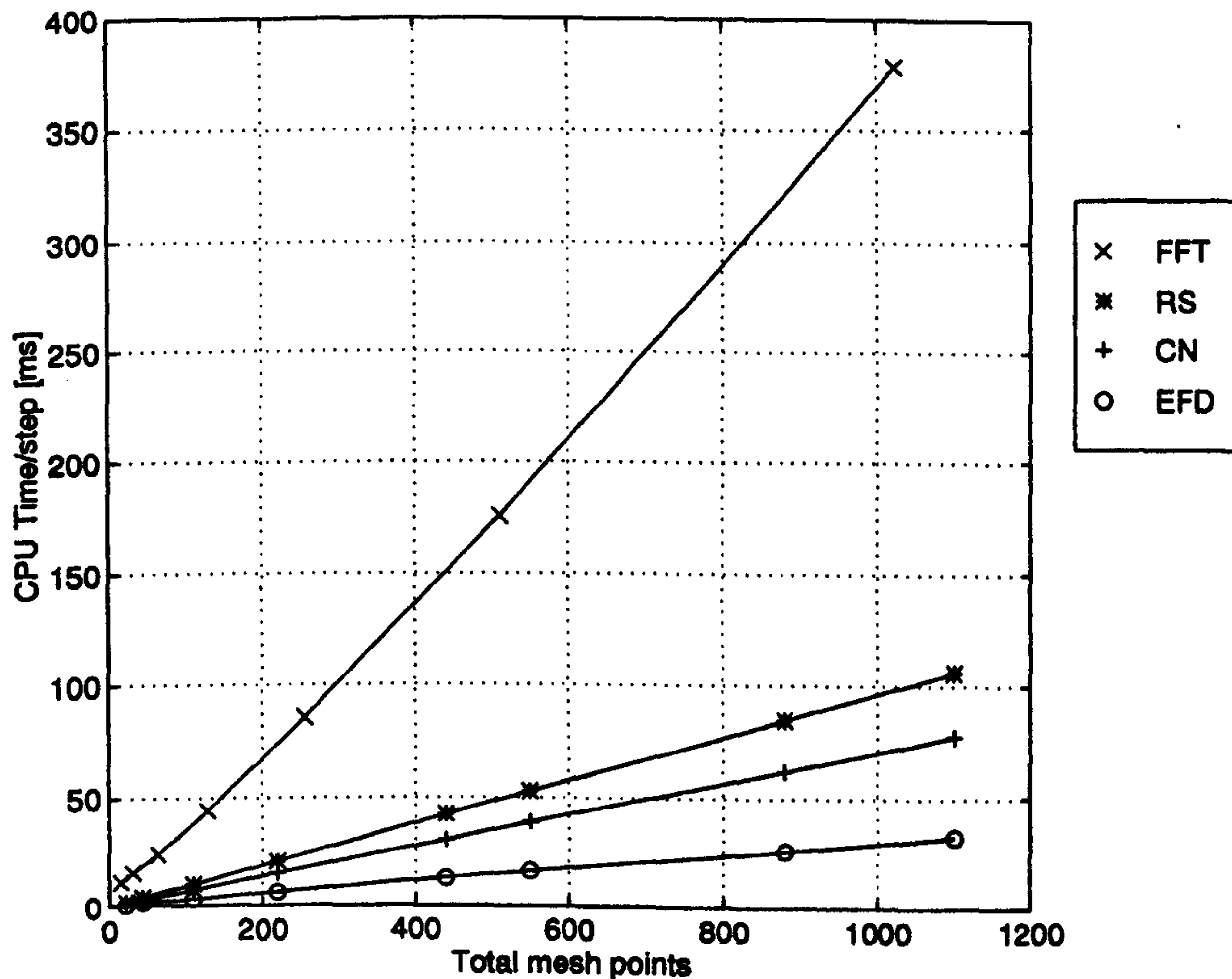


Fig. 5.17 Comparison between the 2-D speed of the CN-BPM, the RS-BPM, the EFD-BPM and the FFT-BPM as a function of the transverse mesh points (M_x).

5.4 The Modified EFD-BPM

The simplicity and the high efficiency of the EFD-BPM make it a very attractive technique to analyse optical devices especially for three-dimensional applications. It turns out that a slight modification to the EFD-BPM improves the stability limit on the longitudinal step size of the method by many fold. If the field $\phi_i(z)$ in Eq. 5.21 is replaced by the following relation [18, 29]

$$\phi_i(z) = \frac{1}{2}[\phi_i(z + \Delta z) + \phi_i(z - \Delta z)] \quad (5.29)$$

then this will lead to

$$\phi_i(z + \Delta z) = V'_i \phi_i(z - \Delta z) - V''_i [\phi_{i-1}(z) + \phi_{i+1}(z)] \quad (5.30)$$

where

$$V_i' = \frac{(2 + V_i)}{(2 - V_i)}$$

$$V_i'' = \frac{[(2\Delta z)/(jk_0 n_0 \Delta x^2)]}{(2 - V_i)}$$

$$V_i = \frac{\Delta z}{jk_0 n_0} \left[-\frac{2}{\Delta x^2} + k_0^2 (n_i^2(z) - n_0^2) \right]$$

The above method is referred to in the literature as the Du Fort-Frankel scheme [18, 29], and for simplicity we will refer to it as the Modified Explicit FD or in short as the MEFD-BPM. Tests for the MEFD-BPM showed that the technique is stable for propagational steps higher than those of the EFD-BPM and produced very close results to those of the CN-BPM. In addition the computations in Eq. (5.30), along the direction of propagation, can be divided into two separate sets (see Fig. 5.18) in which one of these sets could be used in the computation. This splitting for the total mesh points reduces the amount of computation by half while retaining the same accuracy. However, during the analysis of the MEFD-BPM, a spurious field was observed which restricts the usefulness of the method. Fig. 5.19 shows the power spectrum for the 2-D slab waveguide in Fig. 5.4, excited with the first guided mode using the MEFD-BPM. Comparison between this figure and the results of the CN-BPM in Fig. 5.14 shows that there exists an extra peak on the right hand side of the guided mode peak of Fig. 5.19, which is believed to belong to a spurious field. The height of this peak gets smaller by reducing the step size Δz . Although the height of the peak is small compared to the height of the guided mode peak, it was concluded that the MEFD-BPM is not useful unless the fake field is fully removed. However, recently the MEFD-BPM has been published in [30] without any reference to this spurious field². From [30], the Von Neumann strategy was used to give the following relation for the stability of the MEFD-BPM for real refractive index

$$\Delta z \leq \frac{n_0}{\sqrt{n^2 - n_0^2}} \Delta x \quad \text{where } n > n_0 \quad (5.31)$$

² This reference was found during the time of writing this chapter.

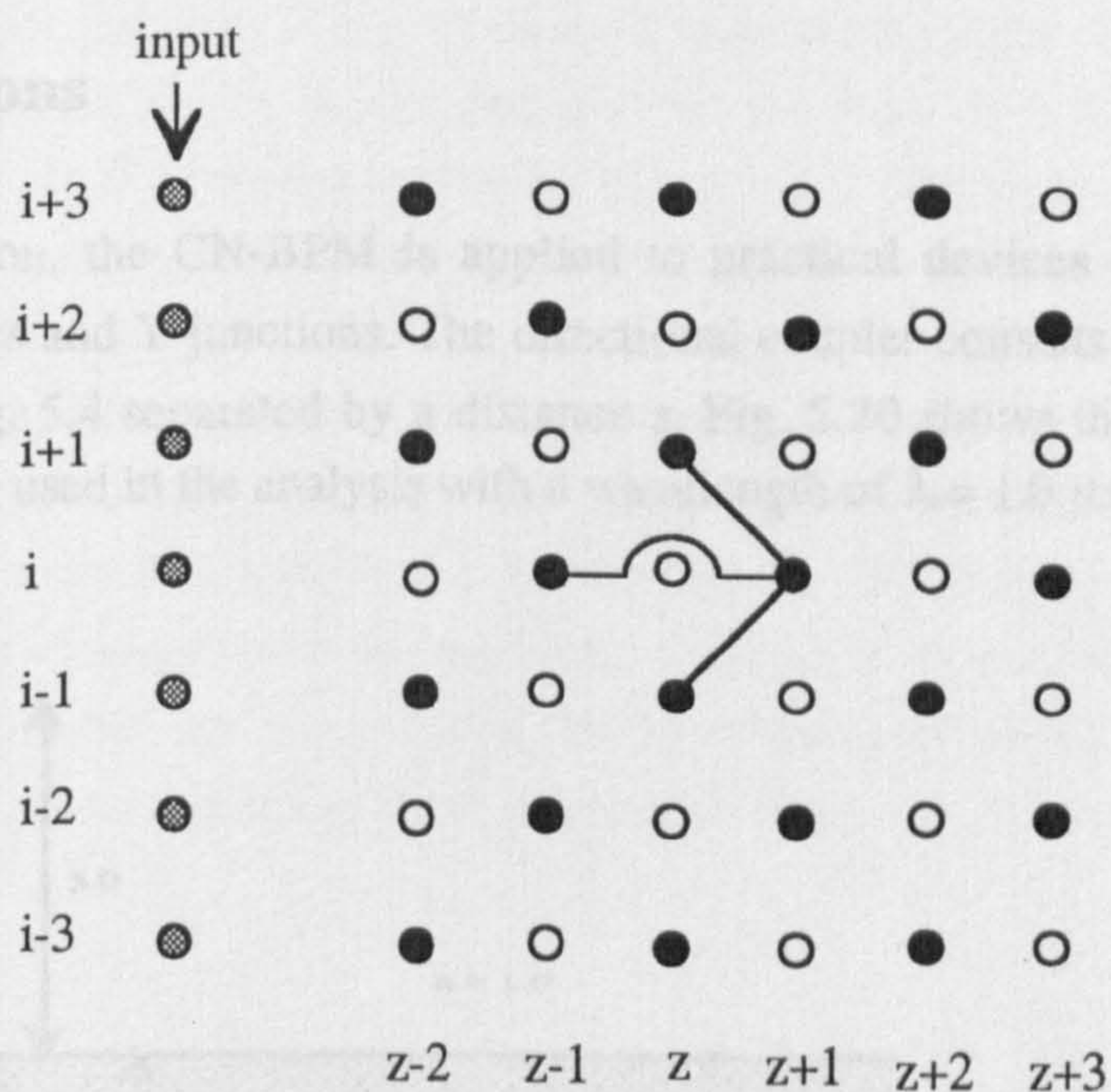


Fig. 5.18 The split of the total mesh points of the MEFD-BPM into two independent sets (black circles set and white circles set) which reduces the total computational time by half. The gray circles indicates the total mesh points of the input.

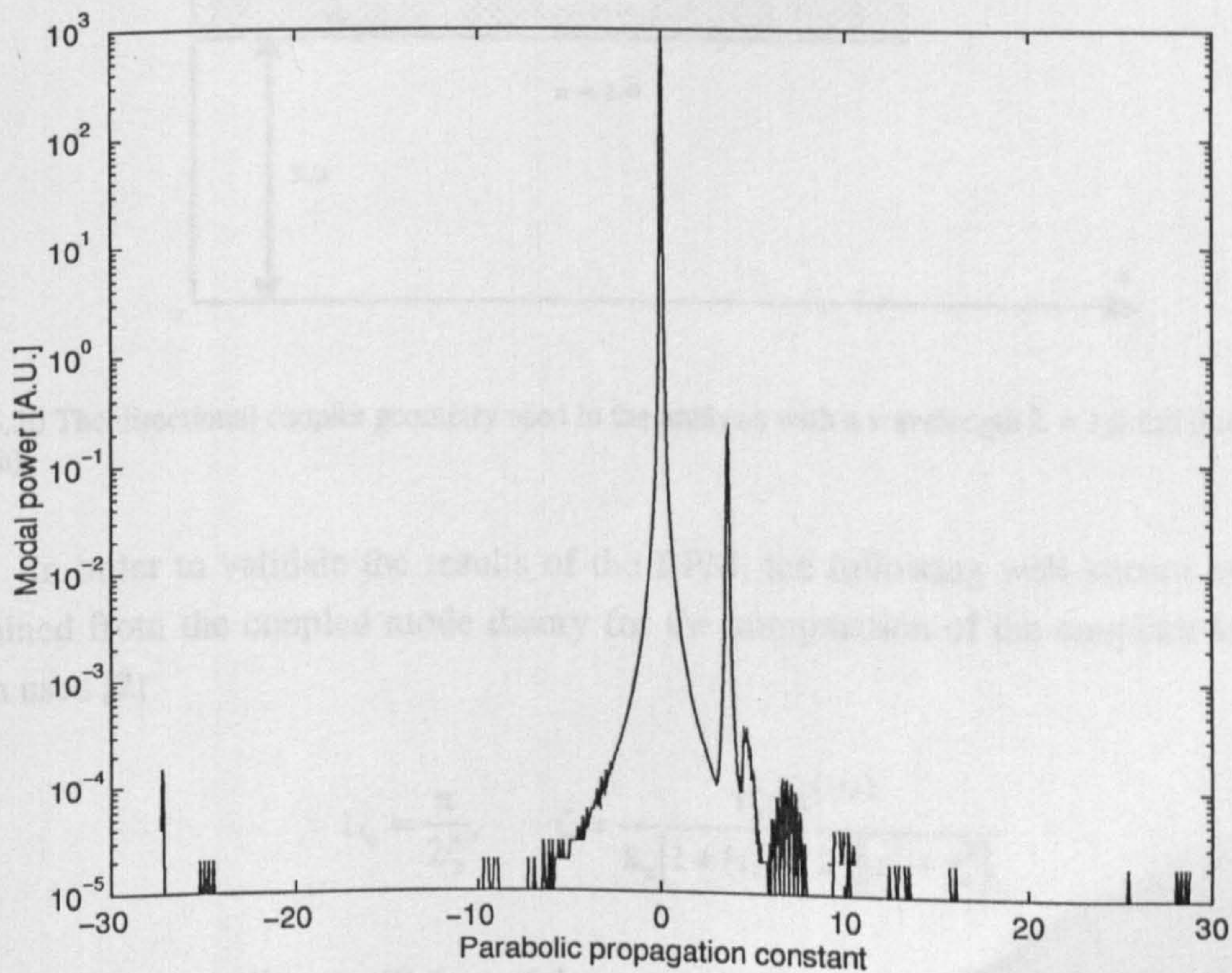


Fig. 5.19 The modal power spectrum for the slab waveguide of Fig. 5.4 excited with the TE_0 guided mode using the MEFD-BPM with $\Delta x = 0.05$, $\Delta z = 0.1$ and a total distance $Z = 102.4$ (all dimensions are in μm).

5.5 Applications

In this section, the CN-BPM is applied to practical devices of two-dimensional directional couplers and Y-junctions. The directional coupler consists of two coupled slab waveguides of Fig. 5.4 separated by a distance s . Fig. 5.20 shows the dimensions of the directional coupler used in the analysis with a wavelength of $\lambda = 1.0 \mu\text{m}$.

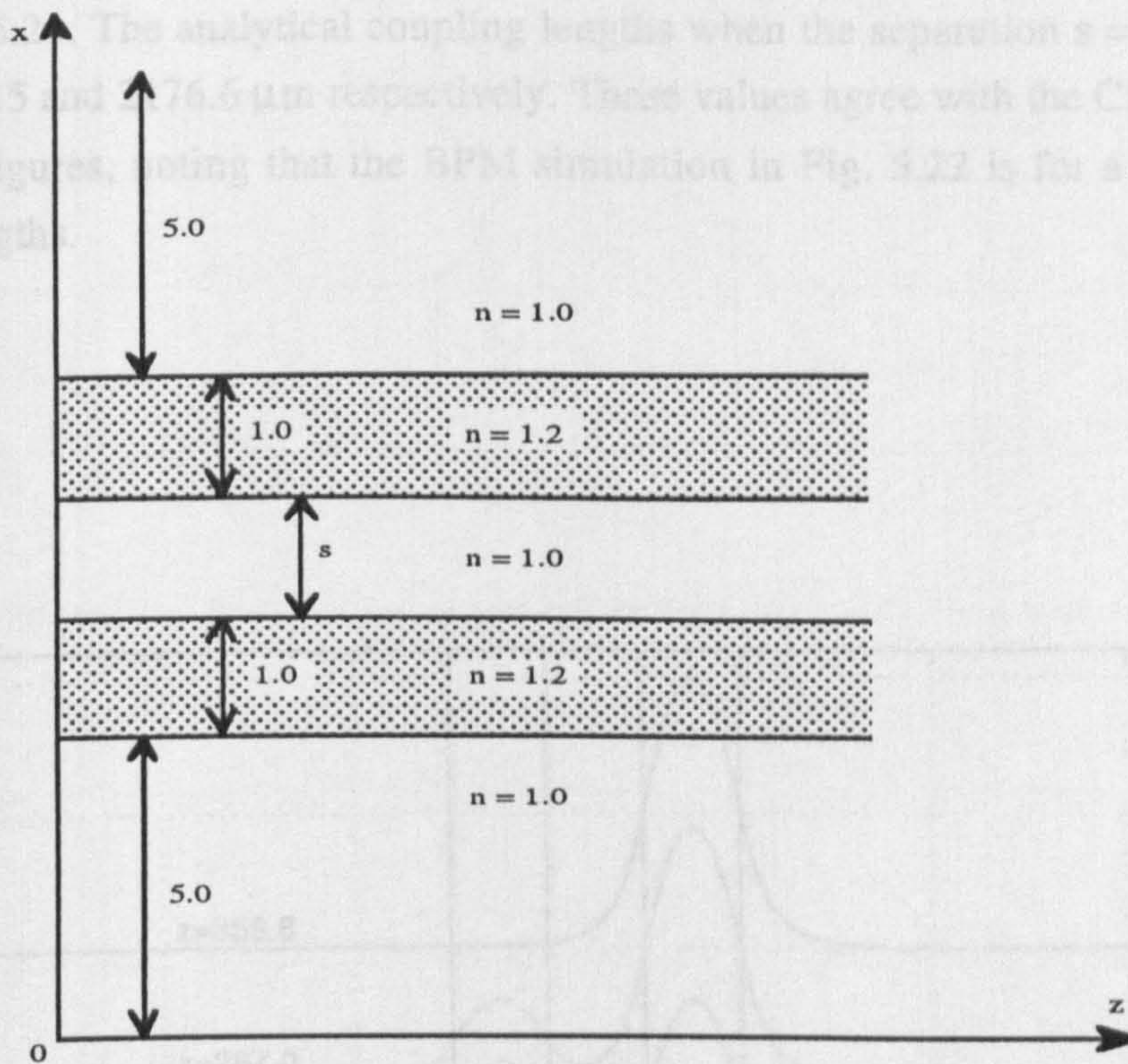


Fig 5.20 The directional coupler geometry used in the analysis with a wavelength $\lambda = 1.0$ (all dimensions are in μm).

In order to validate the results of the BPM, the following well-known expression, obtained from the coupled mode theory for the computation of the coupling length, has been used [2]

$$L_c = \frac{\pi}{2\zeta}, \quad \zeta = \frac{\Gamma_g^2 \Gamma_s^2 e^{-r_s s}}{k_z [1 + (r_s d / 2)] [\Gamma_g^2 + \Gamma_s^2]} \tag{5.32}$$

where ζ is the coupling coefficient of the coupled slab waveguides of Fig. 5.20 (the rest of the parameters are defined below Eq. 2.19). In the following computations, the TE_0 guided

mode of the slab waveguide of Fig. 5.4 was used as an input to one of the slab waveguides of Fig. 5.20. Fig. 5.21 shows the propagation of the field to a distance equal to one coupling length of Eq. 5.32 with a separation of $s = 1 \mu\text{m}$. The figure shows the exchange of power from one waveguide to the other, and also shows the close agreement of the coupling lengths between the analytical, which is equal to $L_c = 358.8 \mu\text{m}$, and the CN-BPM. The same has been repeated to two more analyses, one is by decreasing the separation distance s and the other is by increasing it. The results of Figs 5.22 and 5.23 are for separation distances of $s = 0.5$ and $1.5 \mu\text{m}$ respectively, under the same conditions of that of Fig. 5.21. The analytical coupling lengths when the separation $s = 0.5$ and $1.5 \mu\text{m}$ are $L_c = 59.15$ and $2176.6 \mu\text{m}$ respectively. These values agree with the CN-BPM analysis in the two figures, noting that the BPM simulation in Fig. 5.22 is for a distance of two coupling lengths.

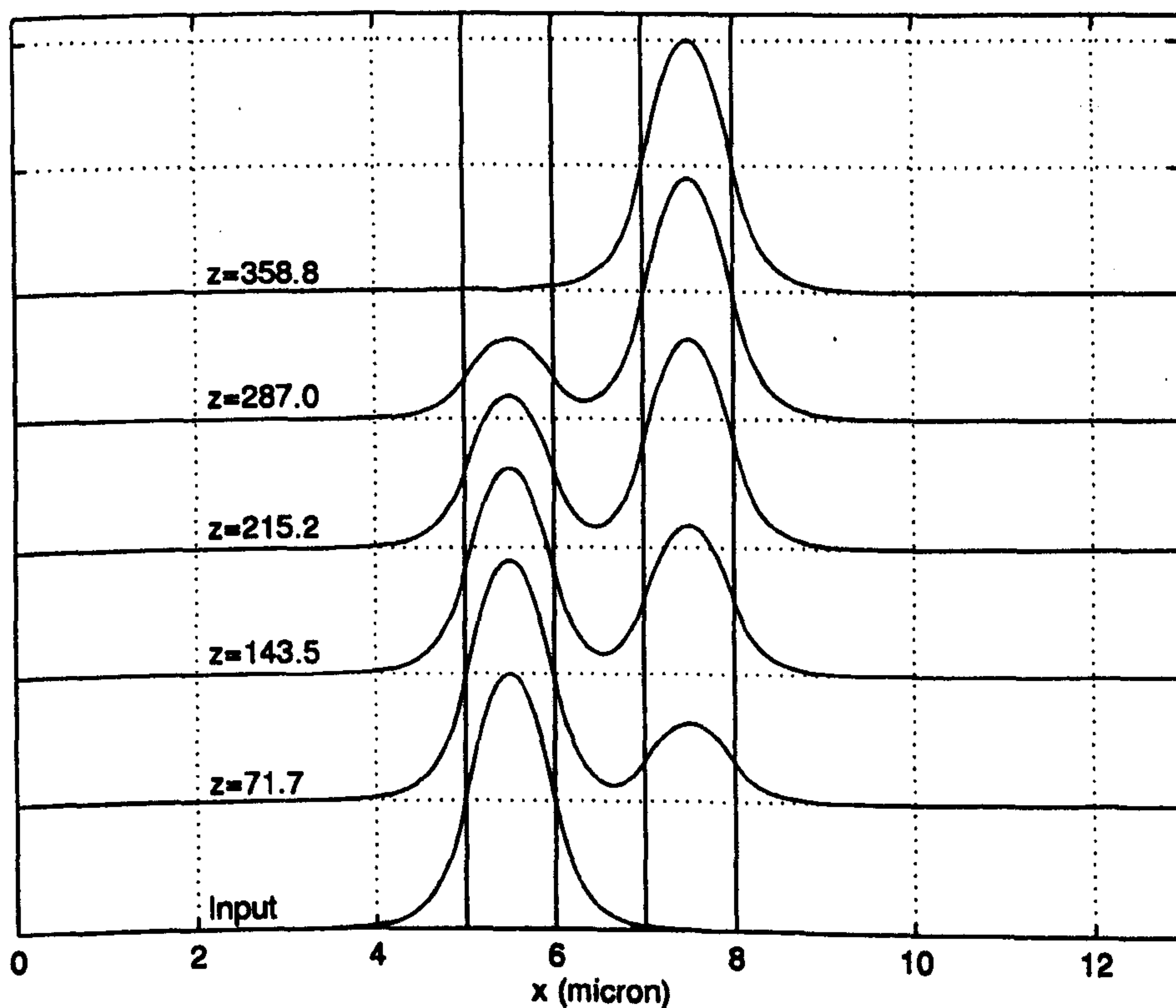


Fig. 5.21 The propagation of the TE_0 guided mode of the single slab waveguide in Fig. 5.4 inside the coupled slab waveguide of Fig. 5.20 with a separation distance $s = 1$. The total propagation distance used is equal to one coupling length, $\Delta x = 0.05$ and $\Delta z = 0.1$ (all dimensions are in μm).

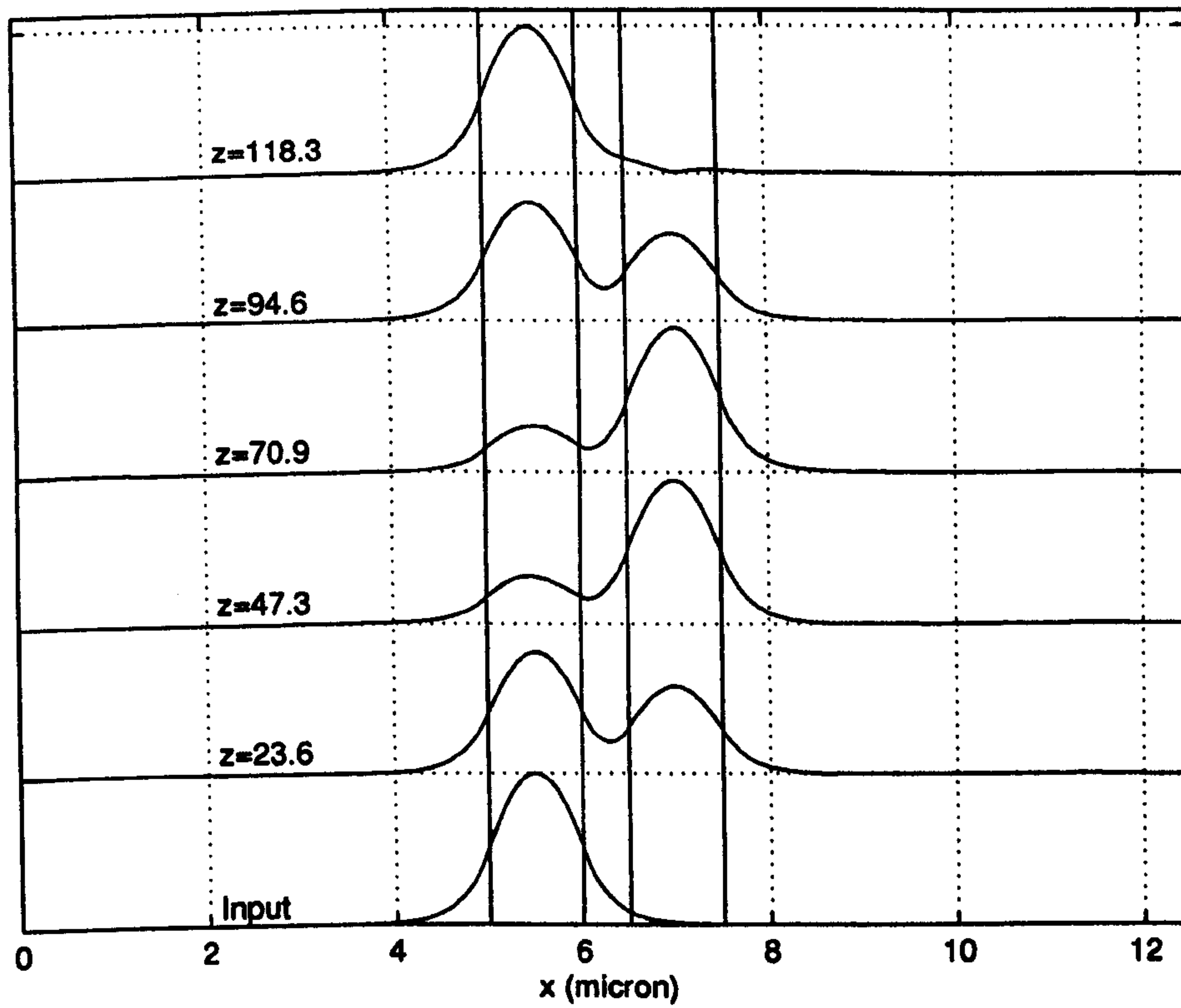


Fig. 5.22 The same as Fig. 5.21 except $s = 0.5 \mu\text{m}$ and the total propagation distance of two coupling lengths.

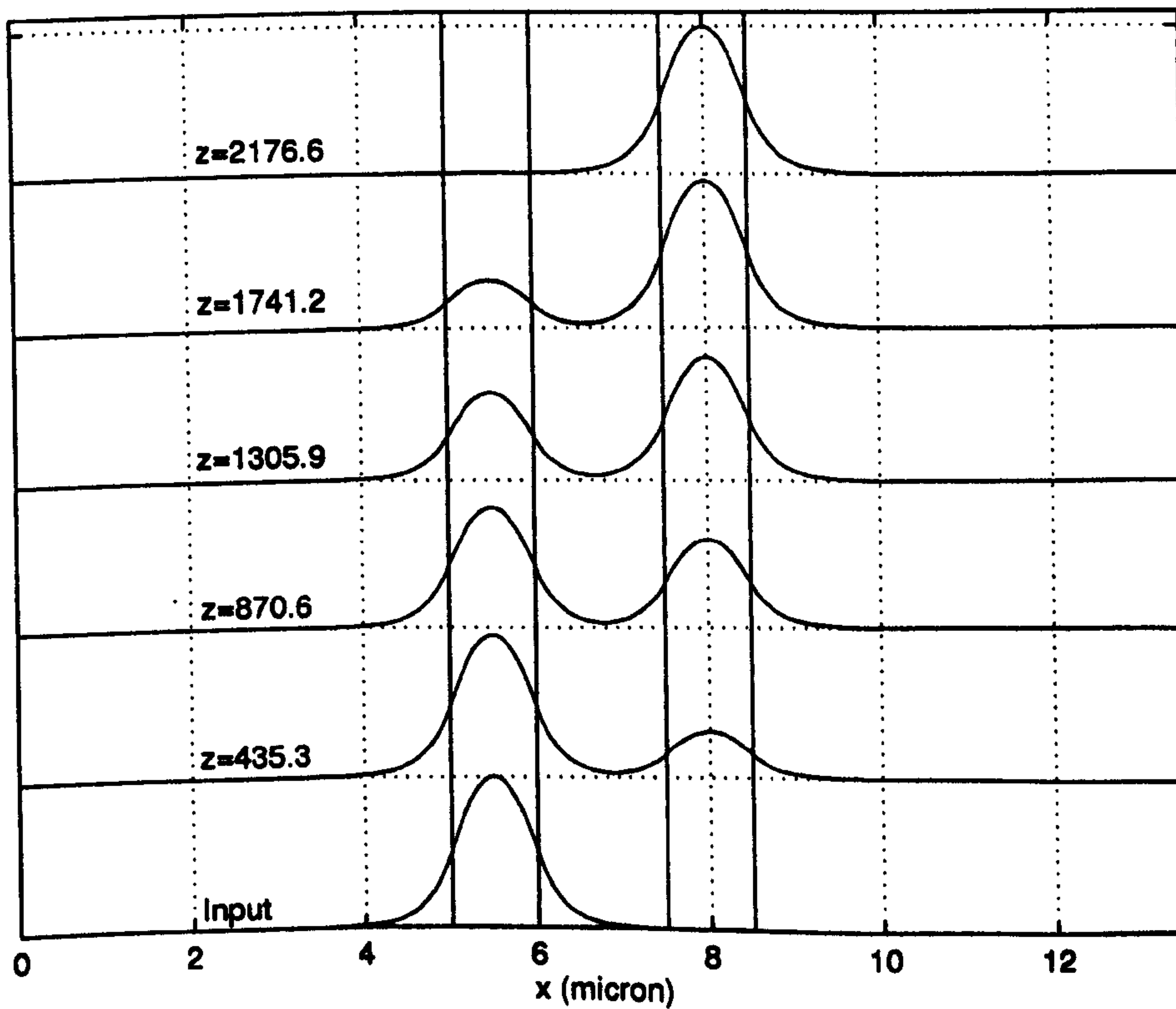


Fig. 5.23 The same as Fig. 5.21 except $s = 1.5 \mu\text{m}$ and the total propagation distance of one coupling length.

The second practical application for the CN-BPM is to simulate the Y-junction in Fig. 5.24 with a wavelength of $\lambda = 1.0 \mu\text{m}$. In the simulation, the TE_0 guided mode of the combined waveguides was used as an input to the Y-junction. Fig. 5.25 shows the simulation, using the CN-BPM, when the half branching angles are equal ($\alpha_1 = \alpha_2 = 2.5$ degrees) forming a symmetric Y-junction. It has to be noticed that the branched waveguides support only one guided mode. From the figure, we may notice that the input field has been separated into two equal and symmetrical output fields, each having the shape of the guided mode of the branched waveguide. The separation between the two output peaks of the fields is equal to $4.875 \mu\text{m}$ which is very close to the expected value. The second test involves the simulation of an asymmetric Y-junction. Fig. 5.26 shows the results of modelling the asymmetric Y-junction where the half branching angles are equal to $\alpha_1 = 0$ and $\alpha_2 = 2.5$ degrees. The figure shows that the two fields split asymmetrically compared to the previous analysis. We may also notice that the peak of the field of the straight waveguide is shifted toward the centre of the branched waveguide.

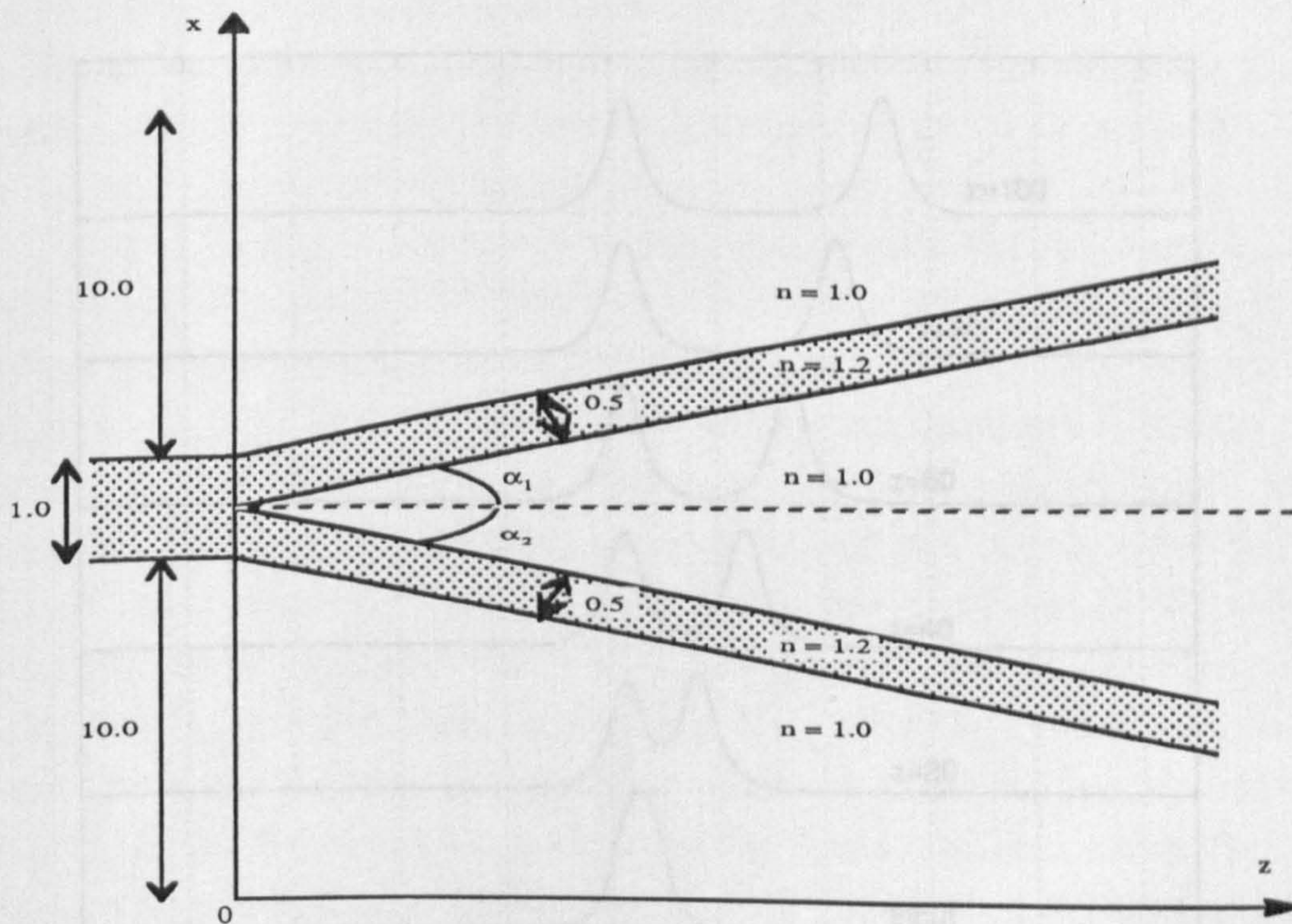


Fig 5.24 The Y-junction geometry used in the analysis with a wavelength $\lambda = 1.0$ (all dimensions are in μm).

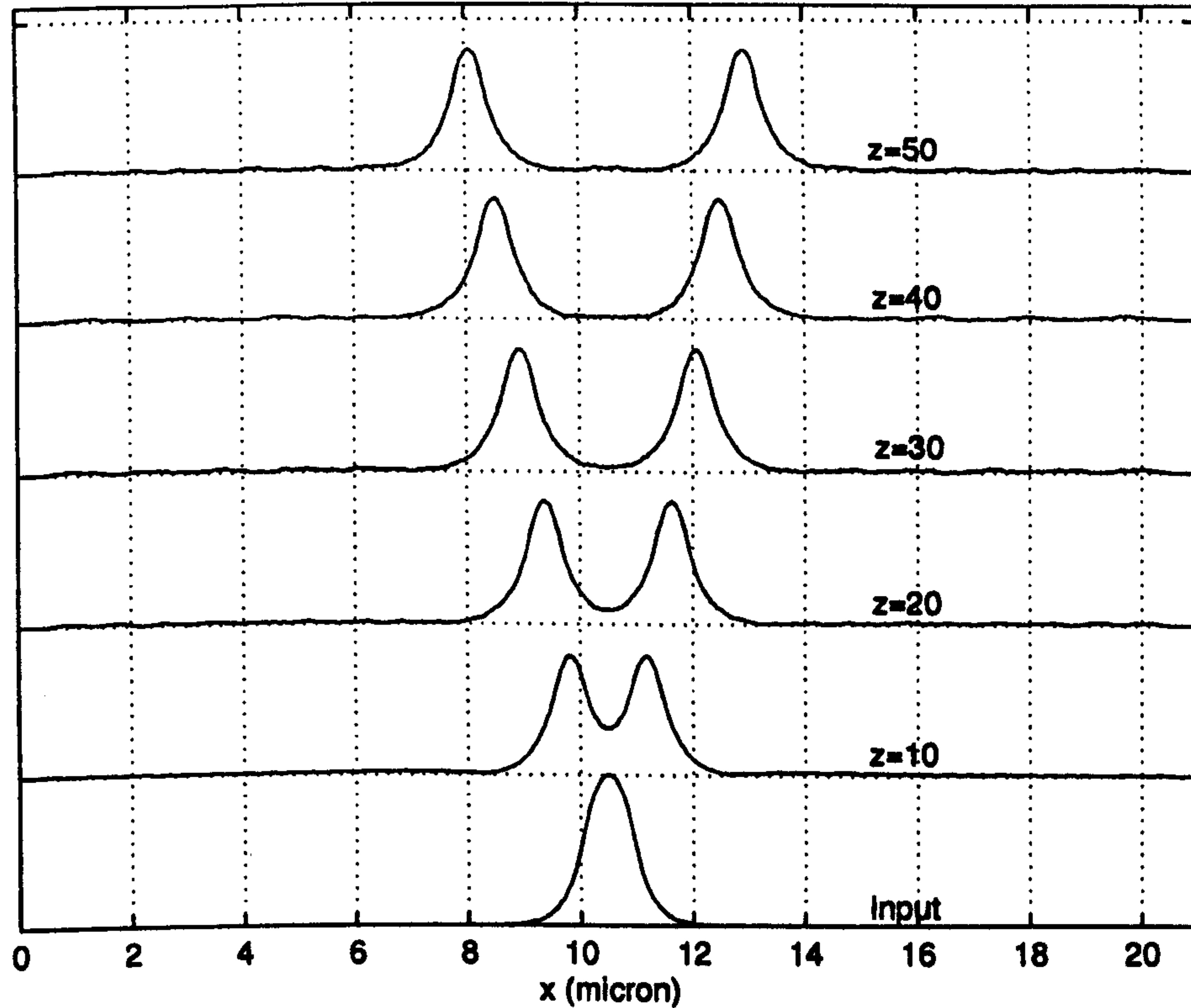


Fig. 5.25 The simulation of a symmetric Y-junction, of Fig. 5.24 where $\alpha_1 = \alpha_2 = 2.5$ degrees, using the CN-BPM with $\Delta x = 0.025$ and $\Delta z = 0.1$ (all dimensions are in μm).

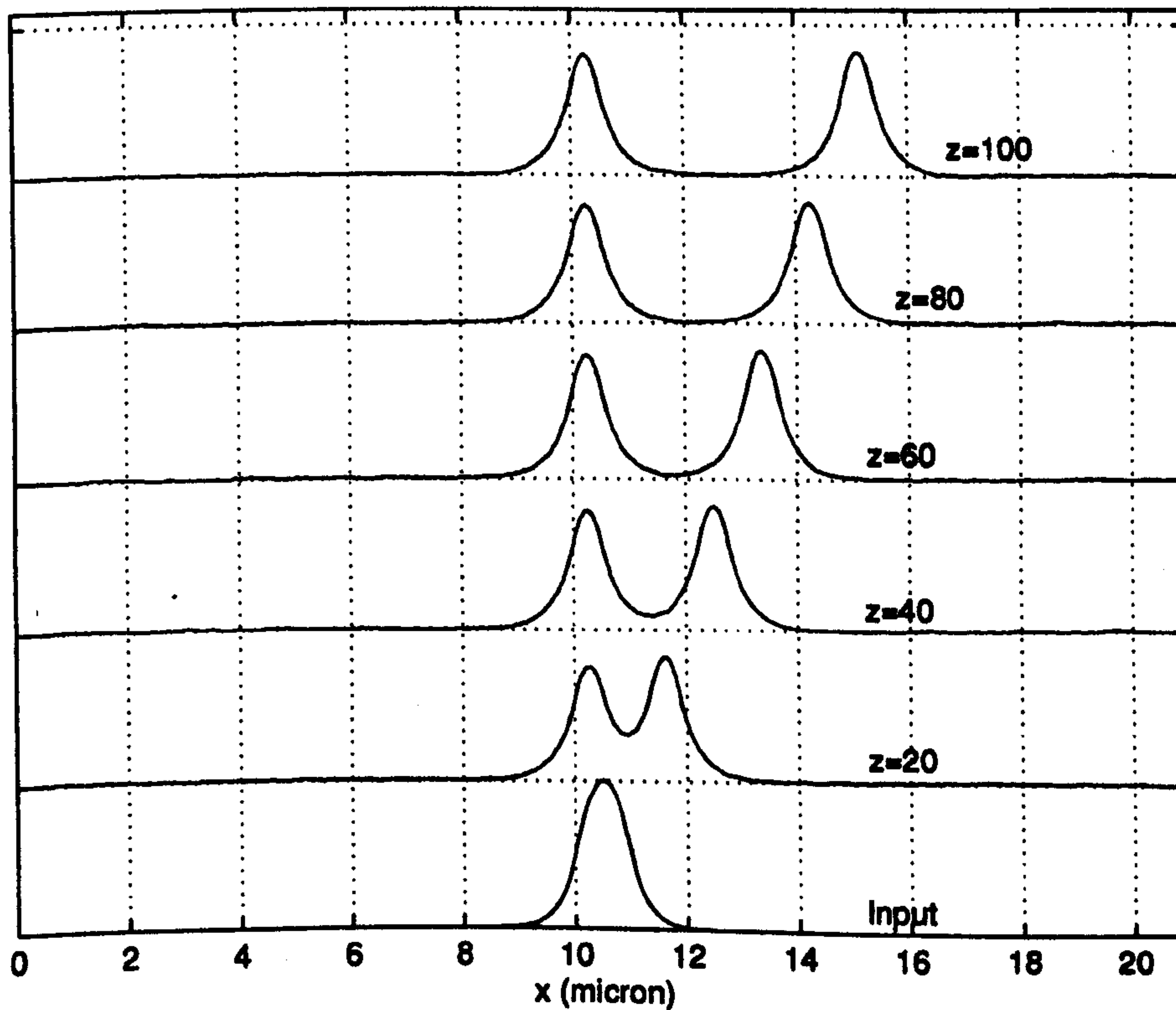


Fig. 5.26 The simulation of an asymmetric Y-junction, of Fig. 5.24 where $\alpha_1 = 0$ and $\alpha_2 = 2.5$ degrees, using the CN-BPM with $\Delta x = 0.025$ and $\Delta z = 0.1$ (all dimensions are in μm).

5.6 Conclusion

In this chapter, the formulations and the numerical implementations of the three 2-D FD-BPMs of chapter 3, have been discussed. A recursive numerical technique has been used which solves the tridiagonal system of equations, created by the CN-BPM algorithm, and which uses a very low computer storage memory. The three FD-BPM computer codes have been tested, on a normal serial PC, to propagate different initial fields in a homogenous medium and a 2-D slab waveguide. To assess the accuracy analysis for the three methods, the spectral technique was tested and used to compute the propagation constants from the BPM fields for the propagation of guided modes of z-invariant slab waveguides. It has been observed that the accuracy, compared to the analytical solutions, of the three algorithms improves by decreasing the mesh spacing Δx . On the other hand, while the accuracy of the RS-BPM converges by reducing the longitudinal step Δz , the accuracy of both the CN-BPM and the EFD-BPM is not sensitive to the change in Δz , provided that Δz of the EFD-BPM is less than the stability limit of the method. Also it has been noticed that Δz for both the RS-BPM and EFD-BPM must be decreased when Δx is reduced for the first to converge and for the second to remain stable. Then a comparison between the speed of all the three FD-BPMs with the speed of the 2-D FFT-BPM showed that the FD-BPMs are much faster, per propagational step, than the FFT-BPM. On the other hand, comparison between the speed per propagational step of the three FD-BPMs showed that the EFD-BPM is faster than the other two methods and the CN-BPM is faster than the RS-BPM. Finally, it has been concluded that both the EFD-BPM and the CN-BPM are more efficient than the RS-BPM since the later converges using smaller Δz than the first two and also the speed of the RS-BPM is slower per propagational step. The comparison between the EFD-BPM and the CN-BPM shows that the EFD-BPM is more efficient than the CN-BPM at large Δx , since both algorithms produced similar results using the same Δz and the EFD-BPM is faster, per propagational step, than the CN-BPM. On the other hand as Δx decreases, the EFD-BPM requires smaller Δz for stability constraint, which might suggest that the CN-BPM is more efficient for very small Δx . To improve the stability limit of the propagational step of the EFD-BPM, a slight modification to the EFD-BPM equation was needed. Although this modification has improved the stability of the EFD-BPM, it also introduced a spurious field to the true solution. Then, as a final check on the mechanism and the accuracy of the FD-BPMs, the CN-BPM was used to model practical 2-D devices, where it has been applied to simulate the propagation of guided modes in a directional coupler and a Y-junction.

REFERENCES

- [1] Chung, Y. and Dagli, N., "An Assessment of Finite Difference Beam Propagation Method", *IEEE J. Quant. Elect.*, **26**, 1335-1339, 1990.
- [2] Scarmozzino, R. and Osgood, R. M., "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.
- [3] Hendow, S. T. and Shaker, S. A., "Recursive Numerical Solution for Nonlinear Wave Propagation in Fibers and Cylindrically Symmetric System", *Appl. Opt.*, 1759-1764, **25**, 1986.
- [4] Haung, W. P., Xu, C. L. and Chaudhari, S. K., "A Vector Beam Propagation Method for Guided-Wave Optics", *IEEE Phot. Tech. Lett.*, **3**, 910-913, 1991.
- [5] Haung, W. P., Xu, Chu, S. T., C. L. and Chaudhari, S. K., "The Finite-Difference Vector Beam Propagation Method: Analysis and Assessment", *J. Lightwave Tech.*, **10**, 295-305, 1992.
- [6] Haung, W. P., Xu, C. L. and Chaudhari, S. K., "Application of the Finite-Difference Vector Beam Propagation Method to Directional Coupler Devices", *IEEE J. Quant. Elect.*, **28**, 1527-1532, 1992.
- [7] Artiglia, M., Vita, P. D., Potenza, M., Coopa, G., Lapenta, G., and Ravetto, P., "Variable Grid Finite Difference Methods for Study of longitudinally Varying Planar Waveguides", *Elect. Lett.*, **27**, 474-475, 1991.
- [8] Accornero, R., Artiglia, M., Coopa, G., Di Vita, P. Lapenta, G., Potenza, M., and Ravetto, P., "Finite Difference Methods for the Analysis of Integrated Optical Waveguides", *Elect. Lett.*, **26**, 1959-1960, 1990.
- [9] Clauberg, R. and Von-Allmen, P., "Vectorial Beam Propagation Method for Integrated Optics", *Elect. Lett.*, **27**, 654-655, 1991.
- [10] Yevick, D., "A Guide to Electric Field Propagation Techniques for Guided-Wave Optics", *Opt. Quant. Elect.*, **26**, 185-197, 1994.
- [11] Yevick, D. and Hermansson, B., "Split-Step Finite-Difference Analysis of Rib Waveguides", *Elect. Lett.*, **25**, 461-462, 1989.
- [12] Yevick, D. and Hermansson, B., "Efficient Beam Propagation Techniques", *IEEE J. Quant. Elect.*, **26**, 109-112, 1990.
- [13] Chung, Y. and Dagli, N., "Explicit Finite Difference Beam Propagation Method: Application to Semiconductor Rib Waveguide Y-Junction Analysis", *Elect. Lett.*, **26**, 711-713, 1990.
- [14] Chung, Y. and Dagli, N., "Analysis of Z-Invariant and Z-Variant Semiconductor Rib

- Waveguides by Explicit Finite Difference Beam Propagation Method with Nonuniform Mesh Configuration", *IEEE J. Quant. Elect.*, **27**, 2296-2305, 1991.
- [15] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Methods", *IEEE Phot. Tech. Lett.*, **6**, 848-850, July 1994.
- [16] Masoudi, H. M. and Arnold, J. M., "Parallel Three-Dimensional Finite-Difference Beam Propagation Methods", *Int. J. of Num. Mod.*, **8**, 95-107, March-April 1995.
- [17] Masoudi, H. M. and Arnold, J. M., "Parallel-Processing Finite-Difference Beam Propagation Methods", presented at the *Int. Symp. on Guided-Wave Optoelec.: Device Characterization, Analysis and Design*, IEEE and Optical Society of America, in Polytechnic University, New York, October 1994.
- [18] Mitchell, A. R. and Griffiths, D. F., "The Finite Difference Method in Partial Differential Equations", John Wiley & Sons, 1980.
- [19] Goldberg, A., Schey, H. M. and Schwarz, J. L., "Computer Generated Motion Pictures of One-Dimensional Quantum-Mechanical Transmission and Reflection Phenomena", *Am. J. Phys.*, **35**, 177-186, 1967.
- [20] DeRaedt, H., "Product Formula Algorithms for Solving the Time Dependent Schrodinger Equation", *Comp. Phys. Rep.*, **7**, 1-72, 1987
- [21] Yevick, D., Rolland, C. and Hermansson, B., "Fresnel Equation Studies of Longitudinal Varying Semiconductor Rib Waveguides: Reference Wavevector Dependence", *Elect. Lett.*, **25**, 1254-1256, 1989.
- [22] Hadley, G. R., "Transparent Boundary Condition for Beam Propagation", *Opt. Lett.*, **16**, 624-626, 1991.
- [23] Hadley, G. R., "Transparent Boundary Condition for the Beam Propagation Method", *IEEE J. Quant. Elect.*, **28**, 363-370, 1992.
- [24] Milonni, P. W., Eberly, J. H., "Lasers", John Wiley & Sons, New York, 1988.
- [25] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. "The Art of scientific computing: numerical recipes in FORTRAN", Cambridge Uni. Press, 1992.
- [26] Stern, M. S., "Semivectorial Polarised Finite Difference Method for Optical Waveguides With Arbitrary Index Profiles", *IEE Proc. Pt. J.*, **135**, 56-63, 1988.
- [27] Feit, M. D. and Fleck, J. A., "Mode Properties of Optical Fibers With Lossy Components by the Propagating Beam Method", *App. Opt.*, **20**, 848-856, 1981.
- [28] Feit, M. D. and Fleck, J. A., "Computation of Mode Properties in Optical Fiber Waveguides by a Propagating Beam Method", *App. Opt.*, **19**, 1154-1164, 1980.
- [29] Strikwerda, J. C., "Finite Difference Schemes and Partial Differential Equations", Wadsworth & Books, California, 1989.
- [30] Xiang, F. and Yip, G. L., "An Explicit and Stable Finite-Difference 2-D Vector Beam Propagation Method", *IEEE Phot. Tech. Lett.*, **6**, 1248-1250, 1994.

Parallel Three-Dimensional Beam Propagation Methods

6.1 Introduction

The previous chapter showed that the three FD-BPMs are very efficient when applied to the analysis of 2-D structures, because the computation involves only one-dimensional numerical arrays. On the other hand the FD-BPMs are time-consuming when used to analyse 3-dimensional devices accurately, for several reasons [1-3]. First, the 3-D nature of the rectangular waveguides leads to a large computational problem; second, some practical devices contain multiple coupled linear and non-linear waveguides which are generally longer than several hundreds of wavelengths; third, the existence of large-contrast media will force the BPM to use small transverse mesh sizes and/or small longitudinal step sizes for convergence or stability reasons. Obviously, this will multiply the computational effort many-fold. The computational resources used (conventional serial computers) are not adequate for this kind of problem. In these circumstances, it is natural to turn to parallel processing implementations in order to achieve practical run times for realistic devices [1-3]. Already, substantial improvements in the speed of parallel over serial computation have been achieved for a number of major mathematical applications [4,5]. However, not all serial methods are readily adapted to the parallel environment and care needs to be exercised in selecting the appropriate method for parallelisation. Generally, as discussed in chapter 4, problems that can be divided into smaller sub-problems where each is spatially localised to minimise the communication cost, are most suitable for parallel implementations. For example, the 3-D FFT-BPM is inefficient in the parallel environment because it requires, for each propagational step, information from the entire spatial domain for the computation of each spatial point, and this involves extensive communication time [3]. This adds also to the poor efficiency of the FFT-BPM compared to the FD-BPMs¹.

¹ See chapters 3 and 5.

On the other hand, comparison between all of the 3-D FD-BPMs described before shows that both the RS-BPM and the EFD-BPM are explicit and therefore highly parallel due to the locality of the spatial points, which reduces the movement of data between processors [1-3, 6-8]. On the other hand, the ADI-BPM is much less efficient in the parallel domain than the explicit methods because of the required inversion of large matrices for each propagational step which is again not naturally localised [3]. Although the propagational step of the ADI-BPM is considerably larger than that of the explicit methods, the explicit methods are more efficient than the implicit method in two ways: first, they are much more efficient per propagational step in normal serial form; and second, they gain a larger speed up when run on parallel computers [1-3]. These two reasons more than compensate for the constraints imposed on both methods².

This chapter shows that the implementation of two explicit finite-difference BPMs on the transputer array and the Connection Machine (CM) speeds up the execution of these algorithms tremendously. Due to the fact that parallel computers can execute large problems rapidly, the parallel FD-BPM allows optical devices to be modelled accurately, by increasing the number of mesh points, and gives the freedom to study complicated devices that contain multiple waveguides [2].

The organisation of this chapter is as follows. Section 6.2 shows an approximate attempt to parallelise the first order ADI-BPM on the transputer array and the CM. Section 6.3 describes the details of implementing the second order EFD-BPM and the second order RS-BPM on the transputer array and the CM. In addition the section shows the results and full comparisons between the performance of these parallel methods. Section 6.4 contains the accuracy assessment of the two parallel 3-D explicit BPMs, where three different rib waveguides and three different directional couplers have been analysed and the results compared with other serial techniques. Also the section shows some simulation examples for propagation inside three-core directional couplers. Section 6.5 contains discussion and conclusions for the results of the parallel implementations of this chapter.

² Also see section 6.5 for full discussion.

6.2 The Alternating Direction Implicit Method (ADI-BPM)

6.2.1 Formulation

For simplicity, the first order ADI-BPM will be studied for the implementation on serial and parallel computers to analyse the three-dimensional waveguides of chapter 2. From Eq. 3.13, the first order three-dimensional split operator can be written as [9]

$$\phi(x, y, z + \Delta z) = e^{\left(-j\frac{\Delta z}{2a_0}\nabla_x^2\right)} e^{\left(-j\frac{\Delta z}{2a_0}\nabla_y^2\right)} e^{\left(-j\frac{\Delta z}{2a_0}d_0\right)} \phi(x, y, z) + O((\Delta z)^2) \quad (6.1)$$

Using the approximation in (3.14), we may write (6.1), after combining the second and the third terms [10], as (see also Eq. 5.3)

$$\begin{aligned} & \left\{1 + j\frac{\Delta z}{2a_0}\nabla_x^2\right\} \left\{1 + j\frac{\Delta z}{2a_0}(\nabla_y^2 + d_0)\right\} \phi(x, y, z + \Delta z) = \\ & \left\{1 - j\frac{\Delta z}{2a_0}\nabla_x^2\right\} \left\{1 - j\frac{\Delta z}{2a_0}(\nabla_y^2 + d_0)\right\} \phi(x, y, z) \end{aligned} \quad (6.2)$$

Equation (6.2) contains four steps to advance the two-dimensional profile of the field from z to $z + \Delta z$. To simplify the process of Eq. 6.2, let us write these four steps as four separate equations to be executed in the following order

$$\phi^{z^*}(x, y) = \left\{1 - j\frac{\Delta z}{2a_0}(\nabla_y^2 + d_0)\right\} \phi(x, y, z) \quad (6.3a)$$

$$\Omega^z(x, y) = \left\{1 - j\frac{\Delta z}{2a_0}\nabla_x^2\right\} \phi^{z^*}(x, y) \quad (6.3b)$$

$$\left\{1 + j\frac{\Delta z}{2a_0}\nabla_x^2\right\} \phi^{z+\Delta z^*}(x, y) = \Omega^z(x, y) \quad (6.3c)$$

$$\left\{1 + j\frac{\Delta z}{2a_0}(\nabla_y^2 + d_0)\right\} \phi^{z+\Delta z}(x, y) = \phi^{z+\Delta z^*}(x, y) \quad (6.3d)$$

Applying the central difference approximation to the partial derivatives in (6.3) leads to the following four discrete equations

$$\phi_{i,m}^{z*} = -b_y \phi_{i,m+1}^z + c_{i,m}^R \phi_{i,m}^z - b_y \phi_{i,m-1}^z \quad (6.4a)$$

$$\Omega_{i,m}^z = -b_x \phi_{i+1,m}^{z*} + (1 + 2b_x) \phi_{i,m}^{z*} - b_x \phi_{i-1,m}^{z*} \quad (6.4b)$$

$$b_x \phi_{i+1,m}^{z+\Delta z*} + (1 - 2b_x) \phi_{i,m}^{z+\Delta z*} + b_x \phi_{i-1,m}^{z+\Delta z*} = \Omega_{i,m}^z \quad (6.4c)$$

$$b_y \phi_{i,m+1}^{z+\Delta z} + c_{i,m}^L \phi_{i,m}^{z+\Delta z} + b_y \phi_{i,m-1}^{z+\Delta z} = \phi_{i,m}^{z+\Delta z*} \quad (6.4d)$$

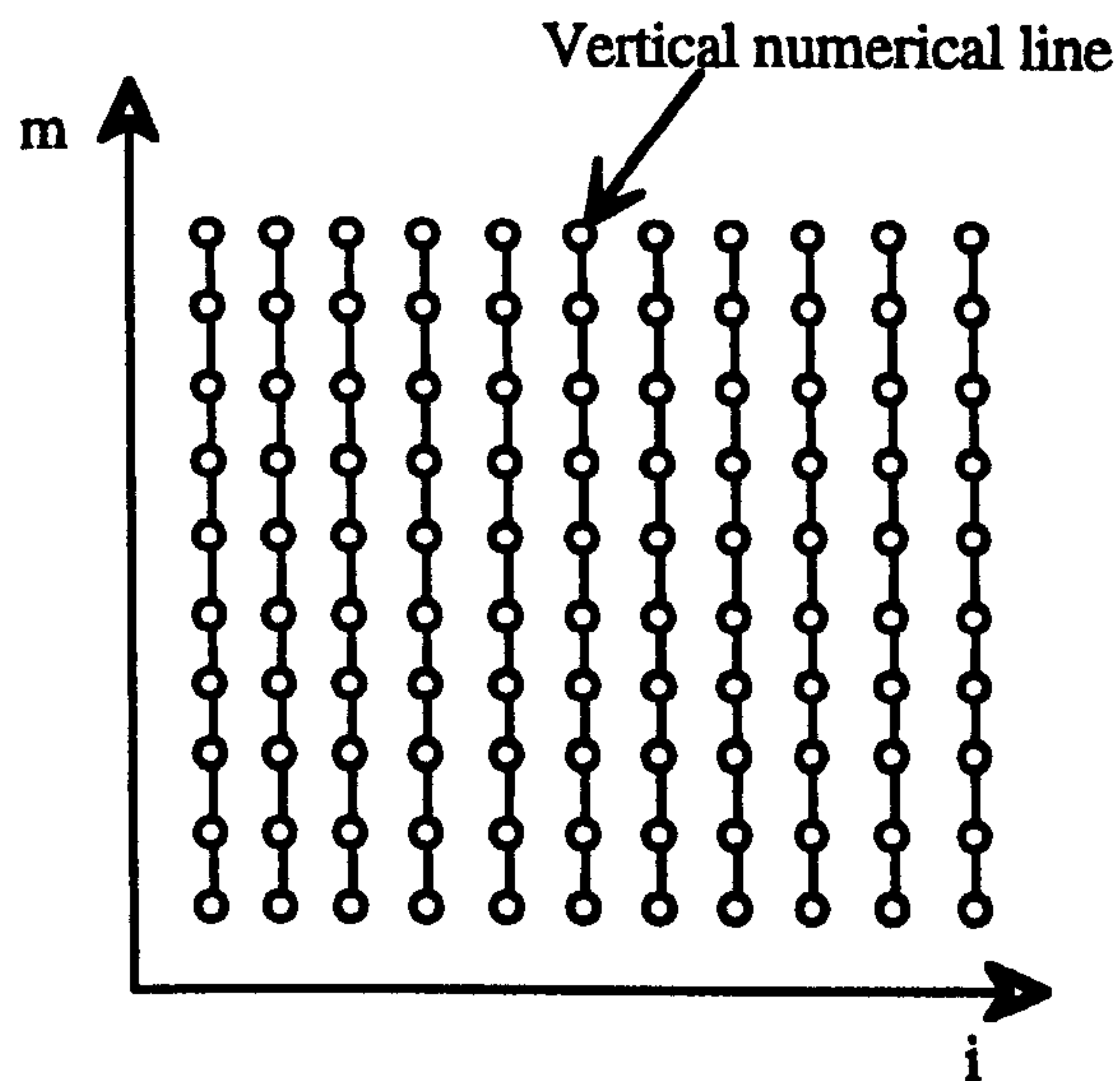
where

$$b_\alpha = \frac{j\Delta z}{2a_\alpha \Delta \alpha^2}, \quad (\alpha = x, y)$$

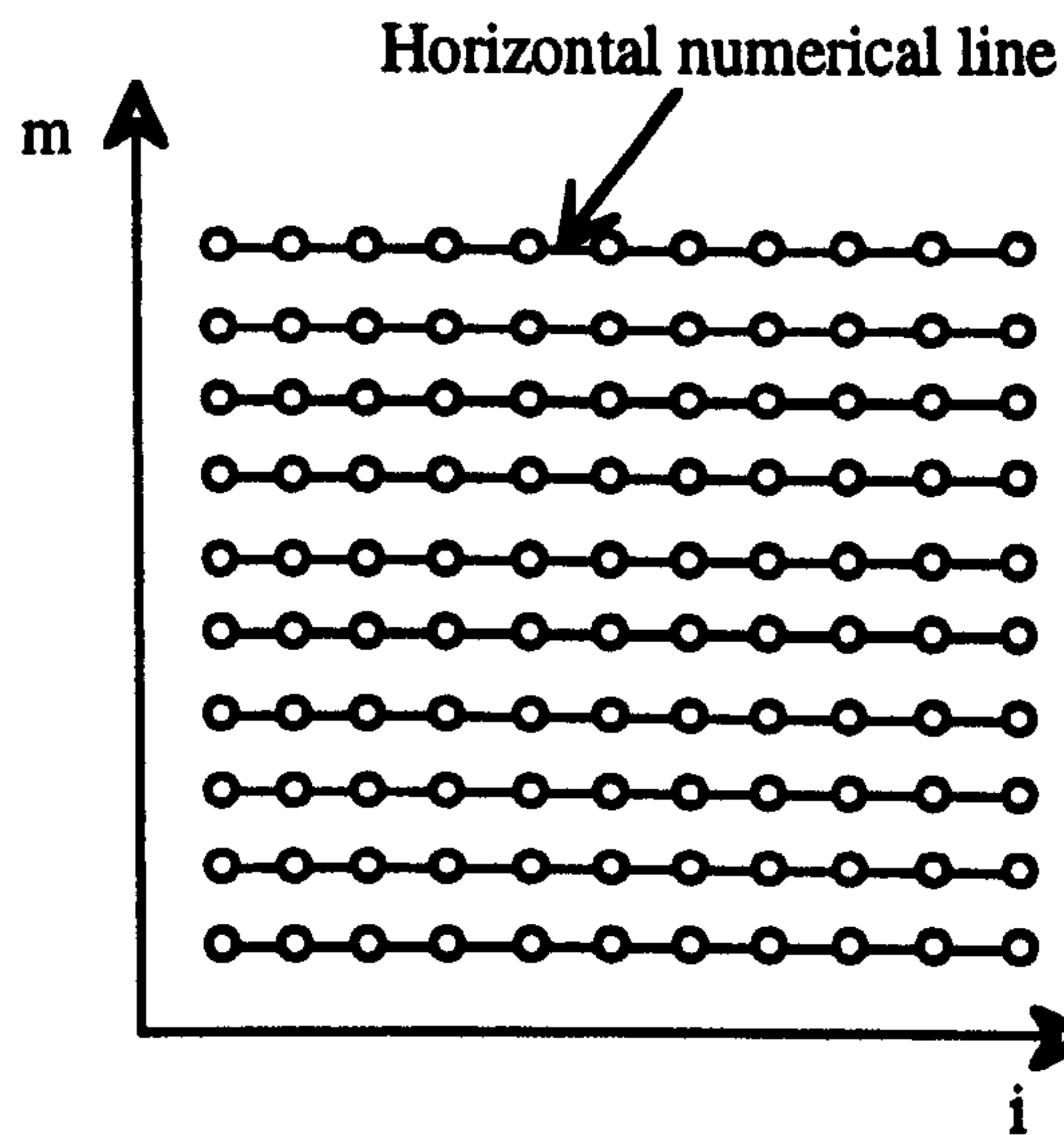
$$c_{i,m}^R = 1 + 2b_y - j \frac{\Delta z}{2a_o} d_o^{i,m}$$

$$c_{i,m}^L = 1 - 2b_y + j \frac{\Delta z}{2a_o} d_o^{i,m}$$

In the above derivations, zero boundary conditions have been assumed at the edges of the computational window. The numerical implementation of the first two equations (6.4a and 6.4b) is straightforward because it involves direct multiplication of the fields with tridiagonal matrices. On the other hand, the execution of Eqs. 6.4c and 6.4d involves inversion of matrices for each propagational step. It should be noticed from the implementation of 6.4 that the splitting of the operators into x and y dependence means that the operation for computing the variation of the field in one direction can be achieved by fixing the variation in the other direction. In other words, the numerical execution of Eqs. 6.4a and 6.4d, which represent the y-variation of the field, are computed by fixing i for all values of m forming independent numerical lines from the y perspective (see Fig. 6.1a). Similarly, Fig. 6.1b shows the independent numerical lines for computing Eqs 6.4b and 6.4c which are the x-variation of the field. From the two figures, we may observe that the computation of each numerical line is equivalent to the numerical computation of the two-dimensional CN-BPM. Thus we may imagine the computation of the split operator of the three-dimensional ADI-BPM as the computation of multiple $(M_x + M_y)$ two-dimensional CN-BPMs for each propagational step.



(a)



(b)

Fig. 6.1 The numerical execution of Eq. 6.4 by forming independent numerical lines in the x and y directions. a) The independent lines for computing Eqs. 6.4a and 6.4d. b) The independent lines for computing Eqs. 6.4b and 6.4c.

In the following implementation of the first order ADI-BPM in Eq. 6.4, the recursive technique in section 5.2 has been used to solve the tridiagonal system of equations in (6.4d) where it has been used for every numerical line of Fig. 6.1a. On the other hand, we may observe from Eq. 6.4c, for a fixed b_x , that the matrices involved are the same and the solution requires the multiplication of $\Omega_{i,m}^2$ with the inverse of one of the matrices. For simplicity, the solution of the tridiagonal matrices of Eq. 6.4c are computed by inverting

one of the matrices, using a simple inversion routine [11], and the inverted matrix is stored and used for multiple steps propagation.

6.2.2 Parallel implementations

The first order ADI-BPM has been implemented on the parallel machines discussed in chapter 4. In order to implement the ADI-BPM algorithm on the transputer array, an efficient topology to connect the processors has to be selected. Fig. 6.2 shows the transputer topology used in the computations. It should be pointed out that the choice of the topology in Fig. 6.2 is made because of its simplicity and flexibility to program the ADI-BPM on the transputer array, where as other topologies may lead to a better performance. Processor 0 in Fig. 6.2 is an interface processor used for user information, input data and to facilitate file operations; this processor is not used for computations and can be ignored in the following discussion.

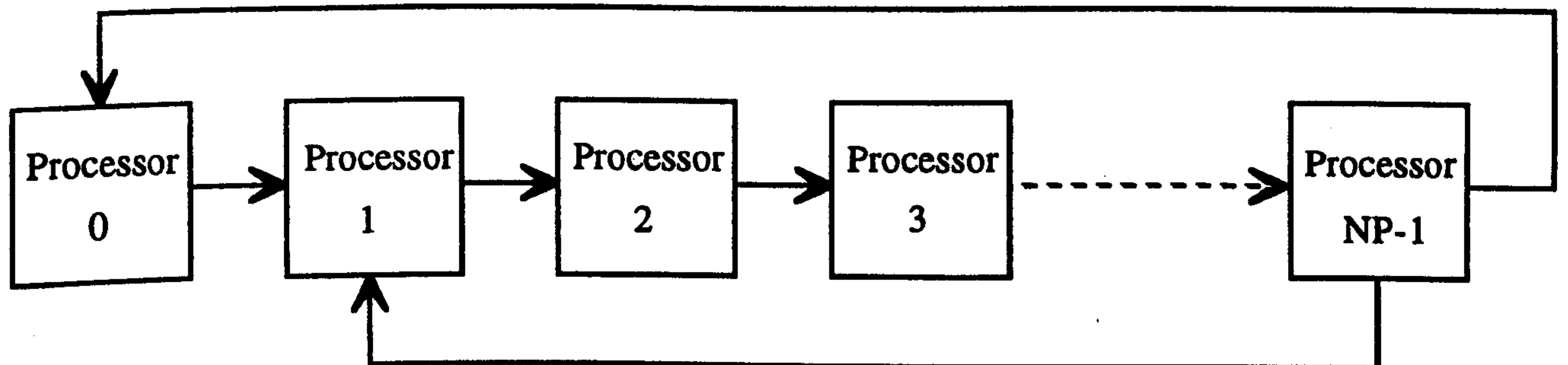
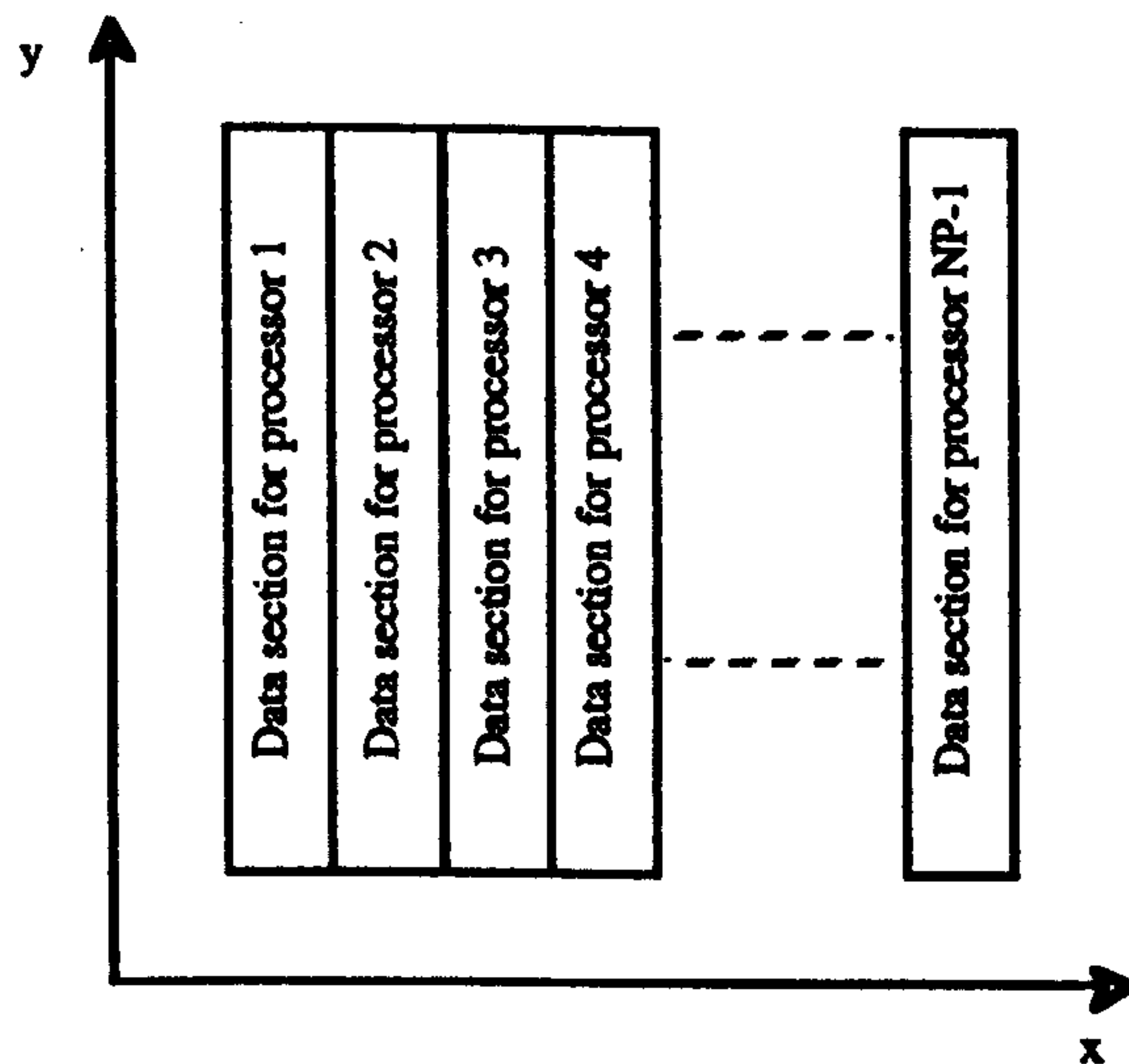


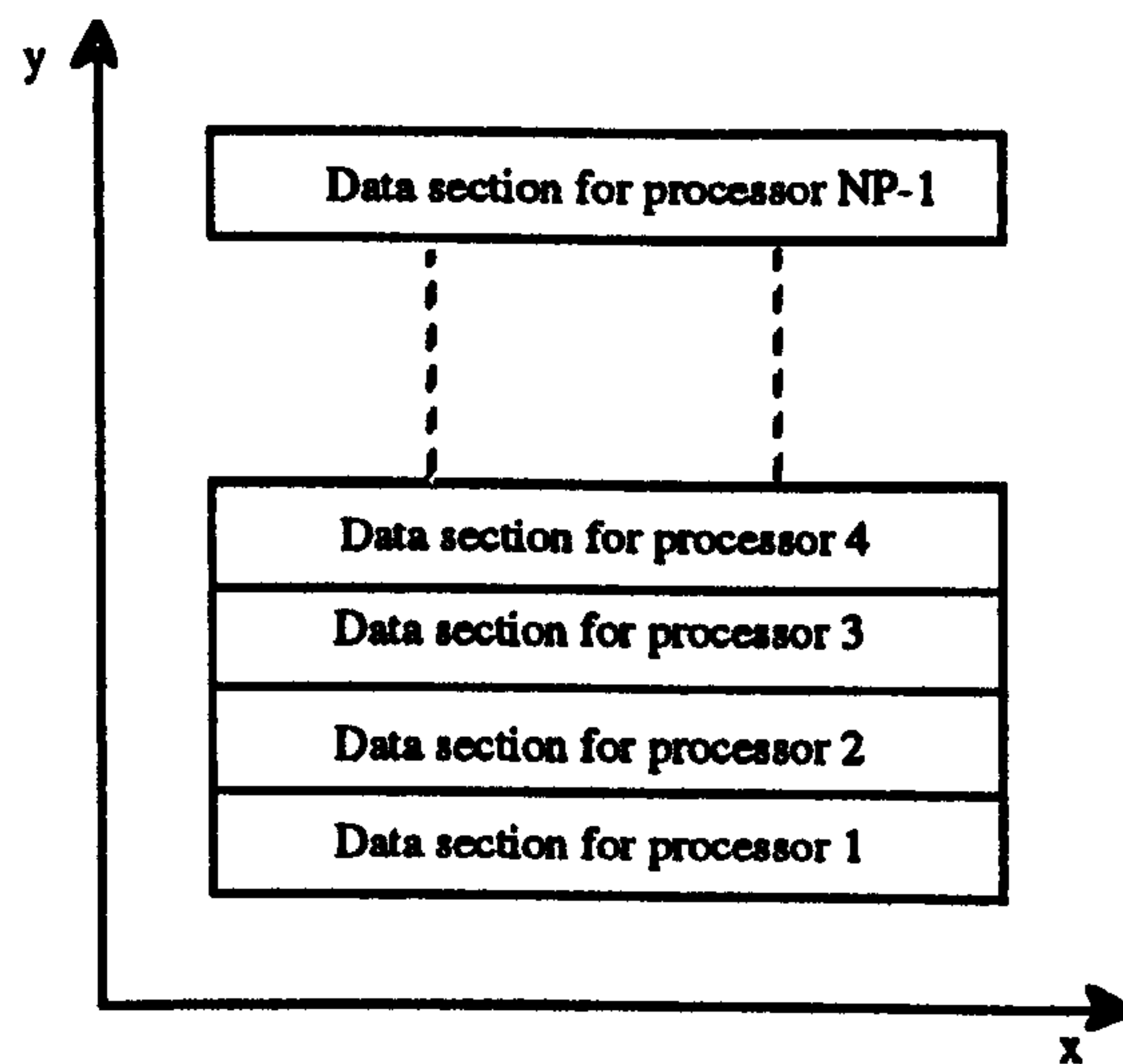
Fig 6.2 The transputer topology used in the implementation of the parallel ADI-BPM where NP stands for the total number of processors.

The idea of the parallel implementation is to distribute the work load of the numerical computations of the four stages in Eq. 6.4 to all processors equally. Before executing the propagation process the matrix in Eq. 6.4c is inverted and stored in the memory of each processor of the topology in Fig. 6.2. To execute the computation in Eq. 6.4a, the discretised initial field $\phi_{i,m}^z$ has to be distributed to all processors. In the topology of Fig. 6.2, processor 1 is called the "master" processor where all the operations begin and end; all the other processors are called the "slaves" and they perform exactly the same instructions. The process of broadcasting the numerical array of the initial field to all processors is executed using send and receive commands. For example, processor N receives the array from processor N-1 and sends a copy of that array to processor N+1. Since Eq. 6.4a can be computed using the idea of independent numerical lines of Fig. 6.1a, each processor can compute a number of these lines concurrently. The 2-D transverse mesh

has been divided horizontally, as shown in Fig. 6.3a, into a number of sections where each section, having $M_x/(NP-1)$ points, is assigned to one processor for the computation of $\phi_{i,m}^{z^*}$ in Eq. 6.4a. This arrangement ensures the overall balance of the system. Then the master processor (1) passes an array, using the process of send and receive, to collect all the computed sections of the array and place them into one array. The process of collecting all the sections of the array is necessary because the next step involves the computation in the other direction. To compute Eq. 6.4b, the collected field $\phi_{i,m}^{z^*}$ is distributed to all processors. Similar to the previous computation of Eq. 6.4a, the 2-D transverse mesh has been divided this time vertically, as shown in Fig. 6.3b, into equal sections of data where each section has $M_y/(NP-1)$ points and each data section has been assigned to one processor for the computation of $\Omega_{i,m}^z$ in Eq. 6.4b. The computation of Eq. 6.4c does not need any movement of data because the local points of Eqs. 6.4b and 6.4c are the same; then the execution of Eq. 6.4c follows immediately after the completion of computing Eq. 6.4b. As mentioned before, this step involves the multiplication of the previously inverted matrix with $\Omega_{i,m}^z$ to find $\phi_{i,m}^{z+\Delta z^*}$. Finally, the computation of Eq. 6.4d is exactly the same as for the computation of Eq. 6.4a in terms of local spatial points. All parts of the computed $\phi_{i,m}^{z+\Delta z^*}$ have been collected into one array, and this array is distributed to all processors for the computation of $\phi_{i,m}^{z+\Delta z}$ in Eq. 6.4d. To summarise the previous process of parallelising the first order ADI-BPM, Fig. 6.4 shows a flow chart describing all the implemented parallel steps.



(a)



(b)

Fig. 6.3 The division of the two-dimensional transverse mesh points of the field into equal sections where each section is assigned to a separate computer. a) For the computation of the y-variation operator (Eqs. 6.4a and 6.4d). b) For the computation of the x-variation operator (Eqs. 6.4b and 6.4c).

A computer code for the previous parallel ADI-BPM has been implemented on the transputer array. The code was written such that the number of processors and the number of mesh points in both directions are variables to test the performance of the parallel algorithm by varying these parameters. In addition, the ADI-BPM of Eq. 6.4 has been also implemented on the Connection Machine of chapter 4. First, as mentioned before, the inverse of the matrix in Eq. 6.4c is computed in the front-end computer serially and the result transformed to the CM memory. Eqs. 6.4a and 6.4b are simple complex number multiplications and are executed numerically using a single instruction mechanism. In Eq.

6.4a, each element of $c_{i,m}^R$ and its corresponding element from $\phi_{i,m}^z$ forced to reside in the same memory of the CM to eliminate data movement. The execution of Eq. 6.4c has been made using global CM matrix multiplication, by multiplying the inverted matrix with $\Omega_{i,m}^z$ to find $\phi_{i,m}^{z+\Delta z^*}$. Finally, the solutions of the independent numerical lines of Eq. 6.4d are computed simultaneously using the mechanism of a single instruction; however the solution of each line is computed serially, since the recursive technique used is a serial method. Also the CM computer code is written such that the number of mesh points in both directions are variables.

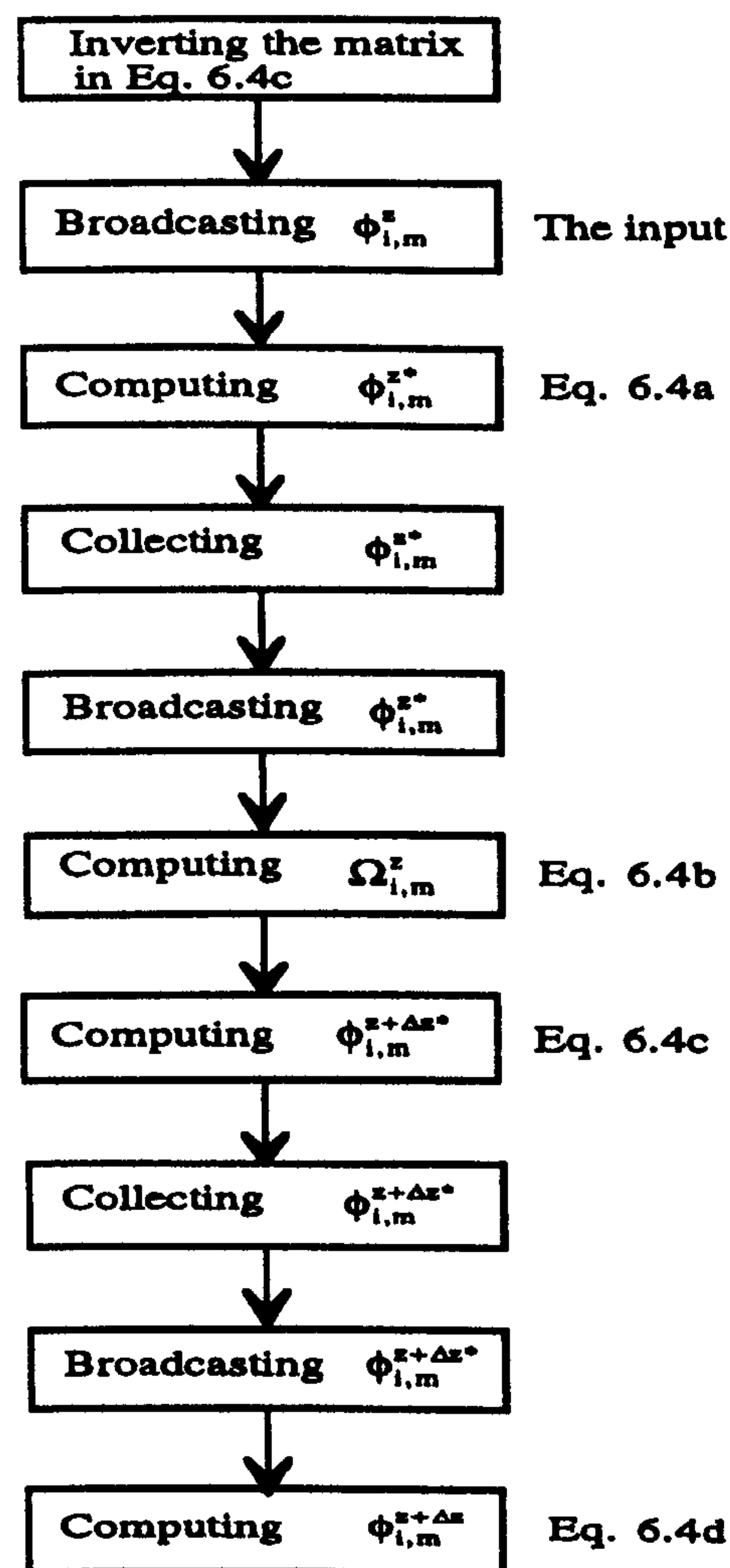


Fig. 6.4 Flow chart for the mechanism of parallelising the first order ADI-BPM of Eq. 6.4 on the transputer array using the topology in Fig. 6.2.

6.2.3 Speed

In order to show the performance of the two parallel implementations of the ADI-BPM, the algorithm has also been implemented serially using one processor of the transputer array. Fig. 6.5 shows the total (computation + communication) execution time per propagational step of the serial, the transputer array and the CM implementations of the first order ADI-BPM versus the number of mesh points in one of the transverse directions M_x ($M_y = M_x$). The two curves marked with "CM-8k" and "CM-16k" mean that half and full CM resources have been used, respectively. We can notice in Fig. 6.5 that the serial and the transputer plots do not contain results exceeding $M_x = 140$, due to memory limitations of the computer. The figure shows that as the number of processors increases the speed of the transputer implementation increases until $NP = 11$ with an optimum speed up of around five times the serial implementation when $M_x = 140$. Increasing the number of processors to more than eleven processors results in a decrease in the speed of the algorithm and this is due to the increase in the communication time between processors. The figure also shows that the CM implementations are far faster than the transputer implementation because of the larger capability of the CM compared to the transputer array.

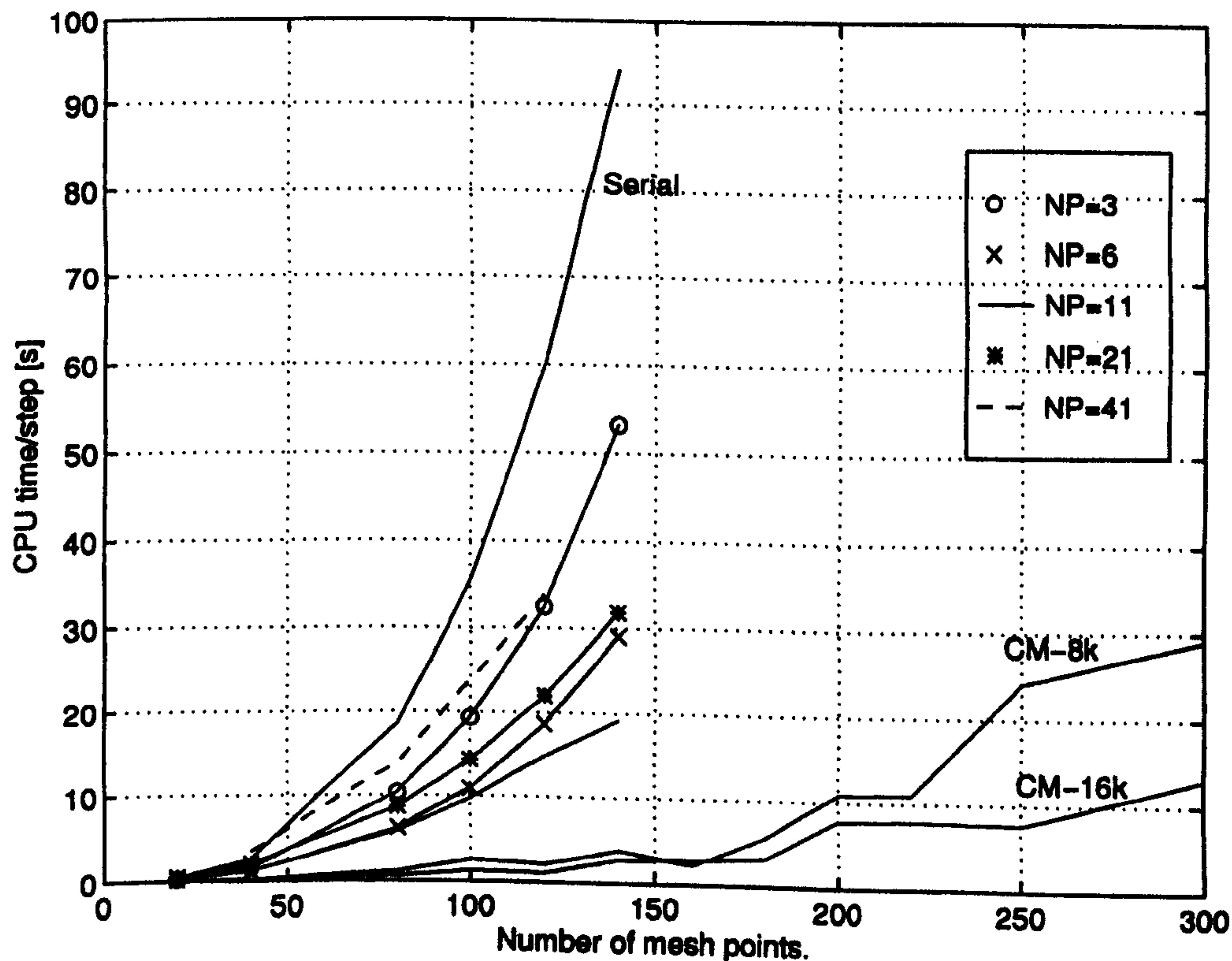


Fig. 6.5 Comparison between the speed, per propagational step, of the serial and the two parallel implementations of the transputer array and the CM for the first order ADI-BPM versus the number of mesh points in one of the transverse directions with $M_x = M_y$ (see text for details).

6.3 The Explicit Finite Difference Methods

6.3.1 Equations

In this section the three-dimensional second order equations of the RS-BPM in Eq. 3.26 and the EFD-BPM in Eq. 3.28 will be used for the parallel implementations. As a matter of convenience these equations are repeated below to show their implementations on the parallel computers. The RS-BPM equation is

$$\phi(x, y, z + \Delta z) = T_y T_x e^{\left(-j \frac{\Delta z}{a_0} U\right)} T_y T_x \phi(x, y, z) \quad (6.5)$$

and the EFD-BPM equation is

$$\begin{aligned} \phi_{i,m}(z + \Delta z) = & \phi_{i,m}(z - \Delta z) - \frac{2j\Delta z}{a_0 \Delta x^2} [\phi_{i-1,m}(z) + \phi_{i+1,m}(z)] \\ & - \frac{2j\Delta z}{a_0 \Delta y^2} [\phi_{i,m-1}(z) + \phi_{i,m+1}(z)] - \frac{2j\Delta z}{a_0} U_{i,m} \phi_{i,m}(z) \end{aligned} \quad (6.6)$$

where

$$\begin{aligned} T_y &= e^{(ps_y^0)} e^{(qs_y^0)} e^{(ps_y^0)} \\ T_x &= e^{(ps_x^0)} e^{(qs_x^0)} e^{(ps_x^0)} \\ U(x, y, z) &= d_0(x, y, z) - \frac{2}{\Delta x^2} - \frac{2}{\Delta y^2} \\ q &= -j \frac{\Delta z}{2a_0} = 2p \end{aligned}$$

It is to be noticed that the implementation of the above two equations is much simpler than the ADI-BPM. We may also observe from these equations, as for the ADI-BPM, that the idea of independent vertical and horizontal numerical lines of Fig. 6.1 is also applicable. However the computation of each numerical line consists of computing independent numerical spatial points. For the RS-BPM, the computation of every two spatial points from each numerical line is independent (see Fig. 5.1) and for the EFD-BPM, the computation of every point is independent (see Fig. 5.2).

6.3.2 Parallel implementations

In order to implement the explicit methods on the transputer cluster, a topology for connecting the processors has to be carefully selected to ensure that maximum efficiency is gained from parallelising these methods. The 2-D grid topology shown in Fig. 6.6 has been chosen for the implementation of both the EFD-BPM and the RS-BPM on the transputer array [1, 2]. It is to be noticed that this topology is the best arrangement to parallelise these algorithms in terms of efficiency and transputer memory distribution. Comparison between the 2-D topology and other topologies (e.g. Ring, Farm), for implementing the two explicit FD-BPMs, will show that the choice of the topology, in Fig. 6.6, reduces the communication overhead between processors to its minimum which is a very important factor in parallel implementations. Excluding the processors at the borders of the topology shown, each processor has four links connected to its neighbours, where each link is a bi-directional communication channel for exchanging information. In the figure, $\text{dim}(x)$ and $\text{dim}(y)$ are the number of processors in the x and y direction, respectively, of the 2-D grid topology.

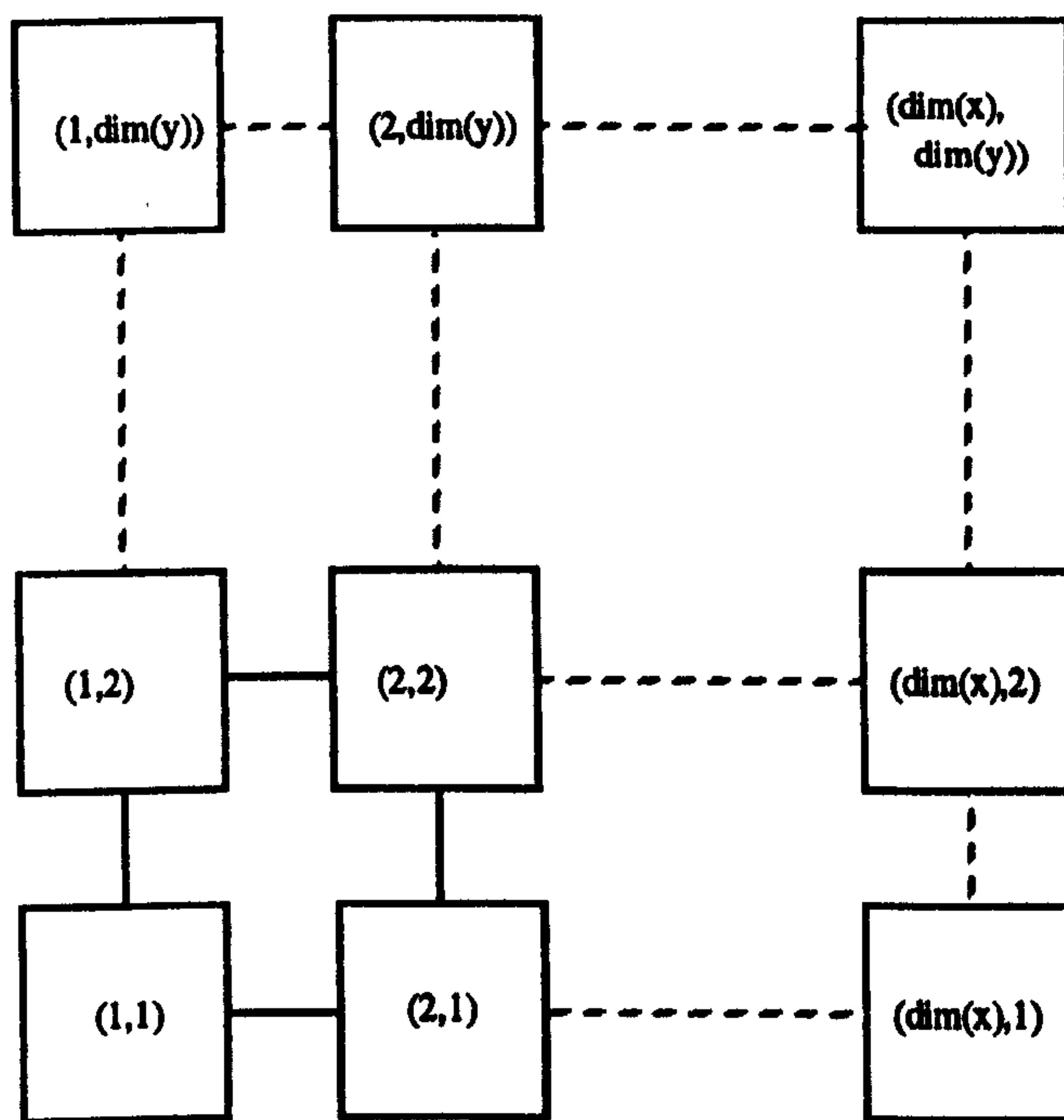


Fig. 6.6 The 2-D grid topology used for the implementation of both the parallel EFD-BPM and the parallel RS-BPM. The number shown indicates the position of each processor in terms of the 2-D grid.

Both the EFD-BPM and the RS-BPM have been implemented on the 2-D grid topology, by dividing the total 2-D transverse mesh into 2-D identical blocks of mesh points (see Fig.

6.7) where the size of each block is equal to $(M_x/\text{dim}(x), M_y/\text{dim}(y))$. Each block of data is assigned to one processor for computation [1, 2].

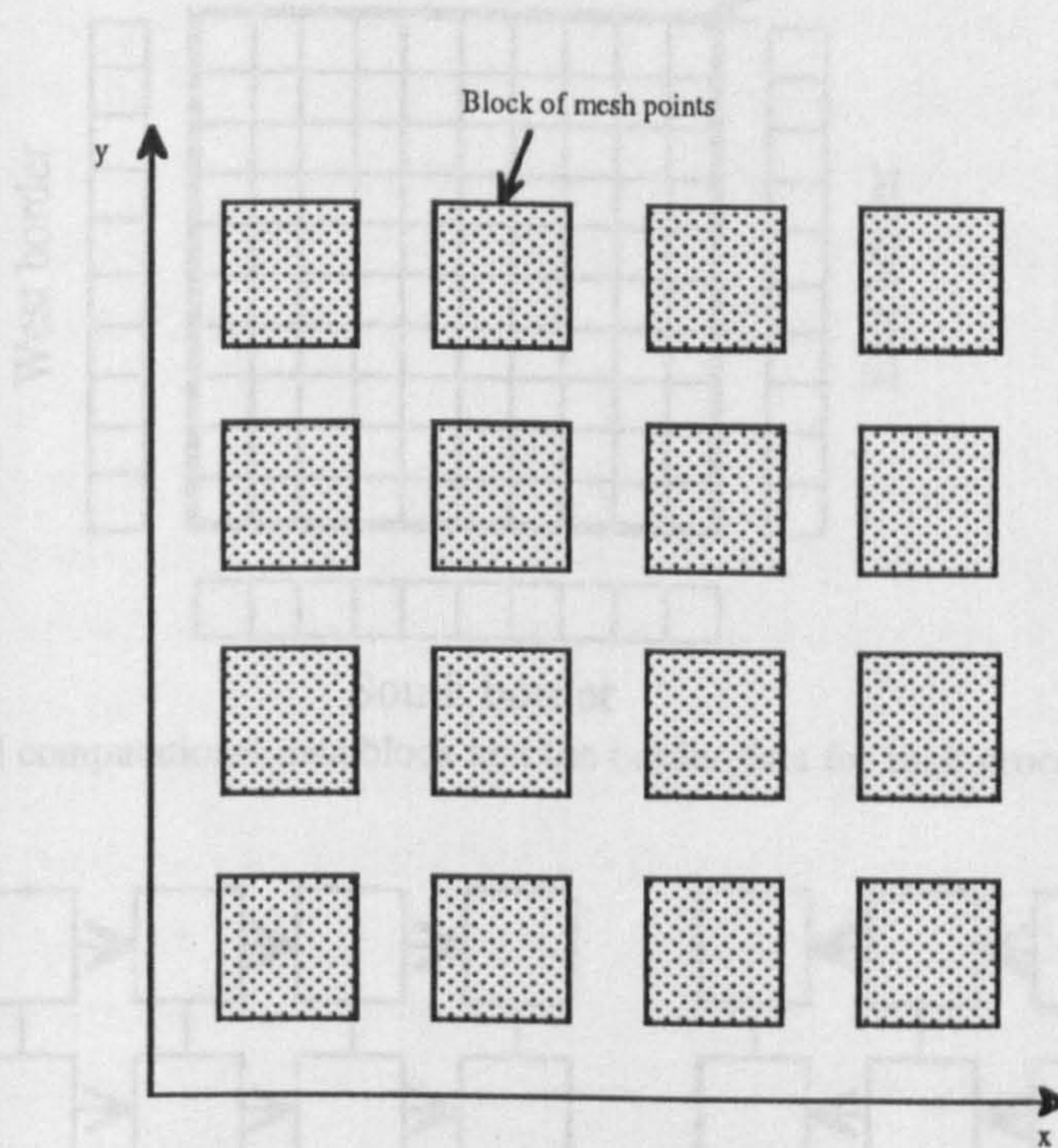


Fig. 6.7 The division of the total transverse mesh points of the EFD-BPM and the RS-BPM into equal blocks of computational data for the implementation on the transputer array.

The arrangement of the 2-D topology in Fig. 6.6 and the 2-D blocks of data in Fig. 6.7 will ensure that all processors carry out equal amounts of computation, without the need to load-balance the transputer system. In addition, it gives the freedom to change both the number of processors and the number of mesh points without altering the parallel computer code. For every propagational step the processors need to exchange the local mesh points at the border of each computational block. The process of transferring the local border data to the neighbouring process is required only once for the EFD-BPM and twice for the RS-BPM one for each of the spatial variations ($T_x T_y$) of Eq. 6.5. Each processor, as shown in Fig. 6.8, has four numerical arrays to store the local border data in addition to the main numerical array for the computation of local data block. These four local border arrays are transferred to all four neighbouring processors using the mechanism of send and receive between processors where four separate subroutines to execute the communication process have been used. Fig. 6.9 shows the mechanism of transferring the four local border arrays where the arrows refer to send and receive commands.

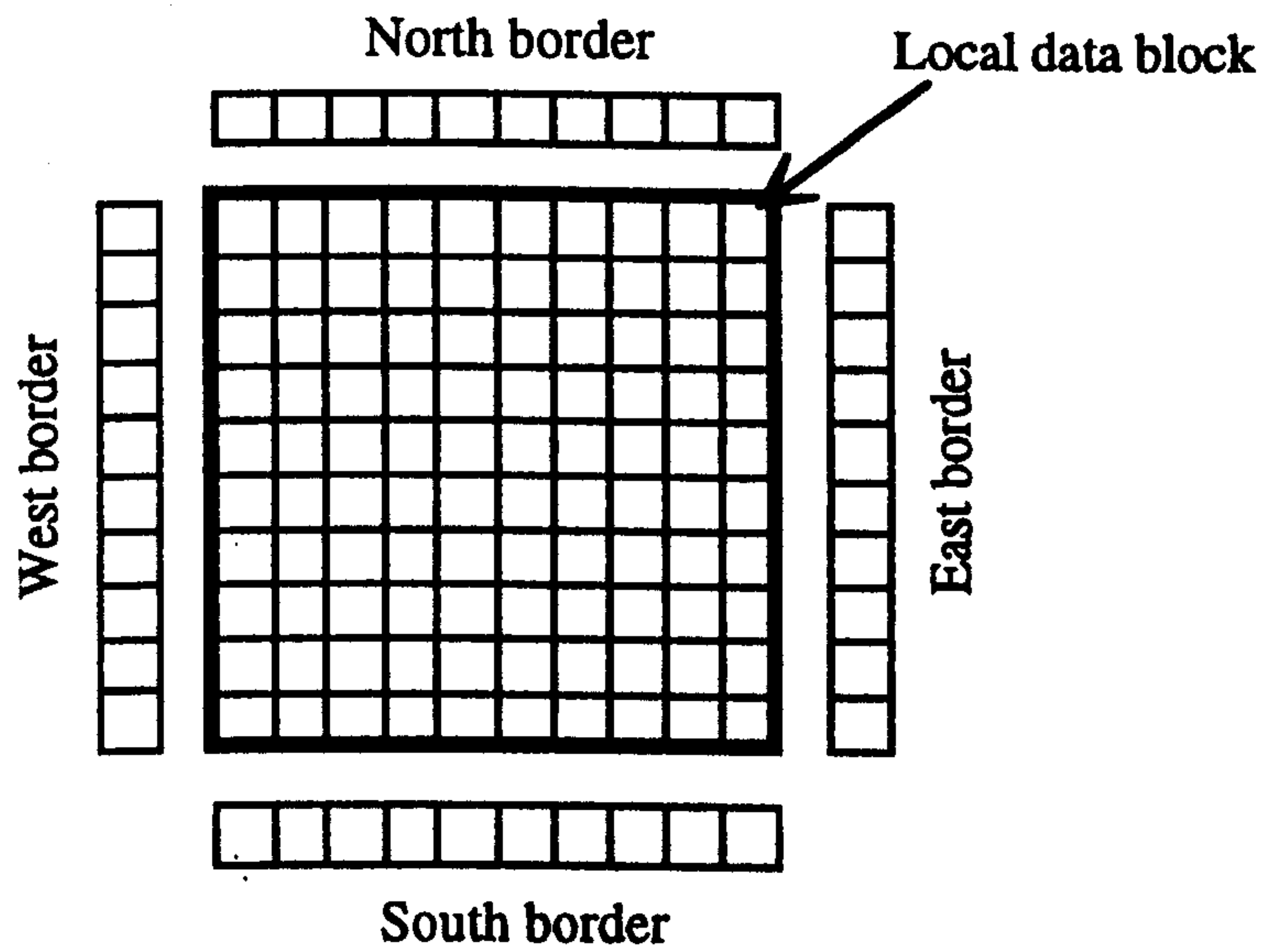


Fig. 6.8 The local computational data block and the border data for each processor of the 2-D grid topology of Fig. 6.6.

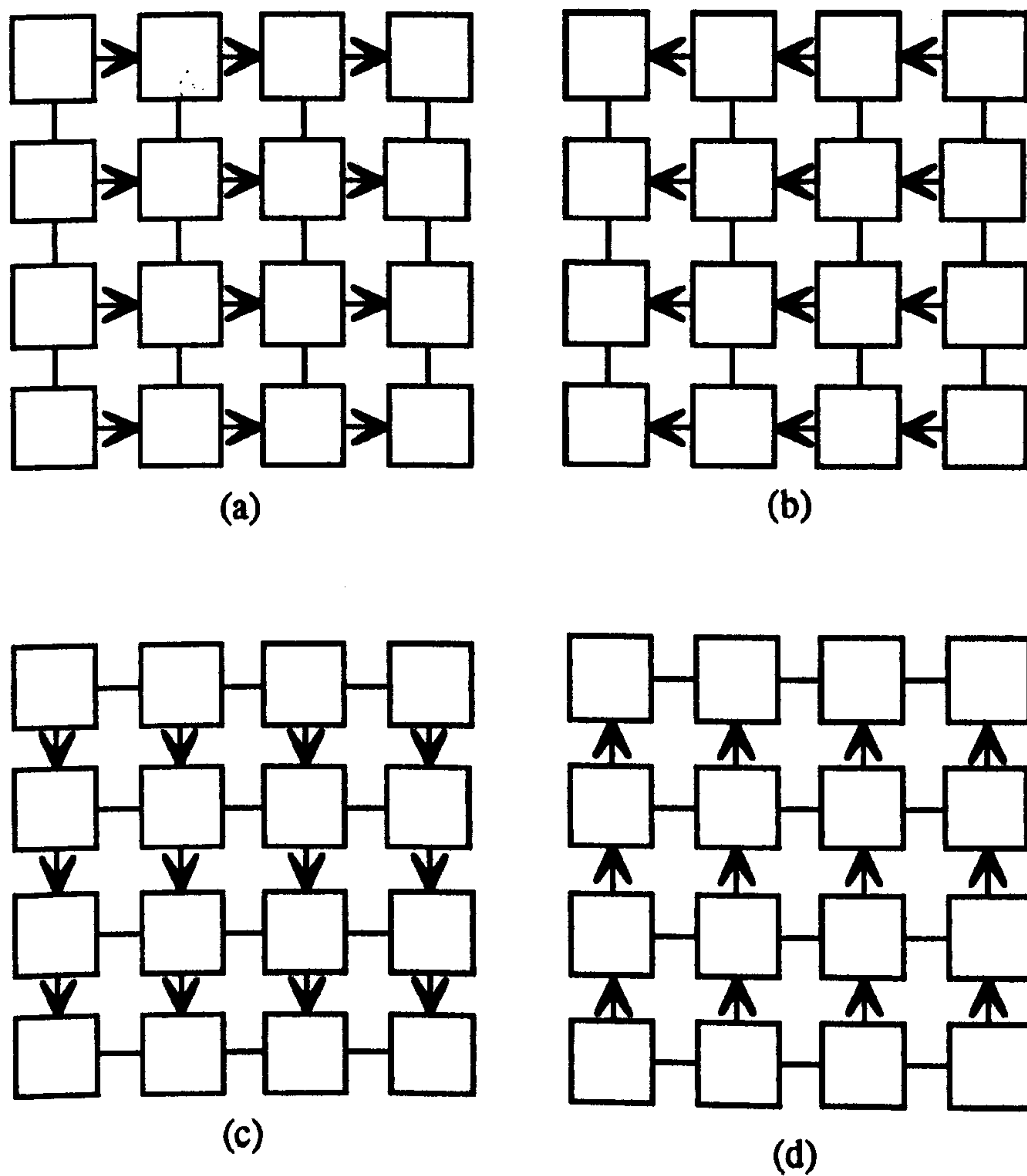


Fig. 6.9 The communication mechanism for transferring the border data arrays to neighbouring processors of the 2-D grid topology for the execution of each propagational step of the parallel EFD-BPM and the parallel RS-BPM; a) local east border transfer, b) local west border transfer, c) local south border transfer and d) local north border transfer.

These communication instructions were written such that each instruction is executed simultaneously along the vertical (of Figs. 6.9a and 6.9b) or the horizontal (of Figs. 6.9d and 6.9c) lines of the 2-D topology of the processors. For example, to transfer the east border arrays of Fig. 6.9a, the process starts from the far left processors, along the vertical line of the 2-D grid, to the far right processors where all vertical instructions are executed in parallel. The same has been done for the other three border arrays (see Fig. 6.9).

In order to study the gain in speed by using the transputer array both, of the explicit methods were also implemented on a single processor (serial computation) of the transputer in addition to the parallel implementations. On the other hand the implementation of the two explicit methods on the Connection Machine of chapter 4 was performed using the single instruction mechanism described before. One global parallel instruction is needed to execute the EFD-BPM equation in (6.6). The execution of the RS-BPM, however, consists of thirteen serial stages three stages: for each T_x or T_y of Eq. 6.5 and one stage for the middle operator that contains the refractive index profile. Each stage was executed using a global parallel instruction [1, 2]. All computer codes were written in FORTRAN with a double precision accuracy³.

6.3.3 Speed

The computer codes of the explicit methods have been tested to analyse practical rectangular waveguides⁴. In this section, the speeds of both parallel explicit methods are examined. To concentrate on the efficiency issue of parallel machines, the number of mesh points in both directions is set to be equal ($M_x = M_y$) with uniform grid spacing (Δx and Δy). The best way to compare between the two explicit methods and between the performance of the two machines is by computing the total CPU time per propagational step for different numbers of transverse mesh points and by changing the number of processors as well. Fig.6.10 (in log-log scale) shows the total CPU time of both the second order EFD-BPM and the second order RS-BPM per propagational step versus the number of mesh points in one of the transverse direction x [1, 2]. The figure shows both the serial and the transputer results in addition to the CM results for comparison. In the figure, 'serial' means a single processor and '2 x 2' means that four processors of the transputer array have been used in the computations; other notations can be understood accordingly. 'CM-8k' and

³ More details are in chapter 4.

⁴ See the following section for this analysis.

'CM-16k' means that half and full CM resources have been used, respectively. It can be seen from the figure that for a fixed M_x the speed of both methods increases as the number of processors of the transputer increases. On the other hand, generally the CM speed is faster than the best performance of the transputer array. For example at $M_x = 400$, the CM-16k is faster than the 64 processors of the transputer array by 11 times for the EFD-BPM and 6.8 times for the RS-BPM. We can notice in Fig. 6.10 that the serial plot does not contain results exceeding $M_x = 240$ for the EFD-BPM and $M_x = 250$ for the RS-BPM, due to memory limitations of the computer.

To assess the performance of the parallel implementations, let us define the following terms:

$$\text{Speed up} = \frac{\text{Serial Speed}}{\text{Parallel Speed}} \quad (6.7)$$

and

$$\text{Efficiency} = \frac{\text{Speed up}}{\text{Number of processors}} \quad (6.8)$$

It has to be noticed that the efficiency calculation, defined by Eq.6.8, is only for the transputer array results; it is meaningless to do that for the Connection Machine because the processors are different from those of the transputer array. Fig. 6.11 shows the speed up for both of the explicit methods. The figure shows that the speed up for both algorithms, for a fixed number of mesh points, increases as the number of processors increases. At $M_x = 240$, the speed-up factors for the EFD-BPM when using the full transputer size (8 by 8) and the full CM resources (CM-16k) are around 54 and 547.3 respectively, while the speed-up of the RS-BPM when using 8 by 8 of the transputer array and the CM-16k is around 60 and 369 respectively. The other performance indicator of the transputer, the efficiency, is shown in Fig. 6.12 for both of the explicit methods. It can be seen from the figure, for both methods, that for a fixed number of processors the efficiency increases as the number of mesh points increases. For $M_x = 240$ the percentage efficiencies of the EFD-BPM and RS-BPM are around 98% and 100%, respectively, when using four processors. On the other hand, at the same number of mesh points ($M_x = 240$) but using 64 processors, the percentage efficiencies of the EFD-BPM and the RS-BPM are around 84% and 94% respectively. It can be also seen from Fig. 6.12 that the efficiency of the RS-BPM is always higher than that of the EFD-BPM. The reason for this is that the ratio of the computational time to the communicational time between processors for the RS-BPM is higher than that of the EFD-BPM.

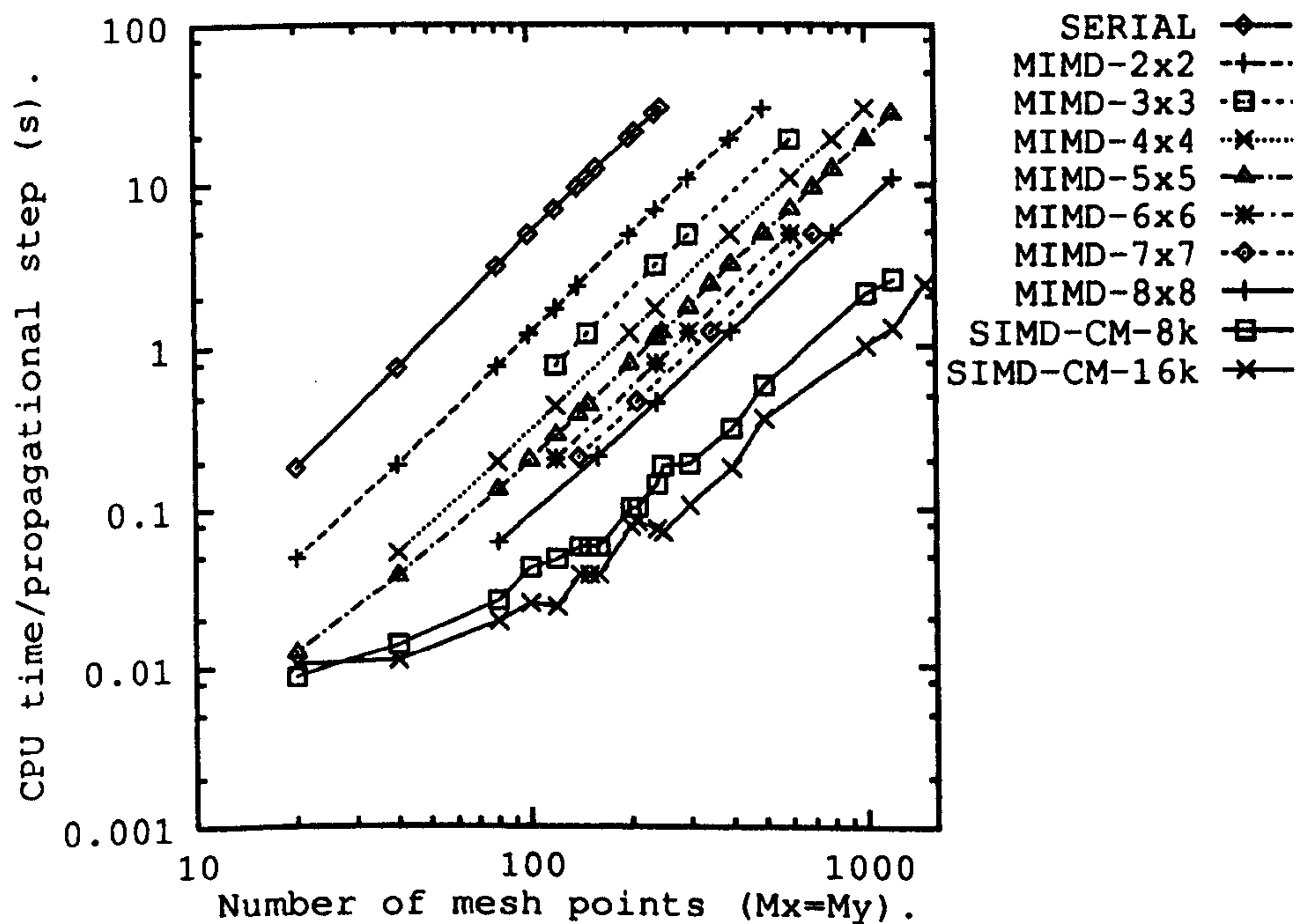
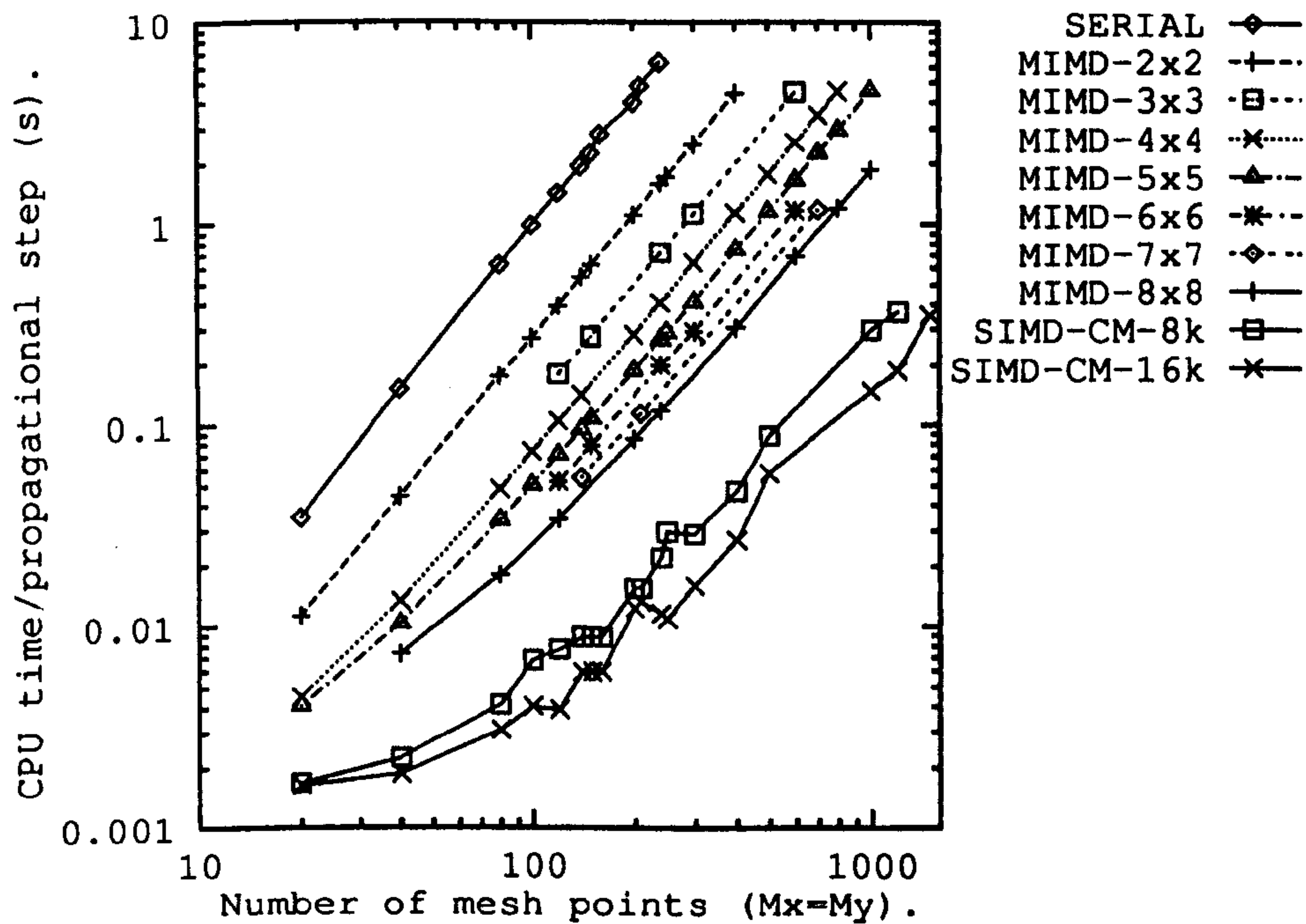


Fig. 6.10 Comparison between the speed of the serial and the parallel implementations of the explicit methods using the transputer array (MIMD) and the Connection Machine (SIMD-CM). Top) the EFD-BPM; Bottom) the RS-BPM.

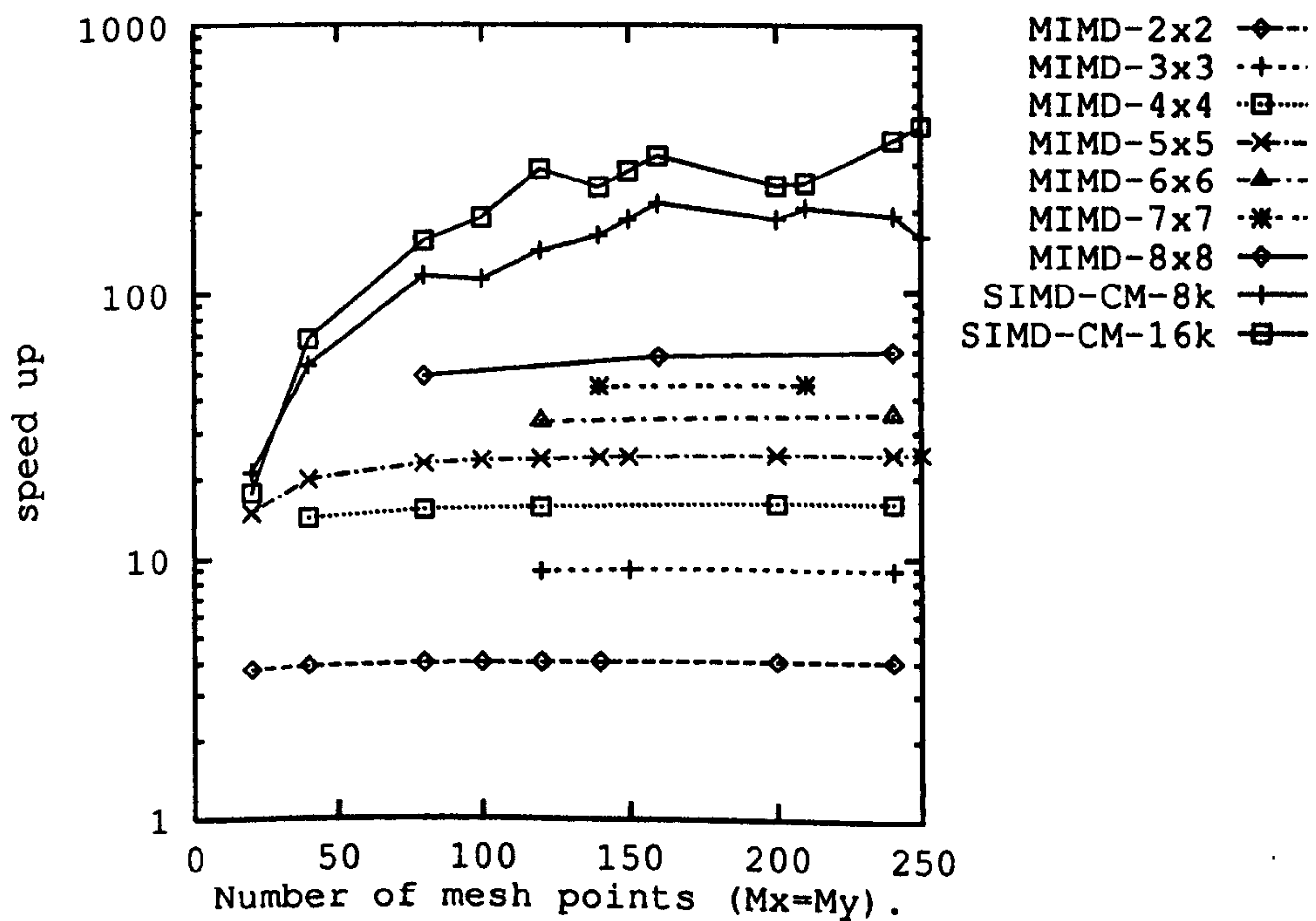
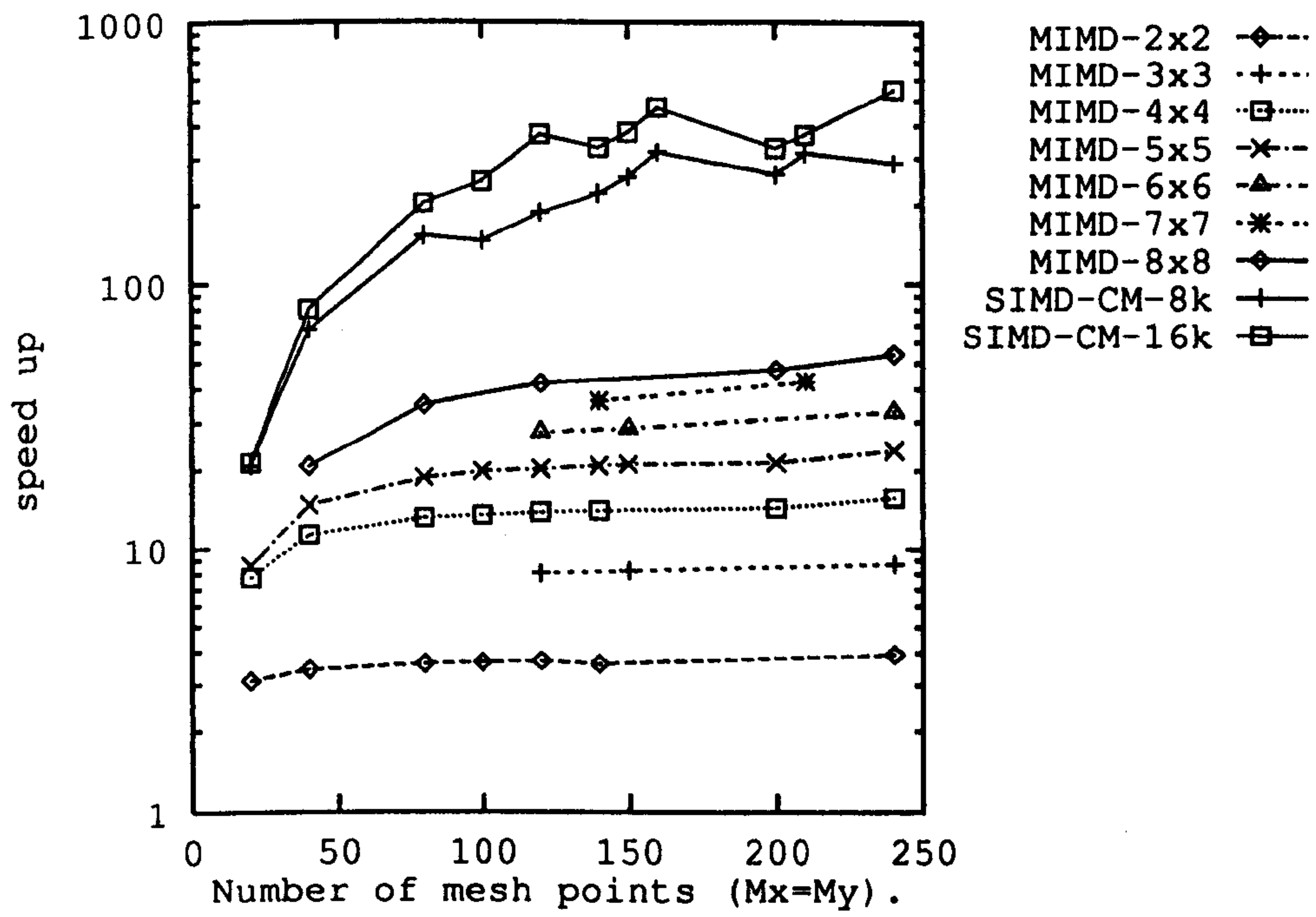


Fig. 6.11 Speed up of both of the parallel explicit methods using the MIMD and the SIMD computers, Top) the EFD-BPM; Bottom) the RS-BPM.

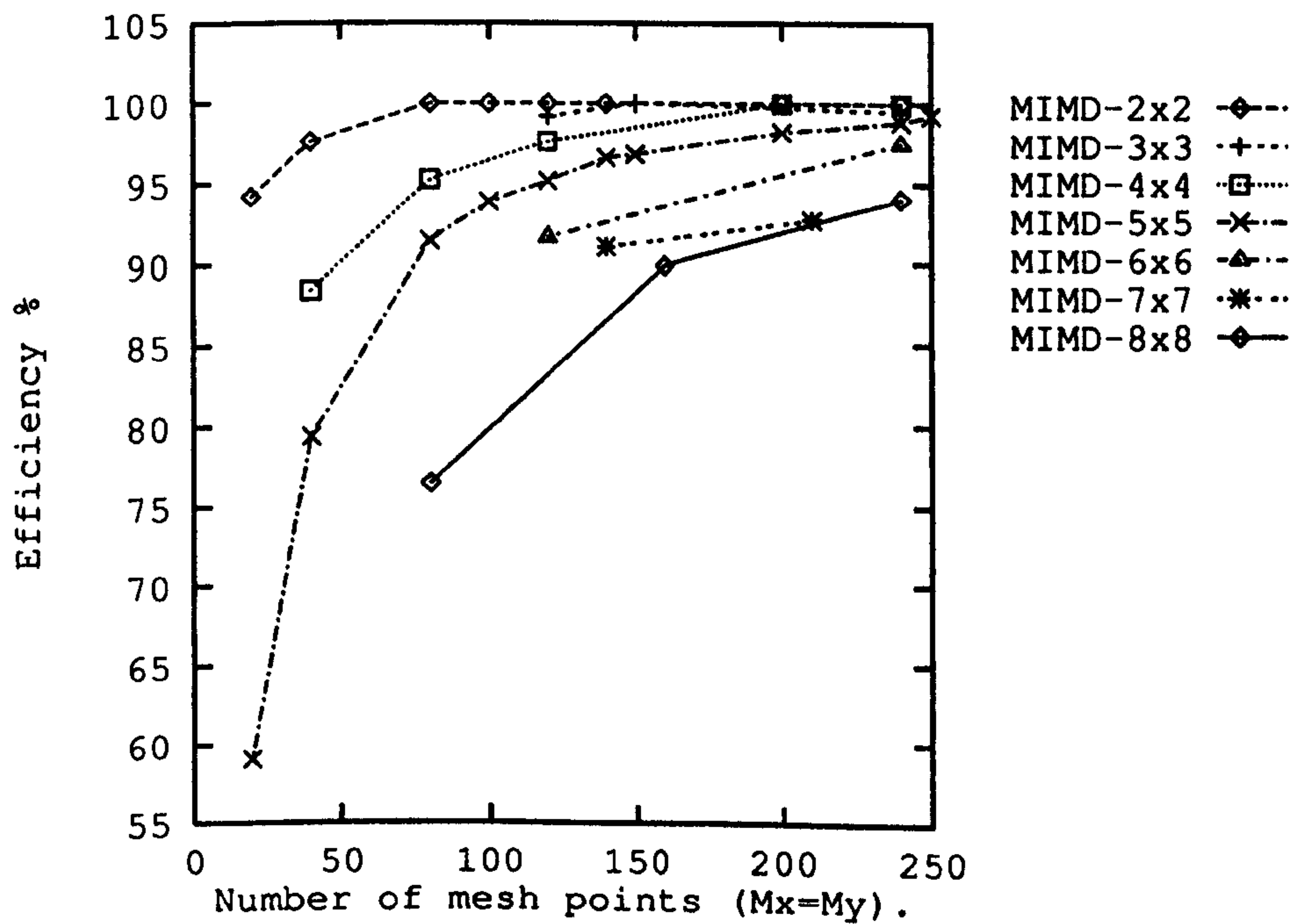
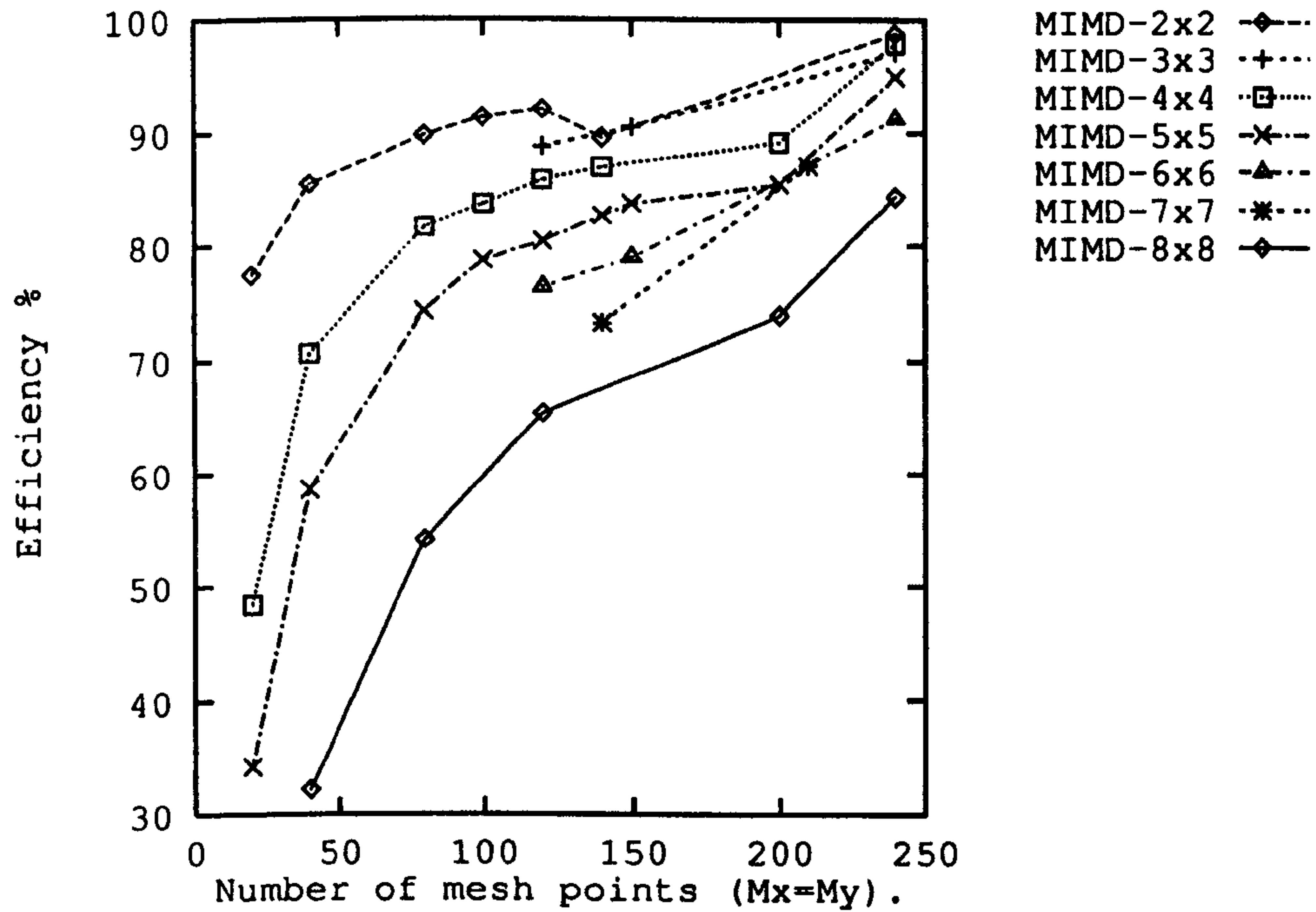


Fig. 6.12 The efficiency in percent of both of the parallel explicit methods of using the MIMD computer, Top) the EFD-BPM; Bottom) the RS-BPM.

Finally, the comparison between the speed of the two parallel methods shows that the parallel EFD-BPM is always faster than the parallel RS-BPM; for example, when using 16k of the CM at $M_x = 1000$ the speed of the EFD-BPM is around 6.8 times faster than the RS-BPM.

6.4 Analysis and Assessments

In this section, the results of the accuracy analysis of testing the two parallel explicit methods are shown [2]. Although the accuracy of the serial EFD-BPM has been verified by [12], in order to compare between the two parallel algorithms both of the parallel explicit methods were tested to analyse three well-known different z-invariant rib waveguides. Fig. 6.13 shows the rib waveguide geometry and Table 6.1 shows the parameters used in the analysis for the three structures. All the three structures were designed for single mode operation using an operating wavelength $\lambda = 1.55 \mu\text{m}$ [13]. The first structure with strong confinement in both lateral and vertical direction is very useful for curved waveguides where radiation loss is kept to its minimum. The mode of the second structure extends laterally because of the small size of the rib height; this waveguide is useful for directional coupler structures, as strong coupling between adjacent waveguides shortens the coupling length. The third structure was designed to give a relatively symmetric circular shape for efficient coupling to optical fibers [13].

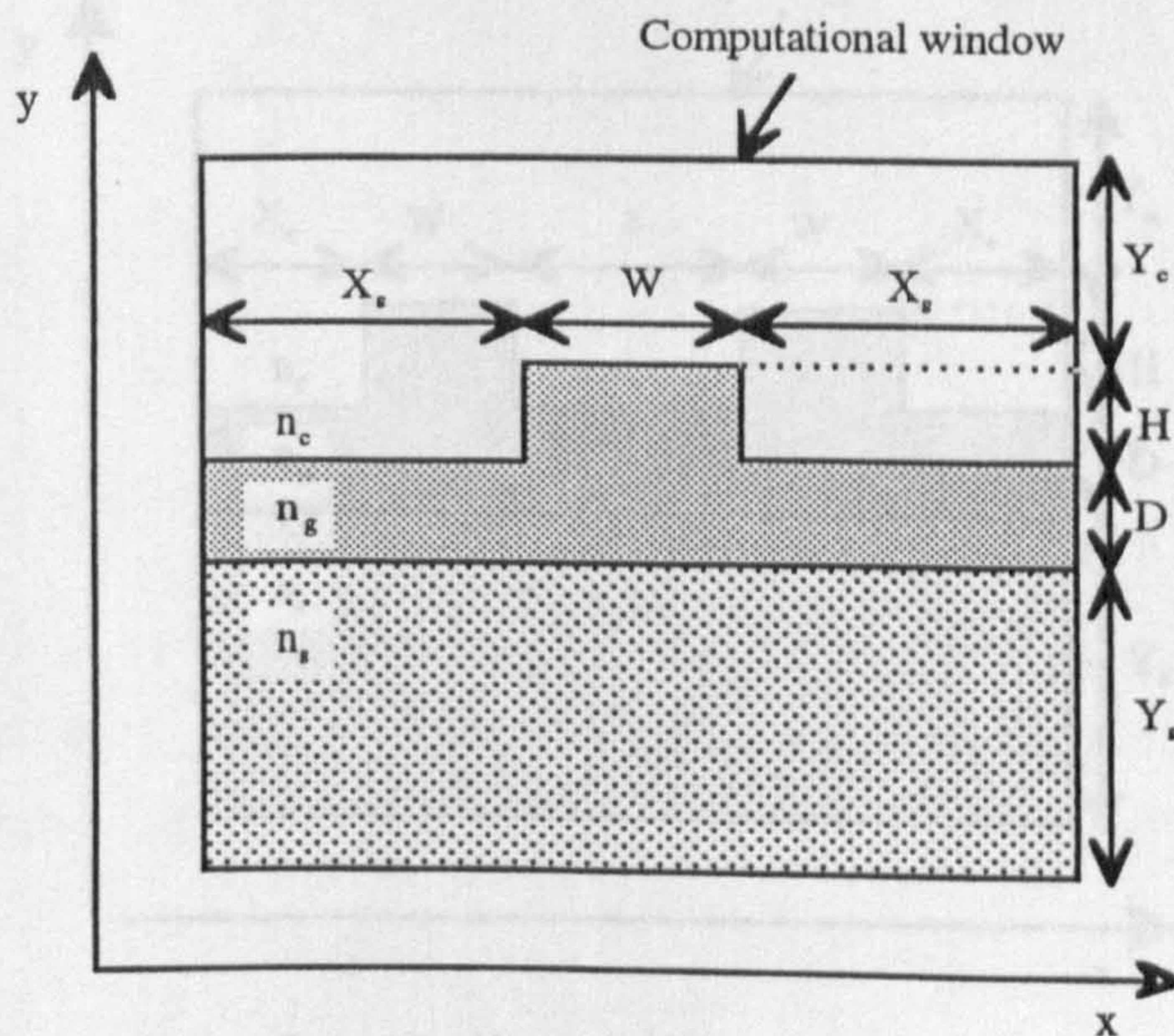


Fig 6.13 The rib waveguide used in the analysis of the three-dimensional FD-BPMs. The parameters are in Table 6.1.

From Table 6.1, the transverse computational window size of structure 1 is small because it is a strong guide, whereas structure 3 is a very weak guide; thus the window size must be large to minimise the influence of the boundary. In the following calculations, the internal interfaces of layers of different refractive indices have been set to be half-way between two adjacent mesh points and the reference index n_0 to be the substrate refractive index n_s [1,2,12,14].

Structure	n_g	n_s	n_c	X_s	W	Y_s	D	H	Y_c	X-window	Y-window
1	3.44	3.34	1.0	3.0	2.0	5.5	0.2	1.1	1.2	8.0	8.0
2	3.44	3.36	1.0	15.0	3.0	3.0	0.9	0.1	1.0	33.0	5.0
3	3.44	3.435	1.0	63.0	4.0	25.0	3.5	2.5	2.0	130.0	33.0

Table 6.1 The rib waveguide parameters used in the computations (see Fig. 6.13) with an operating wavelength of $\lambda = 1.55 \mu\text{m}$.

The first check involves the computation of the fundamental mode indices (n_{eff}) for the three structures. The second test is the computation of the coupling lengths (L_c) of directional couplers consisting of two-core rib waveguides using the same parameters of Fig. 6.13 with a separation gap of $s = 2 \mu\text{m}$ (see Fig. 6.14).

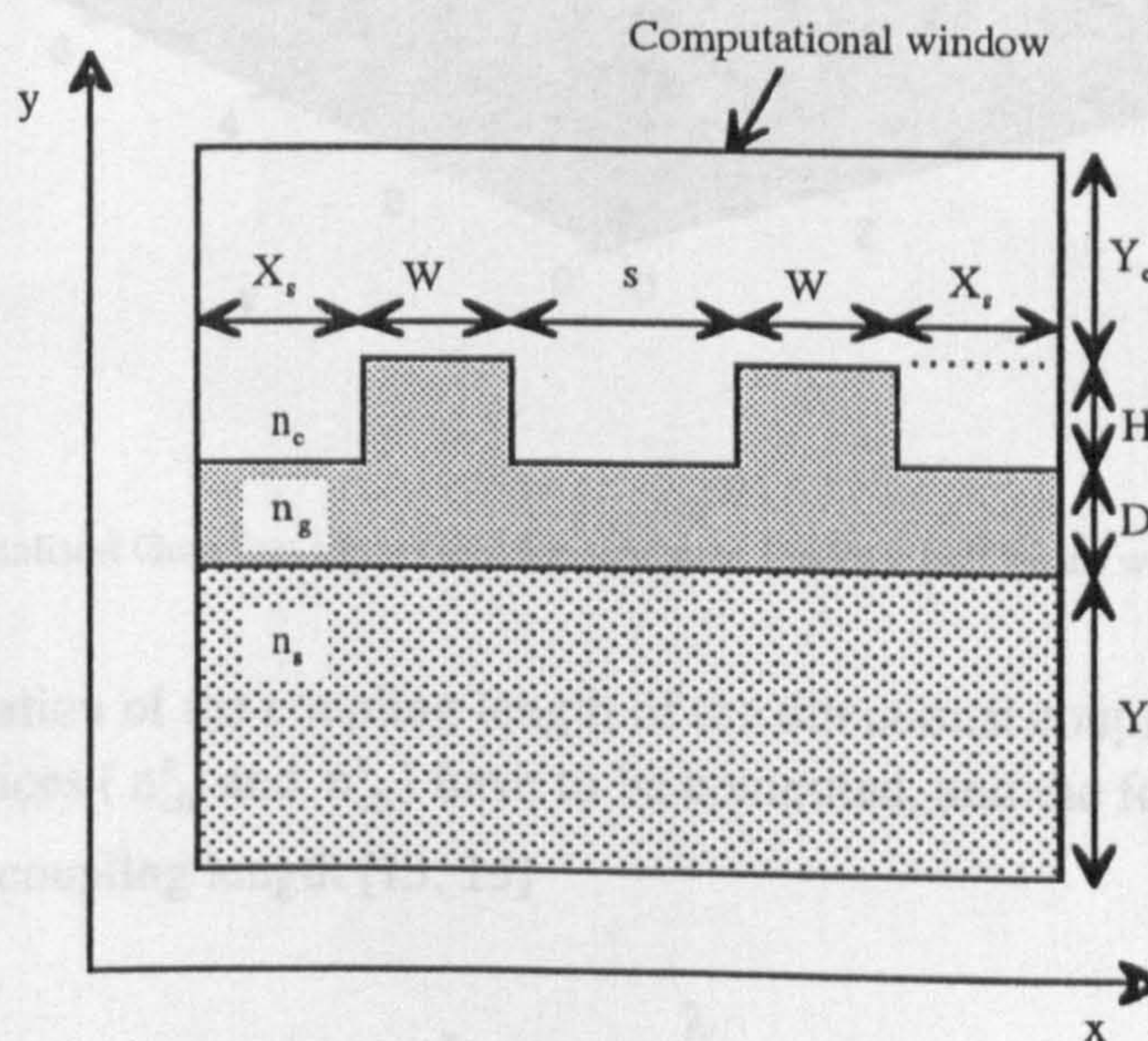


Fig 6.14 The directional coupler rib waveguide used in the analysis of the three-dimensional FD-BPMs. The parameters are the same of Fig. 6.13 with a separation distance of $s = 2 \mu\text{m}$.

The power spectral method⁵ has been used to compute the mode indices of the three structures from the BPM fields. In the following analysis, the input field is set to be a Gaussian field centred in the middle of the guiding layer to excite the fundamental mode of the structure. Fig. 6.15 shows the normalised Gaussian input field used to excite the fundamental mode of structure 1.

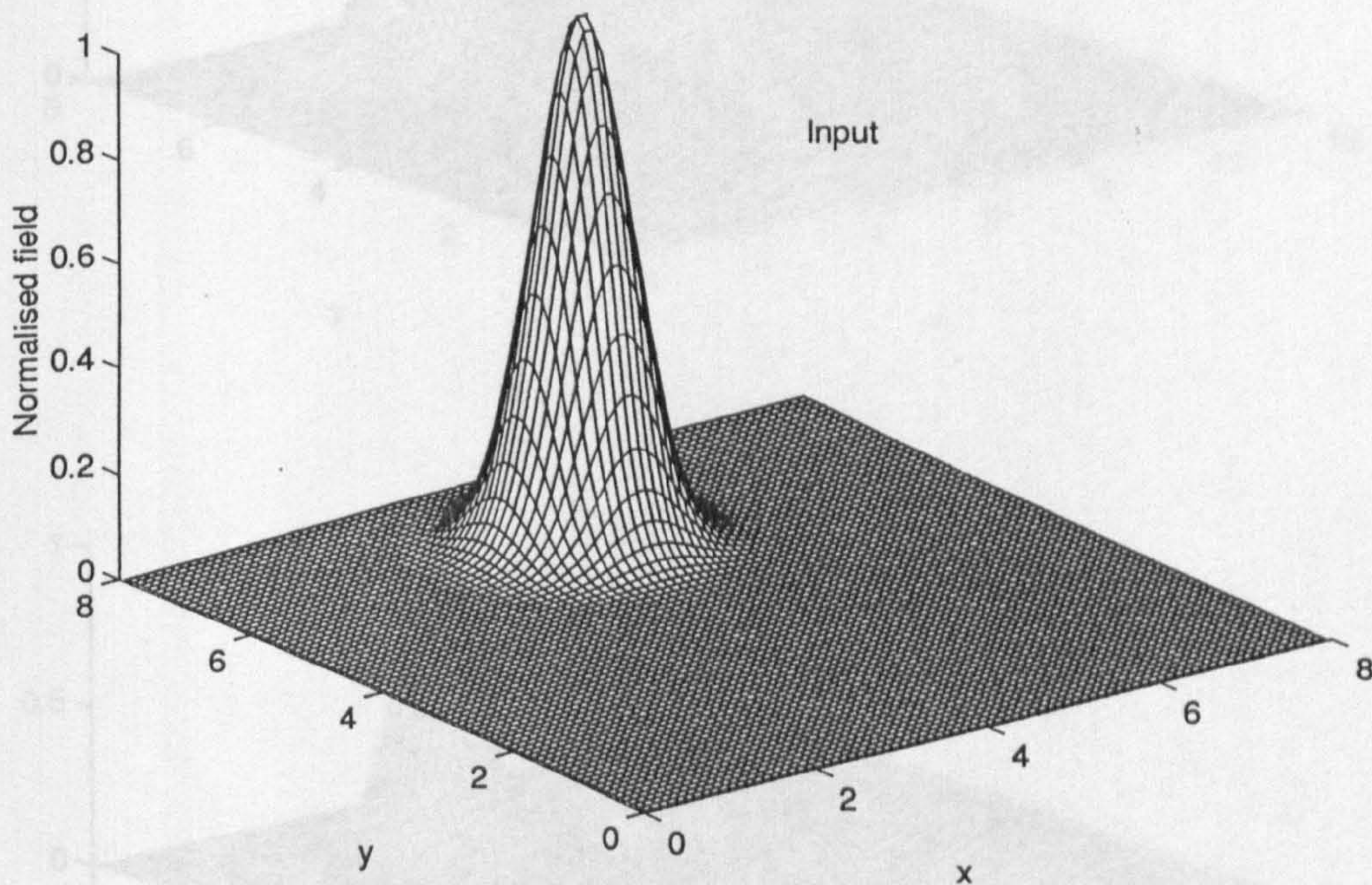


Fig. 6.15 The normalised Gaussian input field for structure 1 with a half beam width of 0.65 μm .

For the computation of the coupling length of the directional coupler, the even and the odd super mode indices (n_{eff}^e and n_{eff}^o) have to be computed, and the following relation is used to compute the coupling length [13, 15]

$$L_c = \frac{\lambda}{2(n_{\text{eff}}^e - n_{\text{eff}}^o)} \quad (6.9)$$

⁵ Discussed in chapters 3 (page 50) and 5 (page 91).

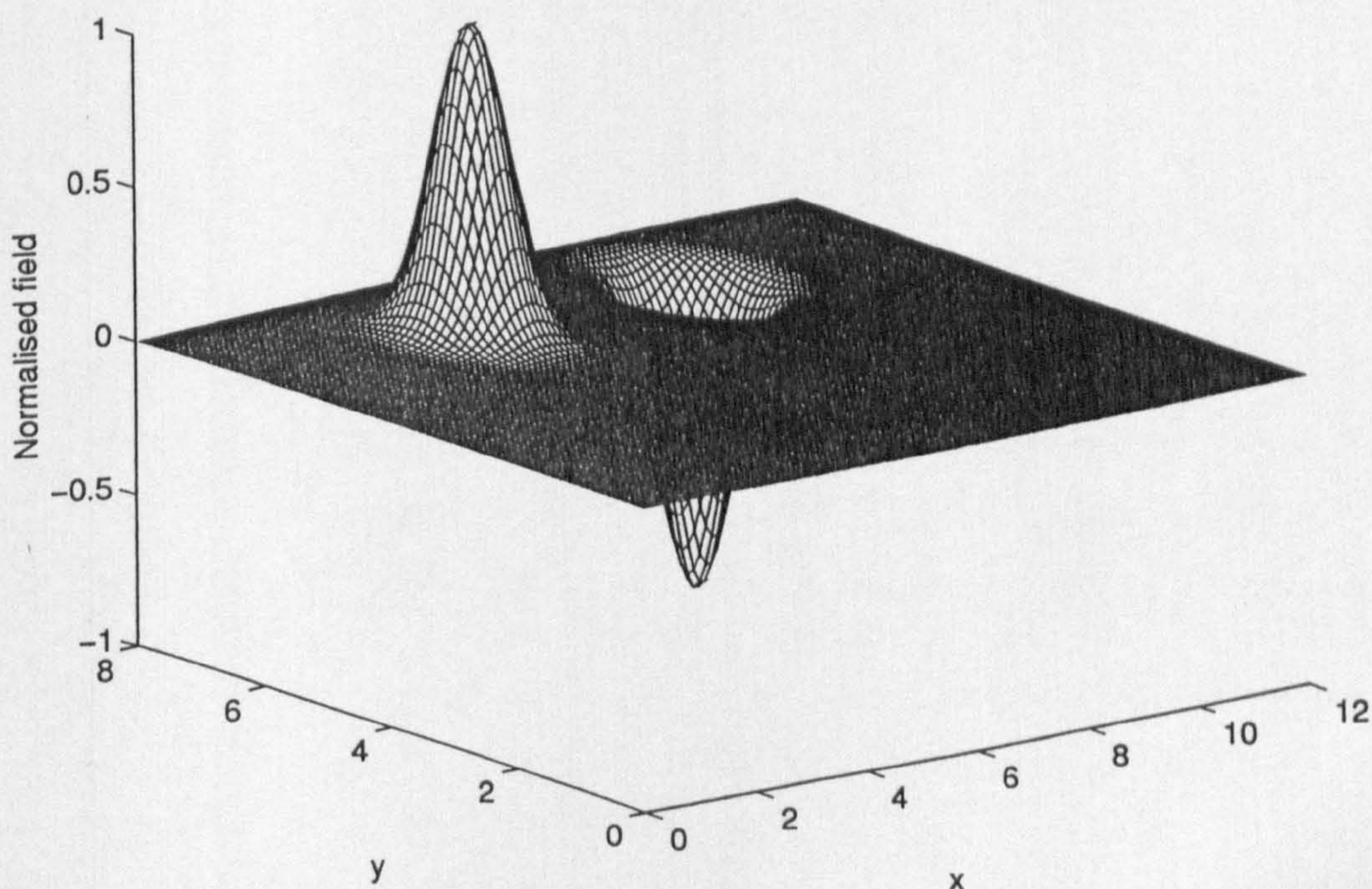
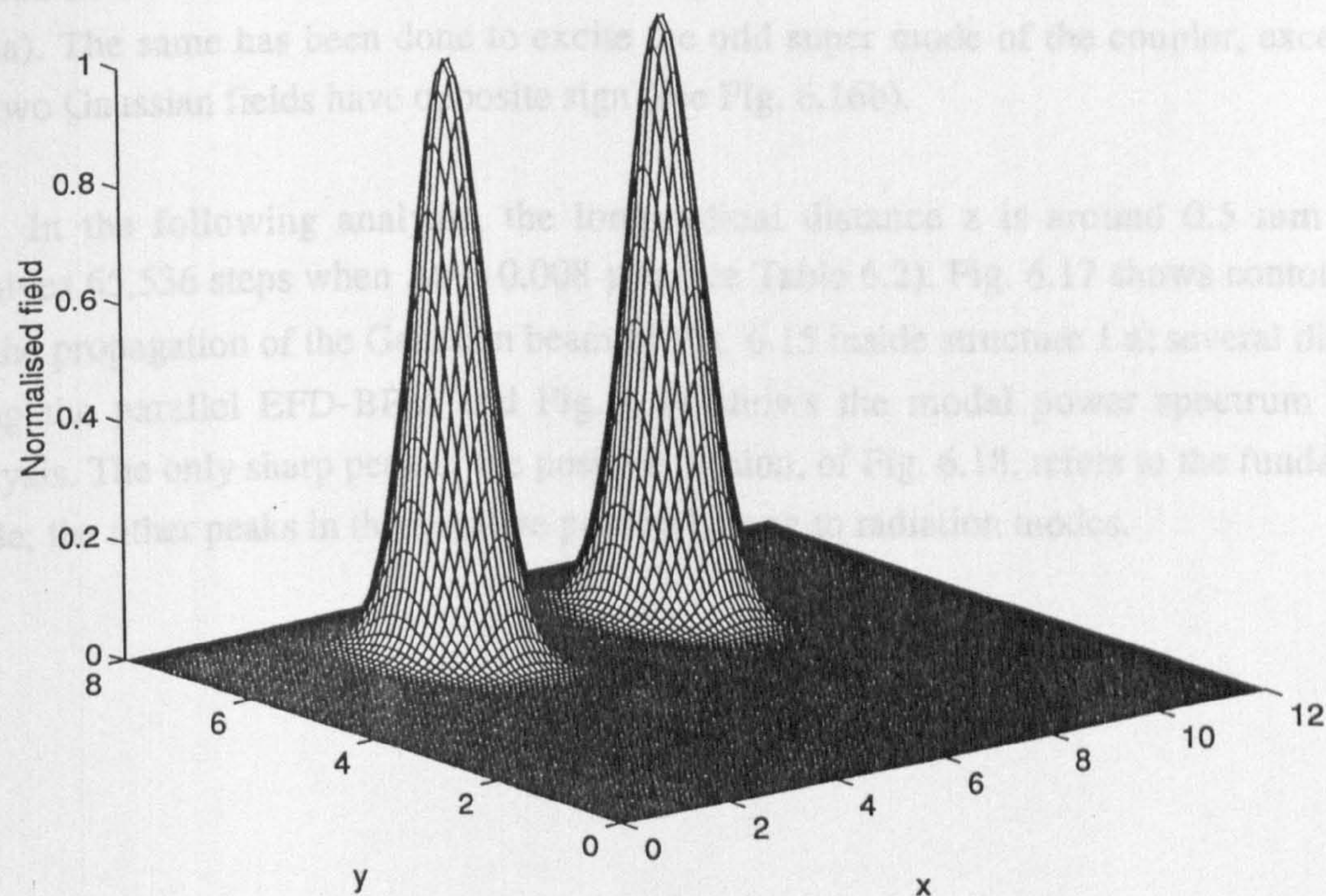


Fig. 6.16 The normalised Gaussian input field for the directional coupler of structure 1 with a half beam width of $0.65 \mu\text{m}$; Top) the even Gaussian input used to excite the even super mode and Bottom) the odd Gaussian input used to excite the super odd mode.

In order to excite the even super mode, a sum of two equal-amplitude Gaussian fields centred in the middle of each arm of the coupler has been launched as an input (see Fig. 6.16a). The same has been done to excite the odd super mode of the coupler, except that the two Gaussian fields have opposite sign (see Fig. 6.16b).

In the following analysis, the longitudinal distance z is around 0.5 mm which involves 65,536 steps when Δz is 0.008 μm (see Table 6.2). Fig. 6.17 shows contour plots for the propagation of the Gaussian beam of Fig. 6.15 inside structure 1 at several distances using the parallel EFD-BPM and Fig. 6.18 shows the modal power spectrum of this analysis. The only sharp peak in the positive portion, of Fig. 6.18, refers to the fundamental mode; the other peaks in the negative portion belong to radiation modes.

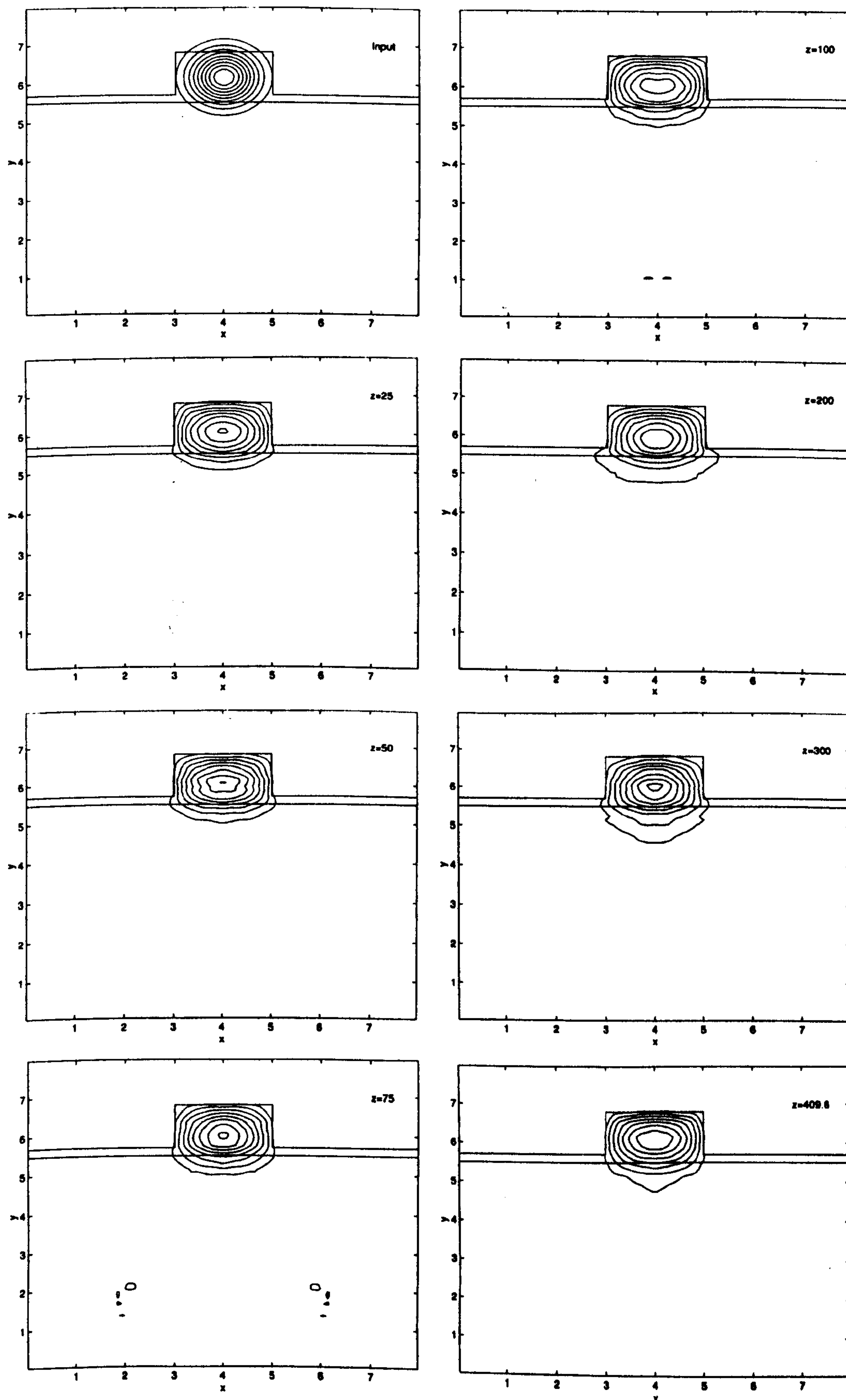


Fig. 6.17 Contour plots for the evolution of the normalised Gaussian beam of Fig. 6.15 inside structure 1 using the parallel EFD-BPM with $\Delta x = \Delta y = 0.1$ and $\Delta z = 0.025$ (all dimensions are in μm).

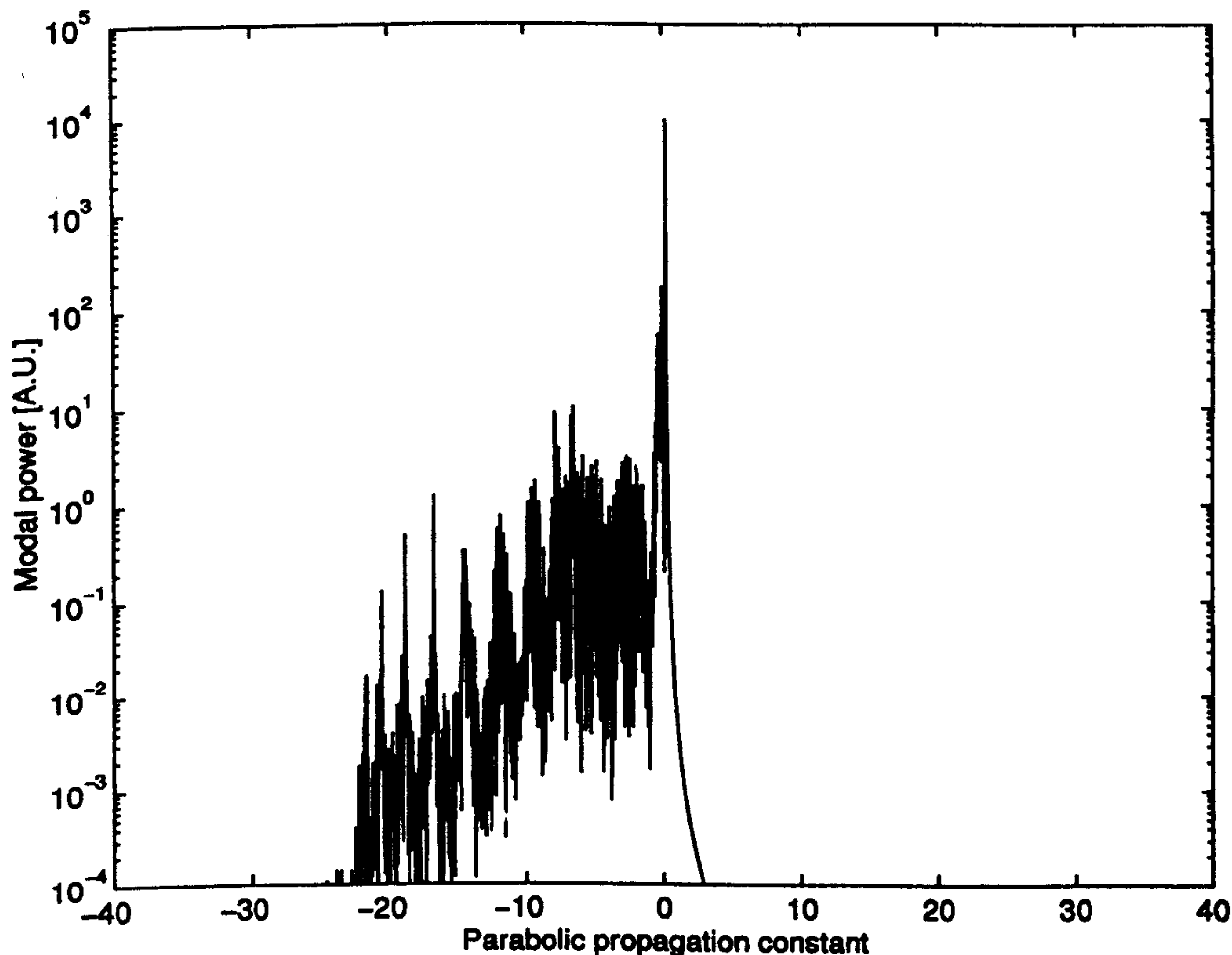


Fig. 6.18 The modal power spectrum for the propagation of the Gaussian beam of Fig. 6.17.

Tables 6.2, 6.3 and 6.4 show the fundamental mode indices n_{eff} and the coupling lengths L_c of structures 1, 2 and 3 respectively, computed by both of the parallel EFD-BPM and RS-BPM [2]. Table 6.2, shows values of n_{eff} and L_c for various transverse mesh sizes Δx and Δy of both algorithms, as well as for various step sizes Δz for the RS-BPM. The EFD-BPM is not stable for values of Δz larger than those given in Table 6.2. As for the 2-D analysis of chapter 5, the results converge as the grid spacing decreases; the convergence of the two methods is clearly demonstrated in Table 6.2. In order for the RS-BPM to converge, Δz should be decreased as the transverse mesh sizes are decreased. It can be observed from Table 6.2 that both methods produce similar results around almost the same step size, thus we can conclude that the parallel EFD-BPM is more efficient than the parallel RS-BPM since it is faster. The results of Tables 6.3 and 6.4 indicate similar conclusions to those of Table 6.2; however, it has been noticed that the values of the mode

index and the coupling length produced by the EFD-BPM for structure 2, are larger than the values of the RS-BPM. To test this further, Δx is decreased to 0.05 for the EFD-BPM and the new values of n_{eff} and L_c calculated, which are 3.395641865 and 0.8299 mm respectively. These are in good agreement with the values of the RS-BPM.

To validate the results of the parallel explicit methods, Table 6.5 contains published results of references [13] and [14] of the same structures for comparisons. FD(1) and FD(2) of [13] are the solution of the scalar wave equation in (2.47) using the finite-difference approximation with different treatment to the interfaces between layers and WAVE is also a solution to the scalar wave equation using a function fitting technique. On the other hand, the results of [14] are based on the solution of the scalar parabolic equation using the FFT-BPM. It can be observed that the results of the EFD-BPM and the RS-BPM are as accurate as the results of Table 6.5 and very close to those in reference [13]. It has been pointed out by [14], and can be seen from Table 6.5, that the coupling lengths of the FFT-BPM are always shorter than any other method. In addition to the discussion given in [14], Δz should be decreased further, as for the RS-BPM, in order to arrive at better results.

Figs. 6.19a, 6.19b and 6.19c show, respectively, the fundamental modes of structures 1, 2 and 3 using the parallel EFD-BPM. Fig. 6.20 shows a three-dimensional plot for the fundamental mode of the directional coupler of structure 1, while the second-order mode profile is the same except that one of the peaks is reversed (not shown). In order to demonstrate the mechanism of the directional coupler, the parallel EFD-BPM was used to propagate the fundamental mode profile of structure 1 where this input field is launched in the left waveguide of structure 1 of the directional coupler in Fig. 6.14 and using a separation distance $s = 0.4 \mu\text{m}$. Fig. 6.21 shows three-dimensional plots for the power exchange between the two waveguides using the parallel EFD-BPM at several propagational distances. The propagation in Fig. 6.21 was allowed to cover two coupling lengths, and the input field is totally recovered at a distance of around $z = 3.808 \text{ mm}$. Finally, the parallel EFD-BPM is used to compute the coupling length in Eq. 6.9 as a function of the separation distance s between the two waveguides of structure 1. Fig. 6.22 shows the computed coupling length versus the separation distance s .

Δx (μm)	Δy (μm)	Δz (μm)	n_{eff}		L_c (mm)	
			EFD-BPM	RS-BPM	EFD-BPM	RS-BPM
0.2	0.1	0.25	-	3.390792451	-	285.45
		0.125	-	3.392857134	-	300.27
		0.0625	-	3.393297053	-	303.45
		0.035	3.393440335	-	304.52	-
		0.03125	-	3.393403764	-	304.28
		0.015625	-	3.393430250	-	304.52
0.1	0.1	0.25	-	3.388004414	-	$n_e < n_o$
		0.125	-	3.391420992	-	293.56
		0.0625	-	3.392132577	-	315.68
		0.03125	-	3.392304294	-	321.44
		0.025	3.392362259	-	323.46	-
		0.015625	-	3.392346865	-	322.78
		0.0078125	-	3.392357486	-	323.19
0.1	0.05	0.125	-	3.382249421	-	-
		0.0625	-	3.389903808	-	299.00
		0.03125	-	3.391376826	-	326.32
		0.015625	-	3.391718884	-	333.05
		0.0124	3.391831397	-	335.50	-
		0.0078125	-	3.391802926	-	334.63
0.05	0.05	0.0625	-	3.388289549	-	256.88
		0.03125	-	3.390725915	-	305.00
		0.015625	-	3.391287509	-	331.34
		0.008	3.391472234	-	340.66	-
		0.0078125	-	3.391425235	-	338.28
		0.00390625	-	3.391459065	-	340.21
0.025	0.025	0.03125	-	3.375348807	-	24.08
		0.015625	-	3.388199074	-	214.33
		0.0078125	-	3.390513402	-	308.15
		0.00390625	-	3.391051924	-	335.208
		0.002	3.391228915	-	358.80	-

Table 6.2 The fundamental mode indices and the coupling lengths of structure 1, computed using the parallel EFD-BPM and the parallel RS-BPM.

Δx (μm)	Δy (μm)	Δz (μm)	n_{eff}		L_c (mm)	
			EFD-BPM	RS-BPM	EFD-BPM	RS-BPM
0.1	0.05	0.125	-	3.385005863	-	0.84325
		0.0625	-	3.393453043	-	0.82987
		0.03125	-	3.395147688	-	0.83523
		0.015625	-	3.395536853	-	0.83702
		0.0124	3.395678682	-	0.91000	-
		0.0078125	-	3.395632373	-	0.83731

Table 6.3 The fundamental mode indices and the coupling lengths of structure 2, computed using the parallel EFD-BPM and the parallel RS-BPM.

Δx (μm)	Δy (μm)	Δz (μm)	n_{eff}		L_c (mm)	
			EFD-BPM	RS-BPM	EFD-BPM	RS-BPM
0.1	0.05	0.125	-	3.4364715439	-	1.40587
		0.0625	-	3.436979326	-	1.81475
		0.03125	-	3.437084489	-	1.24674
		0.012	3.437155943	-	1.25277	-

Table 6.4 The fundamental mode indices and the coupling lengths of structure 3, computed using the parallel EFD-BPM and the parallel RS-BPM.

Stru- cture	n_{eff}				L_c (mm)			
	FD(1) ¹³	FD(2) ¹³	WAVE ¹³	PBM-FFT ¹⁴	FD(1) ¹³	FD(2) ¹³	WAVE ¹³	PBM-FFT ¹⁴
1	3.390617745	3.391291712	3.390449	3.3913	357	341	$n_e < n_o$	65.1
2	3.395166250	3.395429873	3.394888	3.3960	0.797	0.811	0.827	0.71
3	3.436842513	3.436863500	3.436724	3.4365	1.273	1.347	1.968	0.93

Table 6.5 The fundamental mode indices and the coupling lengths in references 13 and 14 used for comparisons.

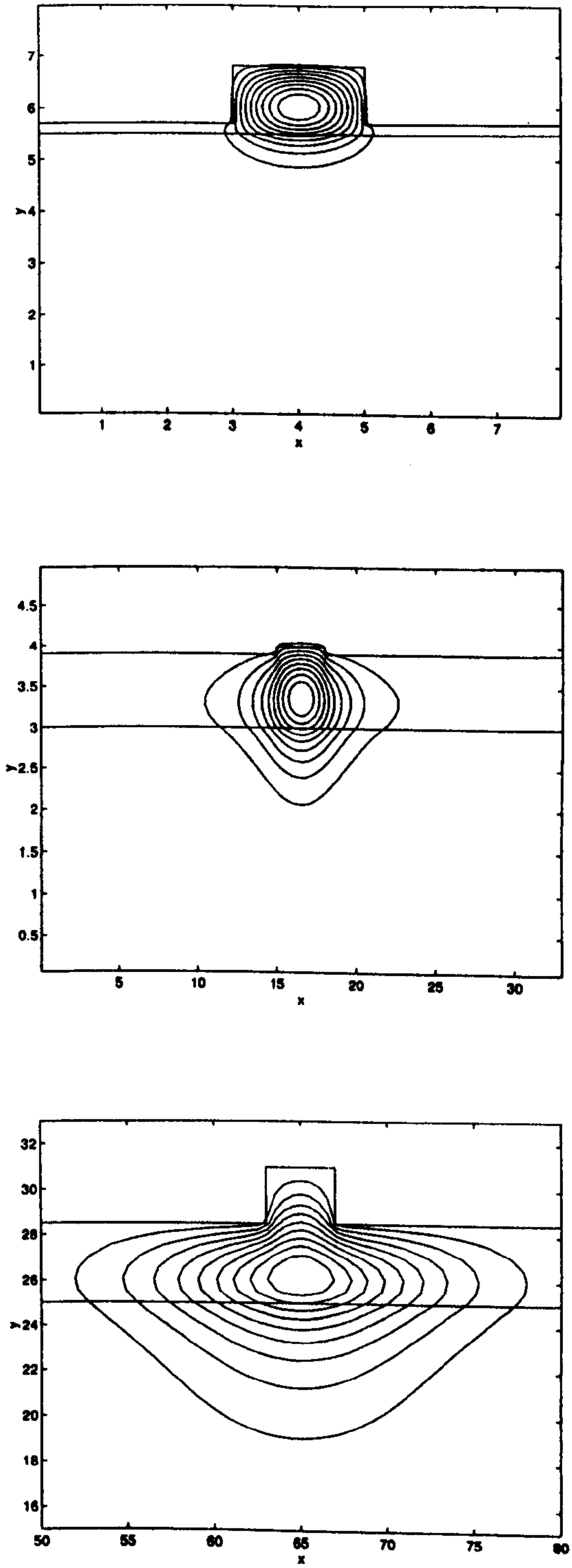


Fig. 6.19 Contour plots for the fundamental modes of (Top) structure 1, (Middle) structure 2 and (Bottom) structure 3.

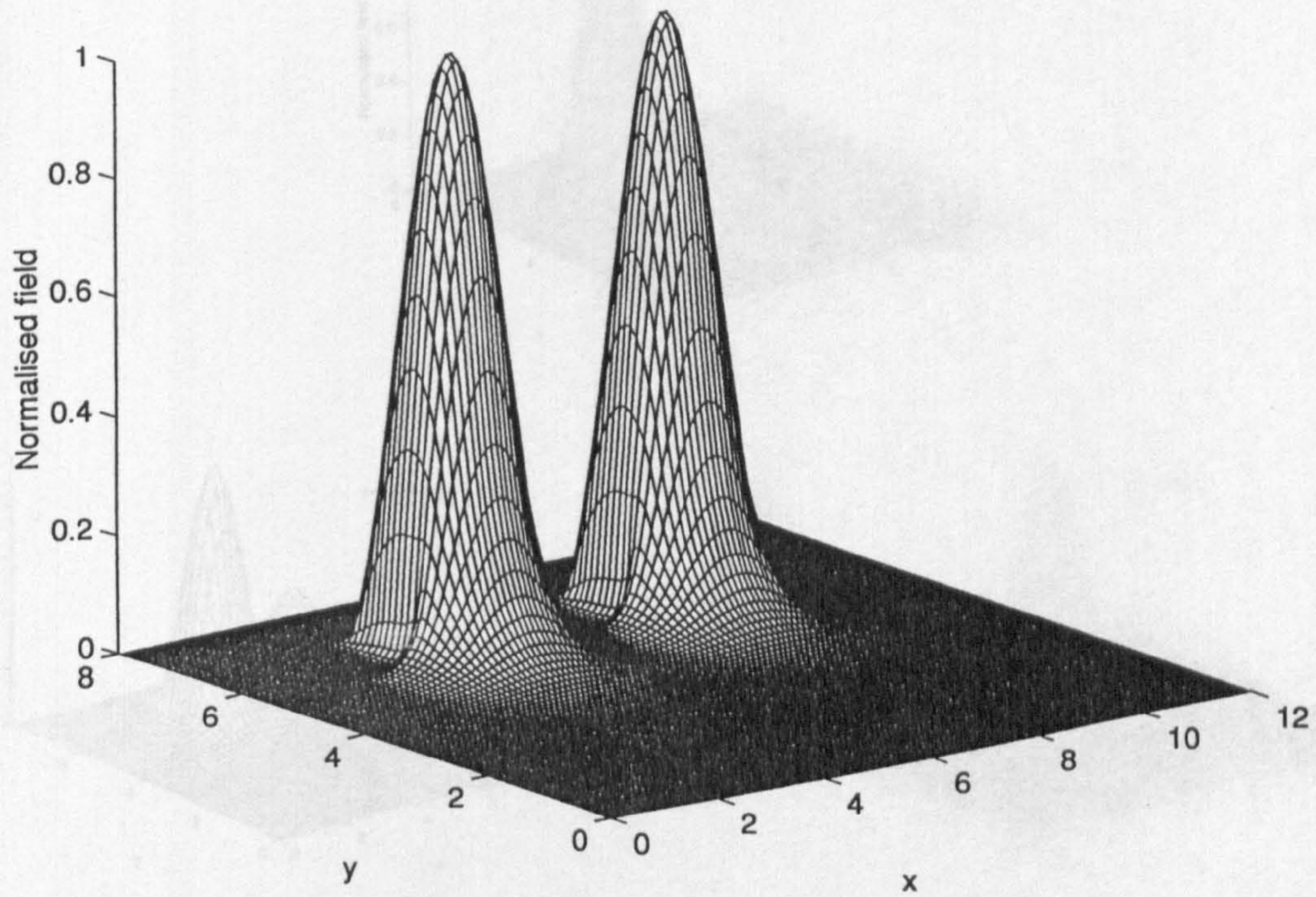


Fig. 6.20 The fundamental super mode profile of the rib waveguide directional coupler of structure 1.

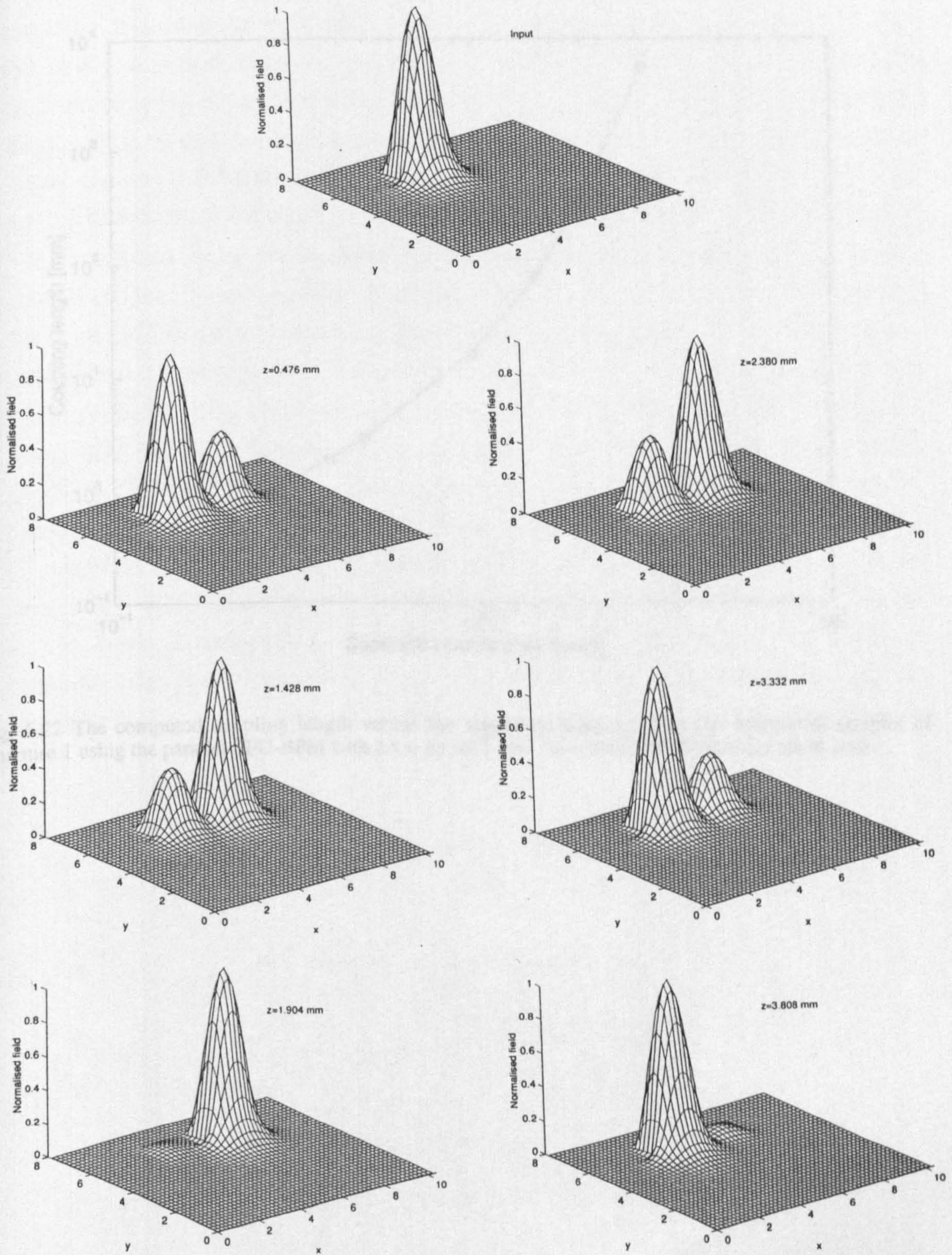


Fig. 6.21 The power exchange between the two waveguides of the directional coupler of structure 1 using the EFD-BPM with $\Delta x = \Delta y = 0.1$ and $\Delta z = 0.025$ (all dimensions are in μm). The input field is the fundamental mode profile of structure 1.

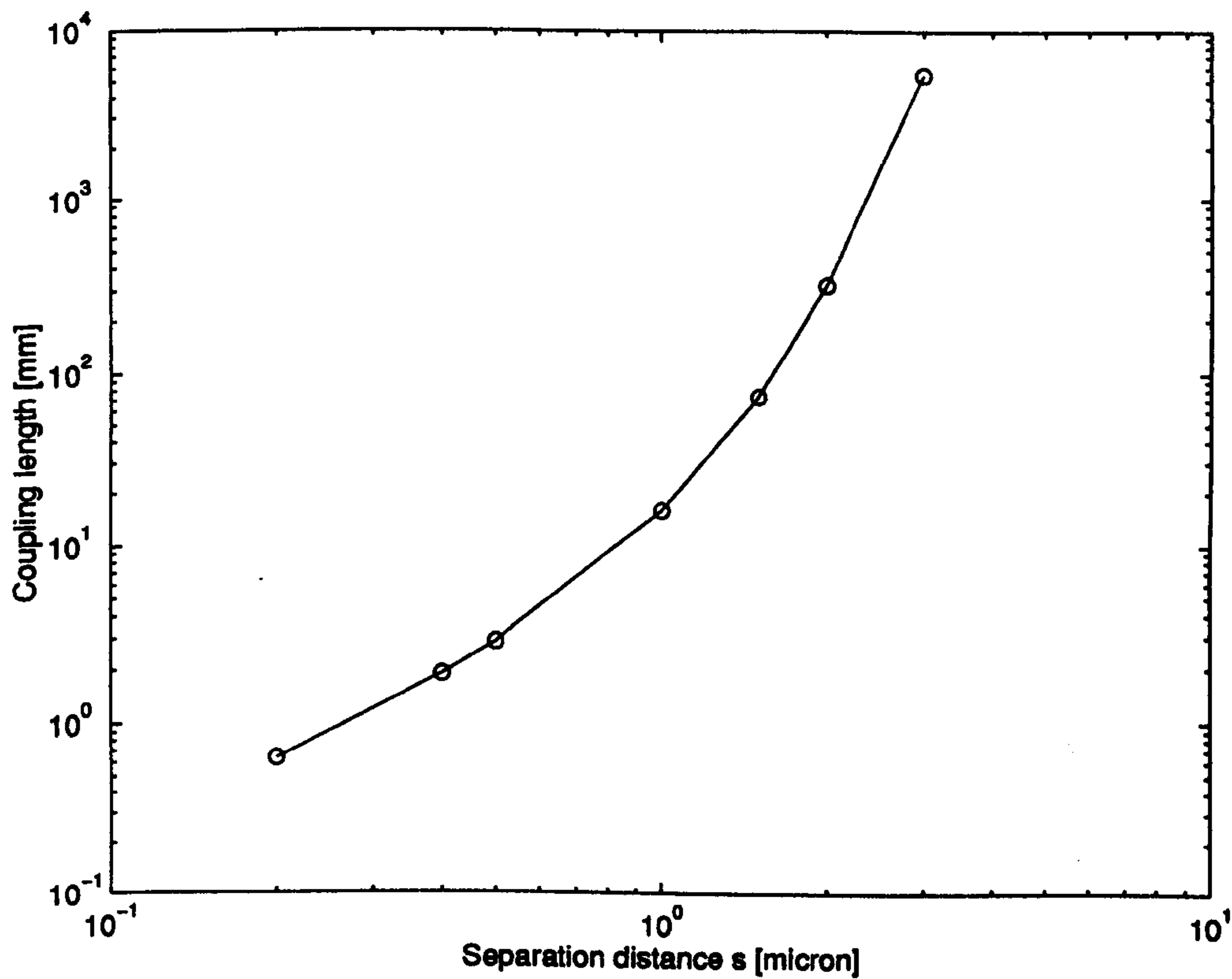


Fig. 6.22 The computed coupling length versus the separation distance s for the directional coupler of structure 1 using the parallel EFD-BPM with $\Delta x = \Delta y = 0.1$ and $\Delta z = 0.025$ (all dimensions are in μm).

Examples

In the following demonstrations, the parallel EFD-BPM is used to show the propagation of the fundamental mode profile of structure 1 inside the three-core coupler shown in Fig. 6.23 with the same parameters of structure 1 in Table 6.1. Three different cases of propagating the same initial field inside the three-core coupler were considered; the first and the second cases are a symmetric structure when the two separating distances are equal ($s_1 = s_2 = 0.4 \mu\text{m}$) and the third case is an asymmetric structure when the two separating distances are not equal ($s_1 = 0.4 \mu\text{m}$ and $s_2 = 0.8 \mu\text{m}$). In the first case, the input field is launched in the left waveguide while observing, during propagation, the coupling to the middle and the right waveguides. The second and the third cases consider the launching of the input field in the middle waveguide. Symmetric propagation in the second case is expected since the waveguide is symmetric. On the other hand, the structure of the third case is asymmetric; thus, an asymmetric propagation effect must be observed. Figs. 6.24, 6.25 and 6.26 show the propagation of the initial field for the three cases 1, 2 and 3 respectively. From Fig. 6.24, the power exchange mechanism from the left to the middle and then to the right waveguides is clearly demonstrated; in addition the opposite process, where the field is coupled from the right waveguide back to the left waveguide, is also shown. In Fig. 6.25, the expected symmetric coupling from the middle waveguide to the outer waveguides is taking place with the reverse process, after a complete coupling to the outer waveguides, is also observable where the initial field recovers its original shape. On the other hand, Fig. 6.26 shows the asymmetric coupling effect compared to the one in Fig. 6.25.

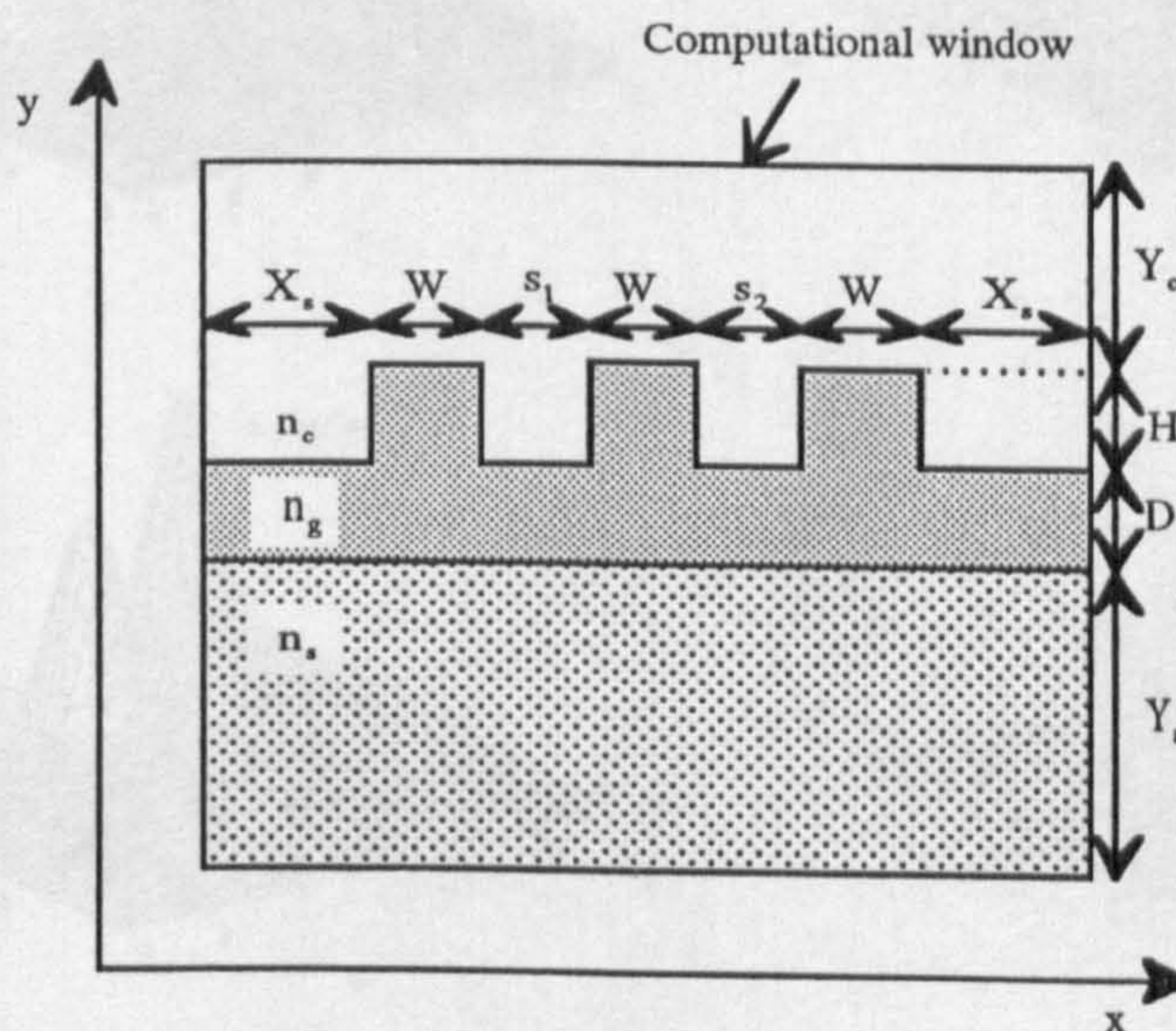


Fig 6.23 The three core coupler rib waveguide. The parameters are the same of Fig. 6.13 and the values of s_1 and s_2 are given in the text.

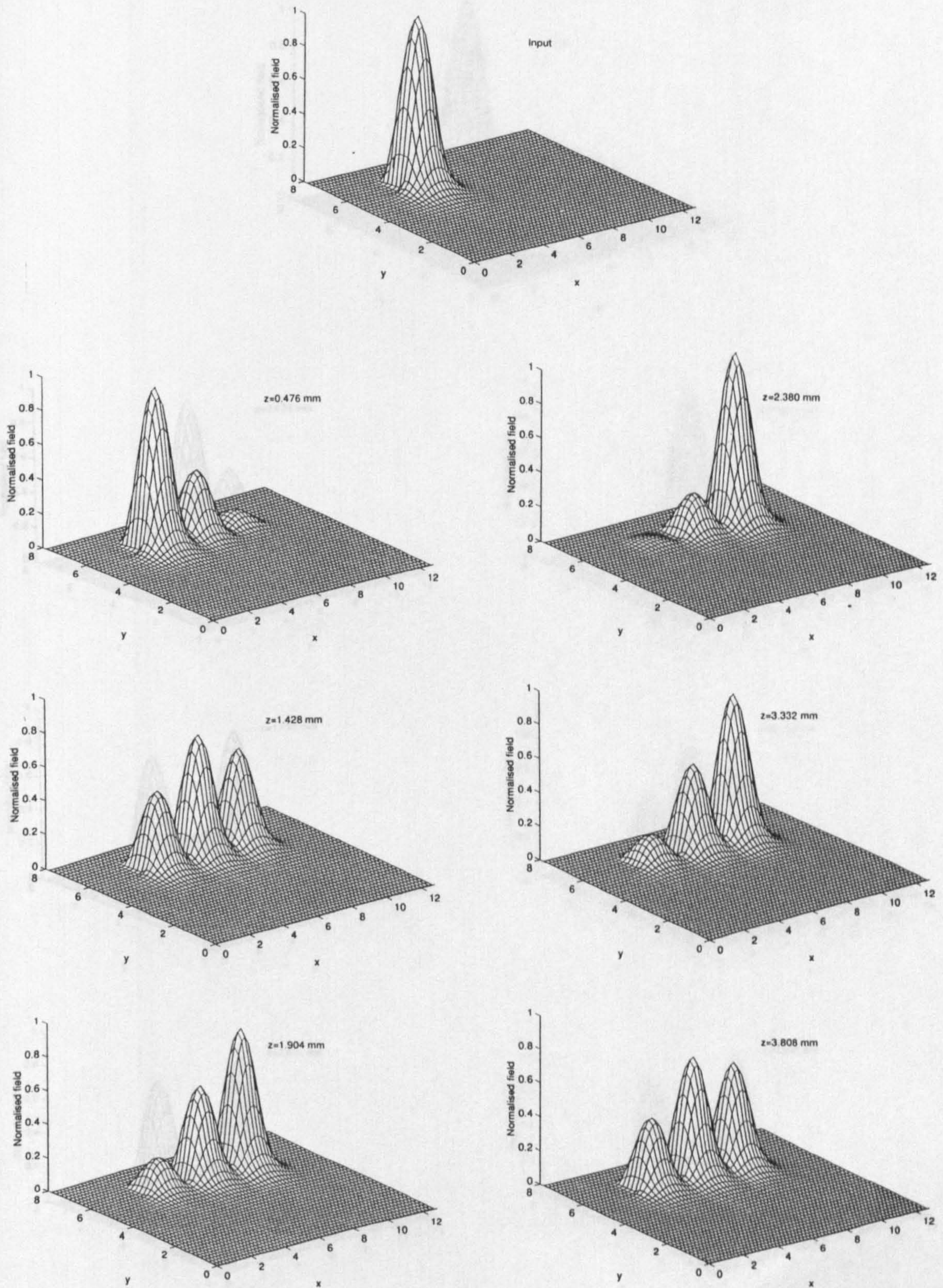


Fig. 6.24 Case 1, the propagation of the fundamental mode profile of structure 1 inside the symmetric three-core coupler of Fig. 6.23 using the parallel EFD-BPM with $\Delta x = \Delta y = 0.1$, $s_1 = s_2 = 0.4$ and $\Delta z = 0.025$ (all dimensions are in μm). The input is launched in the left waveguide.

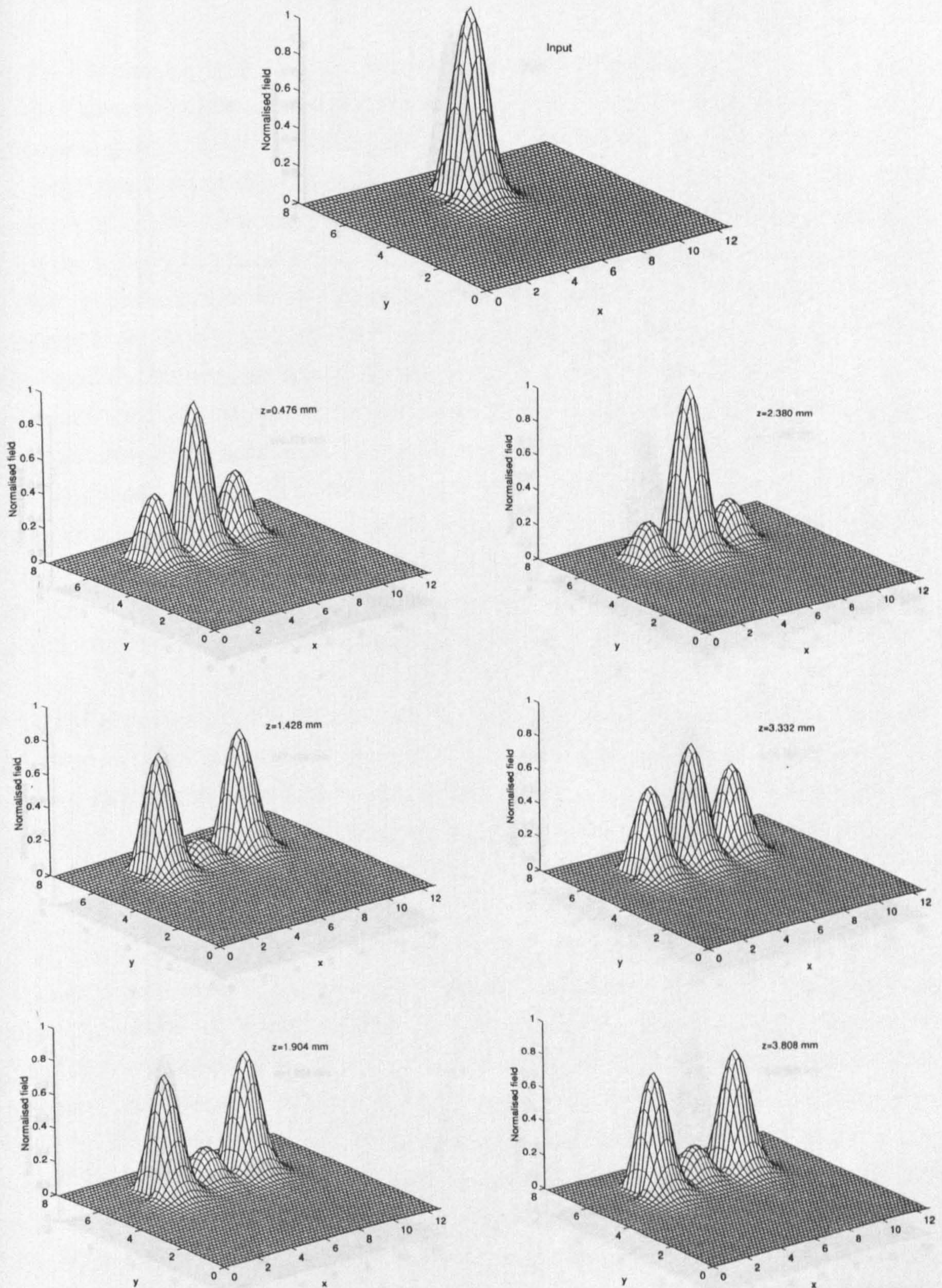


Fig. 6.25 Case 2, the same as in Fig. 6.24 except that the input is launched in the middle waveguide forming a symmetric propagation profile.

6.5 Discussion and Conclusion

In this chapter, the implementation of the parallel beam propagation method and the Connection Machine is described. The results of the parallel implementation are compared with the serial implementation. The parallel implementation is shown to be a large speed-up of the serial implementation. The parallel implementation is shown to be a large speed-up of the serial implementation. The parallel implementation is shown to be a large speed-up of the serial implementation.

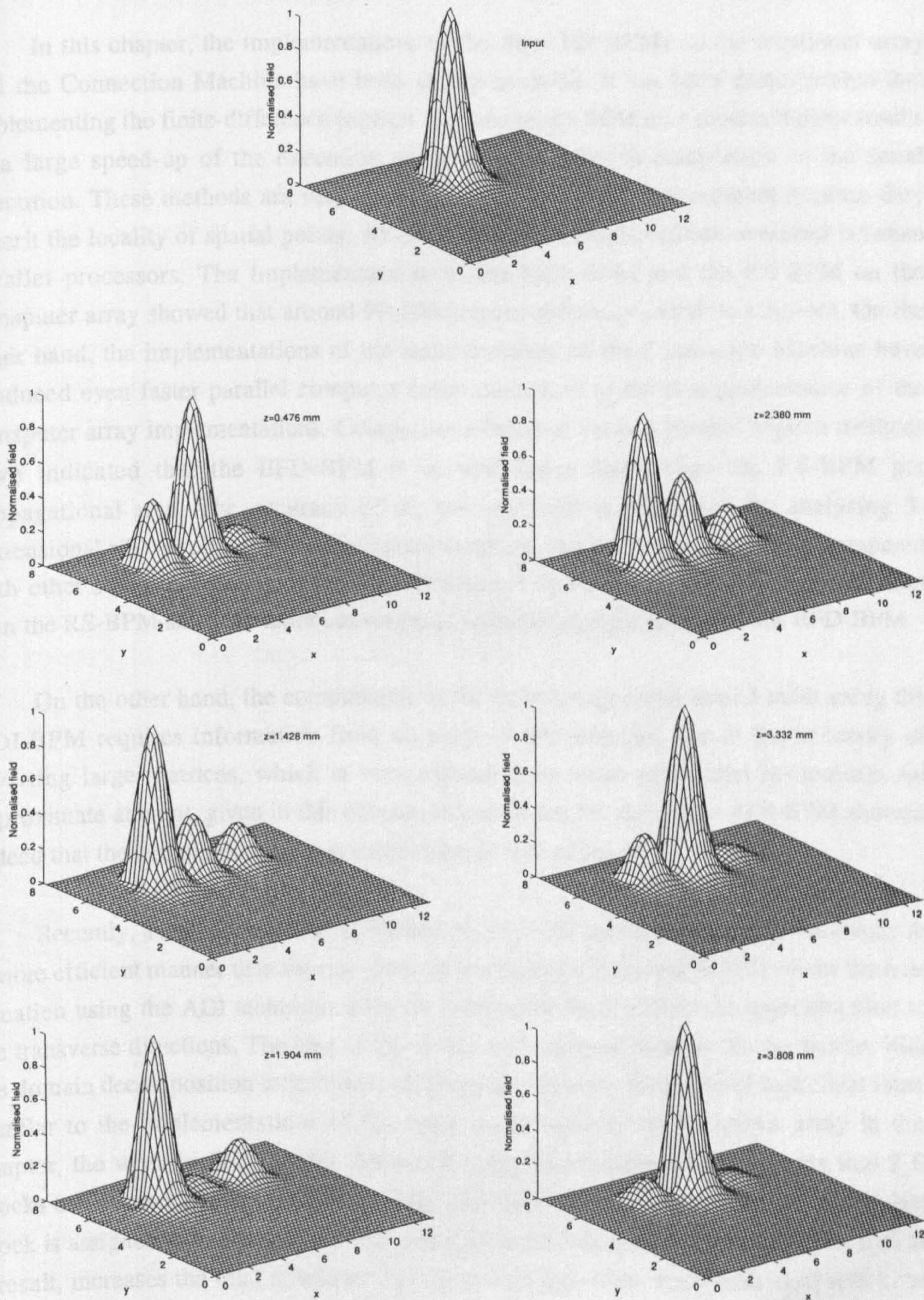


Fig. 6.26 Case 3, the same as in Fig. 6.25 except that $s_2 = 0.8$ forming an asymmetric structure.

6.5 Discussion and Conclusion

In this chapter, the implementations of the three FD-BPMs on the transputer array and the Connection Machine have been shown in detail. It has been demonstrated that implementing the finite-difference explicit versions of the BPM on a supercomputer results in a large speed-up of the execution of these algorithms in comparison to the serial execution. These methods are very well suited to the parallel environment because they inherit the locality of spatial points, which reduces the communication overhead between parallel processors. The implementations of the EFD-BPM and the RS-BPM on the transputer array showed that around 90-100 percent efficiency could be achieved. On the other hand, the implementations of the same methods on the Connection Machine have produced even faster parallel computer codes compared to the best performance of the transputer array implementations. Comparisons between the two parallel explicit methods have indicated that the EFD-BPM is several times faster than the RS-BPM per propagational step. The accuracy of the two methods is confirmed by analysing 3-dimensional rib waveguides and directional couplers, and the results have been compared with other serial techniques. It has been concluded that the EFD-BPM is more efficient than the RS-BPM since the latter converges at a similar step size to that of the EFD-BPM.

On the other hand, the computation of the field at any given spatial point using the ADI-BPM requires information from all parts of the problem, due to the necessity of inverting large matrices, which is very expensive in terms of parallel computing. An approximate attempt, given in this chapter, to parallelise the first order ADI-BPM showed indeed that the communication cost reduces the overall efficiency.

Recently, a method has been published in [16]⁶ that parallelises the ADI technique in a more efficient manner than the one given in this chapter. The work in [16] solves the heat equation using the ADI technique with the fourth-order finite-difference approximation to the transverse directions. The idea of the 'divide-and-conquer' strategy in conjunction with the domain decomposition technique is used to parallelise the inversion of numerical lines. Similar to the implementations of the explicit methods on the transputer array in this chapter, the work in [16] divides the two dimensional transverse mesh points into 2-D blocks of data and also uses a 2-D topology of processors, where each computational data block is assigned to one processor. This arrangement avoids global communication and, as a result, increases the total efficiency of the parallel algorithm. Each numerical line to be

⁶ This work was found during the time of writing this thesis.

inverted is divided, using the 'divide-and-conquer' technique, into several segments in which each segment can be inverted independently. In other words, each numerical line is inverted using one line of the 2-D topology of processors, where each segment is mapped into its corresponding processor. However, using the 'divide-and-conquer' technique to invert each numerical line in parallel adds additional computations to the normal inversion of a system of equations for evaluating, recursively, new parameters. As for the parallel RS-BPM, the four local border boundary must be transferred to neighbouring processors after computing vertical and horizontal numerical lines. In addition, diagonal data passes along the diagonals of the 2-D topology are also needed, to exchange the information at the corners of each local data block. Also another communication mechanism is needed along each line of processors for the serial computation of the new parameters. The above parallel ADI was tested in [16] using CM-5 in the MIMD mode by varying the number of processors and the number of total mesh points to assess the performance of the method. From these tests, the speed of the algorithm increases to an optimum level and decreases afterward where the decrease is due to the increase in the communication cost compared to the computational cost.

As mentioned in the introduction, the propagational step size of the ADI-BPM is considerably larger than that of the explicit methods; however, comparison between the parallel implementations of the explicit methods, in particular the parallel EFD-BPM, with the parallel ADI in [16] shows that the explicit methods are still more efficient than the parallel ADI-BPM for several reasons. The following comparisons more than compensate for the constraints imposed on the explicit methods:

- 1- The ordinary operation counts of the explicit methods are far less than the ADI-BPM.
- 2- The explicit methods use far less computer memory than the ADI.
- 3- All of the parallel-processing performance indicators, for the explicit methods, improve with increasing number of grid points and/or the number of processors where on the contrary, the efficiency of the ADI in [16] reaches a maximum level and decreases afterwards due to the increase in the communication cost. It should be noticed that the method in [16] uses the fourth-order finite-difference approximation for the transverse variations which means an increase in the local computations. In other words, the second-order approximation of the ADI may lead to a less efficient parallel algorithm. In addition, the communication speed for transferring information between processors is around 20 M bytes/s for the CM-5 and is around 1 M bytes/s for the supercluster.

4- The parallel implementation of the ADI in [16] remains restricted to MIMD implementations, but the explicit methods are suitable for both MIMD and SIMD implementations.

5- The parallel explicit methods are very simple to program where, on the other hand, the implementation of the ADI in [16] is very complicated. As a matter of fact, the EFD-BPM equation is the most simple formulation for approximating the parabolic equation. Also the EFD-BPM is easily extendable to analyse nonlinear structures⁷.

6- The implementation of the fourth-order finite-difference approximation to the transverse variations of the EFD-BPM results in a relaxation for the transverse mesh sizes and as a consequence a larger propagational step size could be used. It is to be observed that this extension can be included easily to the existing parallel EFD-BPM for both MIMD and SIMD implementations.

7- There is a possibility to improve the stability of the EFD-BPM using the MEFD-BPM technique⁸.

Finally, the solution of the parabolic equation, discussed in this chapter, is common for many large major mathematical applications; the same implementations of parallel explicit techniques could be used to speed up their execution times.

REFERENCES

- [1] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Methods", *IEEE Phot. Tech. Lett.*, **6**, 848-850, July 1994.
- [2] Masoudi, H. M. and Arnold, J. M., "Parallel Three-Dimensional Finite-Difference Beam Propagation Methods", *Int. J. of Num. Mod.*, **8**, 95-107, March-April 1995.
- [3] Masoudi, H. M. and Arnold, J. M., "Parallel-Processing Finite-Difference Beam Propagation Methods", Presented at the *International Symposium on Guided-Wave Optoelectronics: Device Characterization, Analysis and Design*, IEEE and Optical Society of America in Polytechnic University, New York, October 1994.
- [4] Modi, J. J., "Parallel Algorithms and Matrix Computation", Oxford Series Book, 1988.
- [5] Lewis, T. and El-Rewini, H., "Introduction to Parallel Computing", Printce Hall Int., New Jersey, 1992.

⁷ See chapter 7 (page 148).

⁸ See chapter 5 (page 96).

- [6] DeRaedt, H., "Product Formula Algorithms for Solving the Time Dependent Schrodinger Equation", *Comp. Phys. Rep.*, **7**, 1-72, 1987
- [7] Richardson, J., "Visualizing Quantum Scattering on the CM-2 Supercomputer", *Comp. Phys. Comm.*, **63**, 84-94, 1991.
- [8] Richardson, J., "Computational Physics on the CM-2 Supercomputer", *Phys. Rep.*, **207**, 305-320, 1991.
- [9] Hadley, G. R., "Transparent Boundary Condition for the Beam Propagation Method", *IEEE J. Quant. Elect.*, **28**, 363-370, 1992.
- [10] Yevick, D. and Hermansson, B., "Efficient Beam Propagation Techniques", *IEEE J. Quant. Elect.*, **26**, 109-112, 1990.
- [11] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. "The Art of scientific computing: numerical recipes in FORTRAN", Cambridge Uni. Press, 1992.
- [12] Chung, Y. and Dagli, N., "Analysis of Z-Invariant and Z-Variant Semiconductor Rib Waveguides by Explicit Finite Difference Beam Propagation Method with Nonuniform Mesh Configuration", *IEEE J. Quant. Elect.*, **27**, 2296-2305, 1991.
- [13] Robertson, M. J., Ritchie, S., and Dayan, P., "Semiconductor Waveguides: Analysis of Optical Propagation in Single Rib Structures and Dielectric Couplers", *IEE Proc. Pt. J.*, **132**, 336-342, 1985.
- [14] Feit, M. D. and Fleck, J. A., "Analysis of Rib Waveguides and Couplers by the Propagating Beam Method", *J. Opt. Soc. Am. A*, **7**, 73-79, 1990.
- [15] Marcuse, D., "Theory of Dielectric Optical Waveguides", Academic Press, San Diego, 1991.
- [16] Wang, M. and Vanka, S. P., "A parallel ADI Algorithm for High-Order Finite-Difference Solution of the Unsteady Heat Conduction Equation, and Its Implementation on The CM-5", *Num. Heat Transf. Pt. B*, **24**, 143-159, 1993.

Part III

Nonlinear Optical Waveguides

Chapter 7

Parallel Second Harmonic Generation BPM

7.1 Introduction

The process of converting energy from one optical wave, with a frequency ω , to another wave that oscillates with double (2ω) the original frequency is known as the Second Harmonic Generation (SHG) process. This phenomenon exists in nonlinear materials containing second-order nonlinear susceptibility $\chi^{(2)}$ [1]. In the past, most of the research on this effect was focused on an efficient method to convert to second harmonic [2-4]. However, the existence of second order nonlinearity leads also to another interesting and important effect, in addition to the conversion, which is a nonlinear phase shift to the fundamental beam [5]. At low conversion, and off phase-matching between the fundamental and the second-harmonic, this nonlinear phase shift behaves similarly to the Kerr effect of $\chi^{(3)}$ [5]. Recently this nonlinear phase shift process has received much attention experimentally, and theoretically to achieve all optical switching [5-7]. These advances in the nonlinear effects of optical waveguides show that there is a need to model such devices accurately and efficiently [2-7] in order to understand the behaviour of such effects. On the other hand, waveguides that contain second-order nonlinear susceptibility (and nonlinear effects in general) are very difficult to model using analytically based techniques like Coupled Mode Theory, and even more difficult when the devices contain multiple waveguides in which they have geometrical and/or material change in all three directions. Other methods, based on numerical analysis like the BPM, are much better suited to such devices. For second-harmonic optical devices, the BPM has been used to simulate 2-D [8,9] and 3-D [10] (fibers, with the assumption that the source field is undepleted) devices in both FFT and Finite-Difference (FD) forms.

The previous chapter of this thesis shows that the implementations of the linear versions of the Explicit FD method (EFD-BPM) and the Real Space method (RS-BPM) on

parallel machines will speed up their execution times tremendously because of their nature as explicit methods which reduces communication between the parallel processors [11-13]. On the other hand, other BPM methods do not possess this feature and consequently are not highly parallel and will not gain as much in speed up in the parallel environment. Also it has been demonstrated that the parallel EFD-BPM is several times faster, per propagational step, than the parallel RS-BPM [11-13].

In this chapter, the parallel EFD-BPM is extended to study three-dimensional waveguides in the presence of a second order nonlinearity $\chi^{(2)}$ where the source field is allowed to deplete. In the following, this method is referred to as the Second Harmonic Generation EFD-BPM (or in short SHG-EFD-BPM) [14-17]. The rest of this chapter consists of three parts in addition to a conclusion at the end. Section 7.2 shows the derivation of the nonlinear coupled wave equations both in 3-D (for the SHG-EFD-BPM) and in 1-D, which is commonly used in the literature to validate experimental results and to study possible concepts for optical devices [5-7]. Section 7.3 shows the implementation of the SHG-EFD-BPM on the two parallel machines, the transputer array and the Connection Machine of chapter 4. Then, in section 7.4, the parallel SHG-EFD-BPM is used to analyse semiconductor rib waveguides containing a second order nonlinearity. As a first test, a phase-matching case of a semiconductor rib waveguide where the effective indices of the fundamental and the second harmonic are equal, has been analysed. The second test consists of a non-phased match case in which the two effective indices of the fundamental and the second harmonic are not equal. In this case, the Quasi Phase Matching (QPM) technique has been used to model the relative phase between the fundamental and the second harmonic fields. Also in the same section the parallel SHG-EFD-BPM results have been compared with the solution of the reduced 1-D coupled wave equations where the Runge-Kutta numerical technique has been used for the solution of the 1-D model.

7.2 Formulation

In this section the derivation of the scalar 3-D (BPM) and the 1-D (plane wave) equations for the analysis of second harmonic generation will be shown. The 3-D is based on the EFD-BPM approach where the 1-D is used later in the analysis to validate the results of the 3-D BPM.

7.2.1 The Second Harmonic Generation EFD-BPM

Combining Eqs. 2.3 and 2.14 of chapter 2, the scalar wave equation for the propagation of an electric field E of a given polarisation in a material with a refractive index n and a homogenous second order nonlinear susceptibility $\chi^{(2)}$ can be written as [1,15]

$$\nabla^2 E = \mu_0 \epsilon_0 n^2 \frac{\partial^2 E}{\partial t^2} + \mu_0 \epsilon_0 \frac{\partial^2 P}{\partial t^2} \quad (7.1)$$

where P is the nonlinear polarisation in the material which can be approximated, for c.w. operation, as $P = \chi^{(2)} EE$. Here it is assumed that the fields are linearly polarised and that the vector nature of the fields can be ignored. This is a good first approximation for paraxial problems of the type considered here for which the BPM is appropriate. Let us consider at the moment the propagation of three fields at three different frequencies ω_1 , ω_2 and ω_3 , while extracting a common reference phase in the direction of propagation z . These fields can be written as [1]

$$E^{\omega_1}(x, y, z, t) = \frac{1}{2} \left[\Phi^{\omega_1}(x, y, z) e^{j(\omega_1 t - k_1 n_0 z)} + \text{c.c.} \right] \quad (7.2a)$$

$$E^{\omega_2}(x, y, z, t) = \frac{1}{2} \left[\Phi^{\omega_2}(x, y, z) e^{j(\omega_2 t - k_2 n_0 z)} + \text{c.c.} \right] \quad (7.2b)$$

$$E^{\omega_3}(x, y, z, t) = \frac{1}{2} \left[\Phi^{\omega_3}(x, y, z) e^{j(\omega_3 t - k_3 n_0 z)} + \text{c.c.} \right] \quad (7.2c)$$

where n_0 is a reference refractive index for the problem, k_1 , k_2 and k_3 are the free space wave numbers at the three frequencies ω_1 , ω_2 , and ω_3 respectively, where $\omega_3 = \omega_1 + \omega_2$ and c.c. is the complex conjugate. Inserting (7.2) in (7.1) gives respectively [1]

$$\begin{aligned} \nabla^2 \left[\Phi^{\omega_1} e^{j(\omega_1 t - k_1 n_0 z)} + \text{c.c.} \right] &= \mu_0 \epsilon_0 n_1^2 \left[(-\omega_1^2) \Phi^{\omega_1} e^{j(\omega_1 t - k_1 n_0 z)} + \text{c.c.} \right] \\ &\quad - \frac{\mu_0 \epsilon_0}{2} \left[\chi^{(2)}(\omega_1) (\omega_3 - \omega_2)^2 \Phi^{\omega_3} \Phi^{*\omega_2} e^{j[(\omega_3 - \omega_2)t - (k_3 - k_2)n_0 z]} + \text{c.c.} \right] \end{aligned} \quad (7.3a)$$

$$\begin{aligned} \nabla^2 \left[\Phi^{\omega_2} e^{j(\omega_2 t - k_2 n_0 z)} + \text{c.c.} \right] &= \mu_0 \epsilon_0 n_2^2 \left[(-\omega_2^2) \Phi^{\omega_2} e^{j(\omega_2 t - k_2 n_0 z)} + \text{c.c.} \right] \\ &\quad - \frac{\mu_0 \epsilon_0}{2} \left[\chi^{(2)}(\omega_2) (\omega_3 - \omega_1)^2 \Phi^{\omega_3} \Phi^{*\omega_1} e^{j[(\omega_3 - \omega_1)t - (k_3 - k_1)n_0 z]} + \text{c.c.} \right] \end{aligned} \quad (7.3b)$$

$$\begin{aligned} \nabla^2 \left[\Phi^{\omega_3} e^{j(\omega_3 t - k_3 n_0 z)} + \text{c.c.} \right] &= \mu_0 \epsilon_0 n_3^2 \left[(-\omega_3^2) \Phi^{\omega_3} e^{j(\omega_3 t - k_3 n_0 z)} + \text{c.c.} \right] \\ &\quad - \frac{\mu_0 \epsilon_0}{2} \left[\chi^{(2)}(\omega_3) (\omega_1 + \omega_2)^2 \Phi^{\omega_1} \Phi^{\omega_2} e^{j[(\omega_1 + \omega_2)t - (k_1 + k_2)n_0 z]} + \text{c.c.} \right] \end{aligned} \quad (7.3b)$$

Where $\chi^{(2)}(\omega_1) = \chi^{(2)}(\omega_1; \omega_3, -\omega_2)$, $\chi^{(2)}(\omega_2) = \chi^{(2)}(\omega_2; \omega_3, -\omega_1)$ and $\chi^{(2)}(\omega_3) = \chi^{(2)}(\omega_3; \omega_1, \omega_2)$. Let us consider Eq. 7.3a and rewrite it, after multiplying with $e^{-j\omega_1 t}$, as

$$\nabla^2 \left[\Phi^{\omega_1} e^{-jk_1 n_0 z} \right] + k_1^2 n_1^2 \Phi^{\omega_1} e^{-jk_1 n_0 z} + \frac{k_1^2 \chi^{(2)}(\omega_1)}{2} \left[\Phi^{\omega_3} \Phi^{*\omega_2} e^{-j[(k_3 - k_2)n_0 z]} \right] = 0 \quad (7.4)$$

Expanding the Laplacian operator above into

$$\begin{aligned} \frac{\partial^2 \Phi^{\omega_1}}{\partial z^2} e^{-jk_1 n_0 z} - 2jk_1 n_0 \frac{\partial \Phi^{\omega_1}}{\partial z} e^{-jk_1 n_0 z} - k_1^2 n_0^2 \Phi^{\omega_1} e^{-jk_1 n_0 z} + \nabla_{\perp}^2 \left[\Phi^{\omega_1} e^{-jk_1 n_0 z} \right] \\ + k_1^2 n_1^2 \Phi^{\omega_1} e^{-jk_1 n_0 z} + \frac{k_1^2 \chi^{(2)}(\omega_1)}{2} \left[\Phi^{\omega_3} \Phi^{*\omega_2} e^{-j[(k_3 - k_2)n_0 z]} \right] = 0 \end{aligned} \quad (7.5)$$

where

$$\nabla_{\perp}^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

Eq. 7.5 is similar to the linear counterpart in (3.3). Under the parabolic approximation in (3.4), we may rewrite Eq. 7.5, after neglecting the term containing the second derivative with respect to z , as

$$2jk_1 n_0 \frac{\partial \phi^{\omega_1}}{\partial z} = \nabla_{\perp}^2 \phi^{\omega_1} + k_1^2 (n_1^2 - n_0^2) \phi^{\omega_1} + \frac{k_1^2 \chi^{(2)}(\omega_1)}{2} \left[\phi^{\omega_3} \phi^{*\omega_2} e^{-j[(k_3 - k_1 - k_2)n_0 z]} \right] \quad (7.6a)$$

Using a similar approach to the derivation of Eq. 7.6a, the other two equations in (7.3b) and (7.3c) can be written as

$$2jk_2 n_0 \frac{\partial \phi^{\omega_2}}{\partial z} = \nabla_{\perp}^2 \phi^{\omega_2} + k_2^2 (n_2^2 - n_0^2) \phi^{\omega_2} + \frac{k_2^2 \chi^{(2)}(\omega_2)}{2} \left[\phi^{\omega_3} \phi^{*\omega_1} e^{-j[(k_3 - k_1 - k_2)n_0 z]} \right] \quad (7.6b)$$

$$2jk_3 n_0 \frac{\partial \phi^{\omega_3}}{\partial z} = \nabla_{\perp}^2 \phi^{\omega_3} + k_3^2 (n_3^2 - n_0^2) \phi^{\omega_3} + \frac{k_3^2 \chi^{(2)}(\omega_3)}{2} \left[\phi^{\omega_1} \phi^{\omega_2} e^{j[(k_3 - k_1 - k_2)n_0 z]} \right] \quad (7.6c)$$

where ϕ^{ω_1} , ϕ^{ω_2} and ϕ^{ω_3} are the parabolic fields which are paraxial approximations to the Φ 's (the Helmholtz fields).

For SHG, two of the three fields in (7.6) are of the same frequency, therefore let $\omega_1 = \omega_2 = \omega$ and the third is the summation of these two frequencies with $\omega_3 = \omega_1 + \omega_2 = 2\omega$. and $k_1 = k_2 = k_3/2$. Thus Eqs. 7.6a and 7.6b are the same and one only of them is needed in the analysis. From Eqs. 7.6a and 7.6c we may write the following coupled parabolic wave equations in three-dimensions for both the fundamental and the second harmonic waves [14-17]

$$2jk_0n_0 \frac{\partial \phi^f}{\partial z} = \frac{\partial^2 \phi^f}{\partial x^2} + \frac{\partial^2 \phi^f}{\partial y^2} + k_0^2(n_f^2 - n_0^2)\phi^f + k_0^2\chi^{(2)}\phi^s\phi^{f*} \quad (7.7a)$$

$$4jk_0n_0 \frac{\partial \phi^s}{\partial z} = \frac{\partial^2 \phi^s}{\partial x^2} + \frac{\partial^2 \phi^s}{\partial y^2} + 4k_0^2(n_s^2 - n_0^2)\phi^s + 2k_0^2\chi^{(2)}\phi^f\phi^f \quad (7.7b)$$

where $k_0 = k_1$, $\phi^f = \phi^{\omega_1}$ and $\phi^s = \phi^{\omega_3}$. For convenience, throughout the rest of this work subscripts or superscripts for both f and s are related to the fundamental wave and the second harmonic wave respectively. In Eq. 7.7, it is assumed that the second order nonlinearity is homogeneous and is defined through $\chi^{(2)} = \chi^{(2)}(2\omega; \omega, \omega) = \chi^{(2)}(\omega; 2\omega, -\omega)/2$ [5-7] and the source field ϕ^f is allowed to deplete during propagation.

As for the EFD-BPM, let us use the central finite-difference approximations in Eq. 3.27 to replace the partial derivatives in Eq. 7.7. The discretisation of Eq. 7.7 leads to the following coupled formulation which is second-order accurate with respect to z [14-17]

$$\begin{aligned} \phi_{i,m}^f(z + \Delta z) &= \phi_{i,m}^f(z - \Delta z) + d_x [\phi_{i+1,m}^f(z) + \phi_{i-1,m}^f(z)] \\ &\quad + d_y [\phi_{i,m+1}^f(z) + \phi_{i,m-1}^f(z)] + b_f \phi_{i,m}^f(z) + \alpha_x \phi_{i,m}^s(z) \phi_{i,m}^{f*}(z) \end{aligned} \quad (7.8a)$$

$$\begin{aligned} \phi_{i,m}^s(z + \Delta z) &= \phi_{i,m}^s(z - \Delta z) + \frac{d_x}{2} [\phi_{i+1,m}^s(z) + \phi_{i-1,m}^s(z)] \\ &\quad + \frac{d_y}{2} [\phi_{i,m+1}^s(z) + \phi_{i,m-1}^s(z)] + b_s \phi_{i,m}^s(z) + \alpha_x \phi_{i,m}^f(z) \phi_{i,m}^f(z) \end{aligned} \quad (7.8b)$$

where

$$\begin{aligned}
 d_\rho &= -j \frac{\Delta z}{n_o k_o \Delta \rho^2} \quad , \quad (\rho = x, y) \\
 b_f &= -j \frac{\Delta z}{n_o k_o} \left[-\frac{2}{\Delta x^2} - \frac{2}{\Delta y^2} + k_o^2 (n_f^2 - n_o^2) \right] \\
 b_s &= -j \frac{\Delta z}{2n_o k_o} \left[-\frac{2}{\Delta x^2} - \frac{2}{\Delta y^2} + 4k_o^2 (n_s^2 - n_o^2) \right] \\
 \alpha_x &= -j \frac{\Delta z k_o \chi^{(2)}}{n_o}
 \end{aligned}$$

where i and m represent the discretisation of the transverse co-ordinates x and y . The above two equations represent the propagation of the fundamental and the second harmonic fields or the SHG-EFD-BPM.

7.2.2 The simplified 1-D Second Harmonic Generation

Let us also write the usual approximate 1-D coupled wave equations for the nonlinear wave equation; later in this work the solution of the 1-D equations will be used to compare with the results of the SHG-EFD-BPM. The derivation of the 1-D approximation is the same as for the 3-D approach without x and y variations. If we project the fields onto the 3-D eigenfunctions (the mode distributions) and write the 3-D fields in Eq. 7.7 as [14-17]

$$\left. \begin{aligned}
 \phi^f &= \frac{\Psi^f(z) \mathcal{E}^f(x, y)}{\sqrt{R_1 R_2}} \\
 \phi^s &= \frac{\Psi^s(z) \mathcal{E}^s(x, y)}{R_1}
 \end{aligned} \right\} \quad (7.9)$$

where

$$\begin{aligned}
 R_1 &= \frac{S_o}{S_1} \quad , \quad R_2 = \frac{S_o}{S_2} \\
 S_1 &= \iint_A (\mathcal{E}^f)^2 dA \quad , \quad S_2 = \iint_A (\mathcal{E}^s)^2 dA \\
 S_o &= \iint_A (\mathcal{E}^f)^2 \mathcal{E}^s dA \quad ,
 \end{aligned}$$

the $\psi(z)$'s are dependent only on z , $\mathcal{E}(x, y)$'s are the 3-D z -independent linear mode fields and A is the cross sectional area. Inserting (7.9) in (7.6a) and (7.6c) leads to

$$2jk_0 n_{\text{eff}}^f \frac{\partial}{\partial z} \left(\frac{\psi^f \mathcal{E}^f}{\sqrt{R_1 R_2}} \right) = \nabla_{\perp}^2 \left(\frac{\psi^f \mathcal{E}^f}{\sqrt{R_1 R_2}} \right) + k_0^2 (n_f^2 - (n_{\text{eff}}^f)^2) \left(\frac{\psi^f \mathcal{E}^f}{\sqrt{R_1 R_2}} \right) + k_0^2 \chi^{(2)} \left[\left(\frac{\psi^s \mathcal{E}^s}{R_1} \right) \left(\frac{\psi^f \mathcal{E}^f}{\sqrt{R_1 R_2}} \right)^* e^{-j\Delta k' z} \right] \quad (7.10a)$$

$$4jk_0 n_{\text{eff}}^s \frac{\partial}{\partial z} \left(\frac{\psi^s \mathcal{E}^s}{R_1} \right) = \nabla_{\perp}^2 \left(\frac{\psi^s \mathcal{E}^s}{R_1} \right) + 4k_0^2 (n_s^2 - (n_{\text{eff}}^s)^2) \left(\frac{\psi^s \mathcal{E}^s}{R_1} \right) + 2k_0^2 \chi^{(2)} \left[\left(\frac{\psi^f \mathcal{E}^f}{\sqrt{R_1 R_2}} \right) \left(\frac{\psi^f \mathcal{E}^f}{\sqrt{R_1 R_2}} \right) e^{j\Delta k' z} \right] \quad (7.10b)$$

where the wave vector mismatch is $\Delta k' = 2k_0 (n_{\text{eff}}^s - n_{\text{eff}}^f)$ (note that the extracted reference phase is not common for both fields; the reason for that is to derive the usual 1-D relations used in the literature). Noticing that the summation of the first and the second terms in the right hand sides of Eq. 7.10 is zero (see Eq. 2.47), then multiplying Eq. 7.10a by ψ^f and Eq. 7.10b by ψ^s and integrating over the cross sectional area A leads to

$$2jk_0 n_{\text{eff}}^f \frac{\partial \psi^f}{\partial z} \left(\frac{\iint_A (\mathcal{E}^f)^2 dA}{\sqrt{R_1 R_2}} \right) = k_0^2 \chi^{(2)} \psi^s \psi^{f*} \left(\frac{\iint_A (\mathcal{E}^f)^2 \mathcal{E}^s dA}{R_1 \sqrt{R_1 R_2}} \right) e^{-j\Delta k' z} \quad (7.11a)$$

$$2jk_0 n_{\text{eff}}^s \frac{\partial \psi^s}{\partial z} \left(\frac{\iint_A (\mathcal{E}^s)^2 dA}{R_1} \right) = k_0^2 \chi^{(2)} \psi^f \psi^f \left(\frac{\iint_A (\mathcal{E}^f)^2 \mathcal{E}^s dA}{R_1 R_2} \right) e^{j\Delta k' z} \quad (7.11b)$$

The above two equations may be simplified into the following familiar 1-D nonlinear coupled wave equations as [5-7]

$$\frac{d\psi^f}{dz} = -j \frac{k_0 \chi^{(2)}}{2n_{\text{eff}}^f} \psi^s \psi^{f*} e^{-j\Delta k' z} \quad (7.12a)$$

$$\frac{d\psi^s}{dz} = -j \frac{k_0 \chi^{(2)}}{2n_{\text{eff}}^s} \psi^f \psi^f e^{j\Delta k' z} \quad (7.12b)$$

The relationship between the 3-D model and 1-D model can be extracted from the power relation by inserting Eq. 7.9 in the power equation which leads to [14-17]

$$P^f = \frac{n_{\text{eff}}^f |\Psi^f|^2}{2Z_o R_1 R_2} \iint_A (\mathcal{E}^f)^2 dA = \frac{n_{\text{eff}}^f |\Psi^f|^2}{2Z_o} A_{\text{eff}} \quad (7.13a)$$

$$P^s = \frac{n_{\text{eff}}^s |\Psi^s|^2}{2Z_o R_1^2} \iint_A (\mathcal{E}^s)^2 dA = \frac{n_{\text{eff}}^s |\Psi^s|^2}{2Z_o} A_{\text{eff}} \quad (7.13b)$$

where the effective area A_{eff} is defined as [14-17]

$$A_{\text{eff}} = \frac{S_1}{R_1 R_2} = \frac{S_2}{R_1^2} = \frac{S_1 S_2}{S_o^2} \quad (7.14)$$

and Z_o is the free space impedance.

For perfect phase matching ($\Delta k' = 0$), Eq.7.12 leads to the well known solution for the normalised intensities of the fundamental and the second harmonic as [5]

$$\left. \begin{aligned} \frac{I^f}{I_o} &= \text{sech}^2(\Gamma z) \\ \frac{I^s}{I_o} &= \tanh^2(\Gamma z) \end{aligned} \right\} \quad (7.15)$$

where

$$\Gamma = \frac{k_o \chi^{(2)} |\Psi_o^f|}{2\sqrt{n_{\text{eff}}^f n_{\text{eff}}^s}}$$

and I_o and Ψ_o^f are the initial intensity and field respectively of the fundamental at $z = 0$. For the non-phased-matched case Eq. 7.12 can be solved using Jacobian elliptic functions [3,6] or numerically using the Runge-Kutta method [19]. In the following analysis, the fourth order Runge-Kutta method was used to solve the 1-D coupled wave equation in (7.12) [19].

7.3 Parallel Implementations

Two parallel computer codes for the SHG-EFD-BPM in Eq. 7.8 have been developed which run on the 64-transputer array and the Connection Machine (CM-200) of chapter 4, using the same techniques of the linear EFD-BPM discussed in section 6.3 [11-13]. Two separate computational windows, one for the fundamental field and the other for the

second harmonic field, have been used while allowing them to share common data for the calculations of the coupled terms. The implementation of the SHG-EFD-BPM on the transputer array has been done by dividing the transverse mesh points of the discretised fundamental field into 2-D equal blocks of data where each block is assigned to one processor of the 2-D topology of the transputer array. The same has been done for the second harmonic transverse mesh points field using the same 2-D topology of the fundamental field computations. In other words, each processor computes the local computational block data of both the fundamental and the second harmonic fields and they exchange only values at the local border points at the edges of the blocks with neighbouring processors (see Fig. 7.1) [14,15].

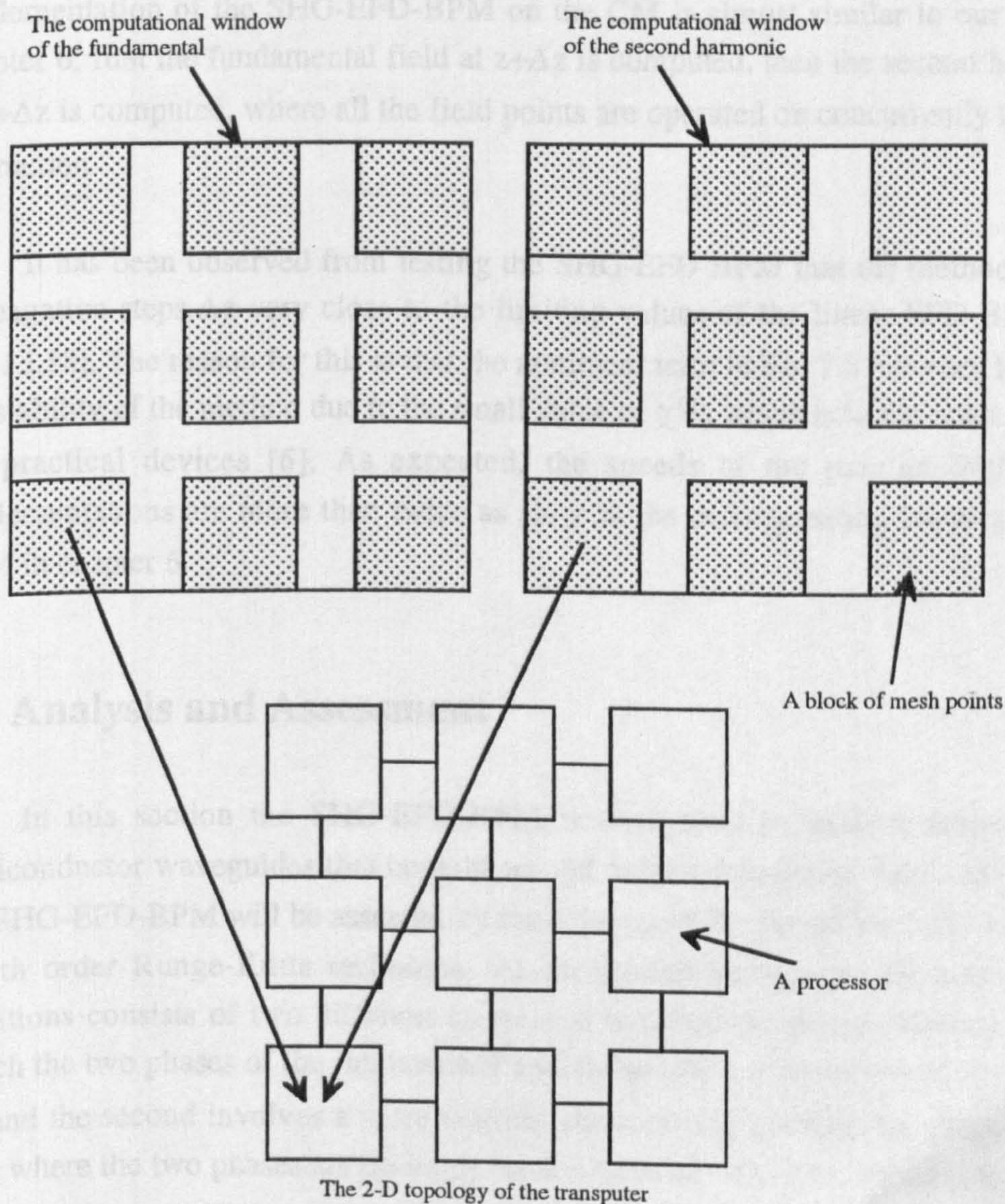


Fig. 7.1 The implementation of the SHG-EFD-BPM on the transputer array (see text for details).

The alternative choice of parallelising the SHG-EFD-BPM on the transputer is to assign the computations of the fundamental and the second harmonic fields to two different 2-D topologies (e.g. each of 32 processors), since the computations of both fields at $z+\Delta z$ are independent and could be solved concurrently. However both topologies, in the latter case, need to exchange all field points values for each propagational step and obviously this involves a massive communication time. Comparison between the two parallel techniques above shows that the first method of using the same 2-D topology for both fields is more efficient because the transverse mesh points of both fields are local to each processor, which means that no time is lost in transferring blocks of data. For this reason the first technique was used in the following analysis. On the other hand, the implementation of the SHG-EFD-BPM on the CM is almost similar to our linear one in chapter 6; first the fundamental field at $z+\Delta z$ is computed, then the second harmonic field at $z+\Delta z$ is computed, where all the field points are operated on concurrently using a single instruction.

It has been observed from testing the SHG-EFD-BPM that the method is stable for propagation steps Δz very close to the limiting values of the linear EFD-BPM in (3.29) [11-13,18]. The reason for this is that the nonlinear term in Eq. 7.8 has very little effect on the stability of the method due to the small value of $\chi^{(2)}$, which is in the order of 10^{-12} m/V for practical devices [6]. As expected, the speeds of the parallel SHG-EFD-BPM implementations are more than twice as slow as the corresponding linear parallel EFD-BPM in chapter 6.

7.4 Analysis and Assessment

In this section the SHG-EFD-BPM will be used to analyse three-dimensional semiconductor waveguides that contain second order nonlinearity. Some of the results of the SHG-EFD-BPM will be assessed by the solution of the simplified 1-D model using the fourth order Runge-Kutta technique. As mentioned before, the solution of the SHG equations consists of two different cases; one is called the phase-matched condition in which the two phases of the fundamental and the second harmonic waves are equal ($\Delta k' = 0$), and the second involves a more realistic situation and is called the non-phased match case where the two phases are different from each other ($\Delta k' \neq 0$). These two cases will be analysed using two different semiconductor waveguides.

The Phase-Matched Case

The first numerical test for the parallel SHG-EFD-BPM involves a nearly perfect phase-matched waveguide. The rib waveguide in Fig. 6.13 has been used with the following parameters (all length dimensions are in μm): $\lambda_f = 1.55$, $\lambda_s = 0.775$, $n_g^f = 3.44$, $n_g^s = 3.4062406$, $n_s = 3.34$, $\chi^{(2)} = 300 \text{ pm/V}$. Other parameters are the same as those of structure 1 in Table 6.1. The second order nonlinear term $\chi^{(2)}$ will be restricted to the guiding region of the waveguide. In this example, n_g^s is chosen such that the two modes' effective indices are equal. Using $\Delta x = \Delta y = 0.1 \mu\text{m}$ and equal lengths of $8 \mu\text{m}$ for the x and y window sizes, the computed first guided mode effective indices of both fields, using $\Delta z = 0.025$ and $n_o = n_s$, are $n_{\text{eff}}^f = 3.39236226$ and $n_{\text{eff}}^s = 3.39236228$. In the following analysis, the first guided mode of the fundamental waveguide is used as an input, and zero initial field is assumed for the second harmonic waveguide. Fig. 7.2 shows the normalised intensities of both the fundamental and the second harmonic versus the longitudinal distance $z(\mu\text{m})$ for the phase-matched case with an input power of 27.2 W.

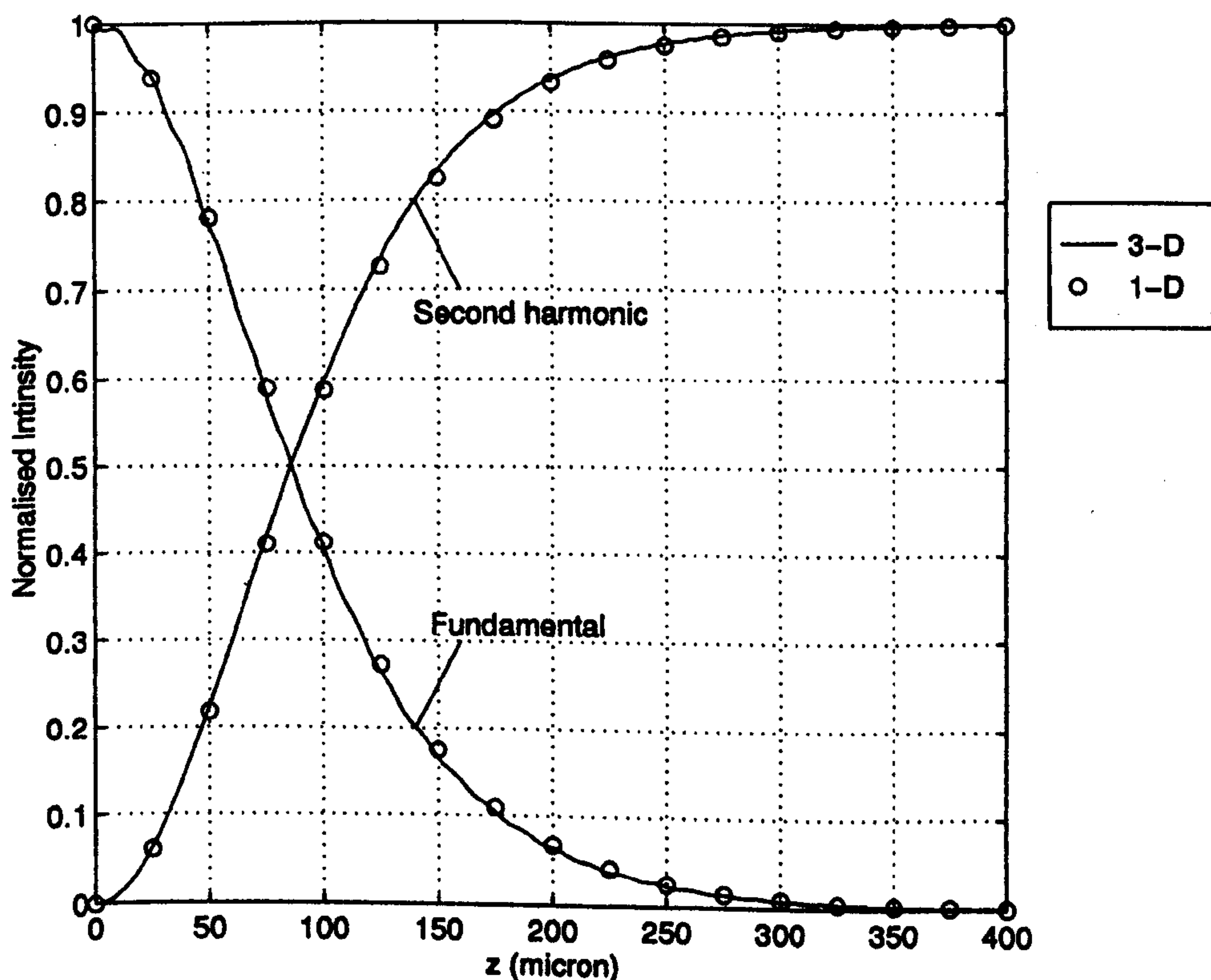


Fig. 7.2 The normalised intensity of a phased matched waveguide for both of the fundamental and the SHG versus the longitudinal distance $z(\mu\text{m})$, using the parallel SHG-EFD-BPM (3-D) and the solution of Eq. 7.12 (1-D), with an input power = 27.2 W.

The figure shows the complete power exchange from the fundamental to the SH due to the matching between the two modes' effective indices ($\Delta k = 0$). The computation of the normalised intensities of the modes has been carried out numerically by evaluating the square of the magnitude of the projected total BPM field onto the normalised mode and then normalising with respect to the input intensity, which can be expressed as

$$I_{\text{norm}}^{f,s}(z) = \frac{\left| \iint_{\Lambda} \mathcal{E}^{f,s} \phi^{f,s}(z) dA \right|^2}{\iint_{\Lambda} |\mathcal{E}^{f,s}|^2 dA \iint_{\Lambda} |\phi^f(0)|^2 dA} \quad (7.16)$$

where f,s means the fundamental or the second harmonic. In the same figure, the solution of Eq. 7.12 has been included for comparison, using an effective area (computed from Eq. 7.14) of the 3-D fields of $1.896 \mu\text{m}^2$. Clearly, the figure shows the excellent agreement of the two results.

The Quasi Phase-Matched Case

The second numerical test for the SHG-EFD-BPM involves a more practical waveguide that contains a second order nonlinear coefficient $\chi^{(2)}$. The rib waveguide in Fig. 6.13 was used to study a nonlinear semiconductor device with the same parameters as for the phased-matched case except for the following alterations (all dimensions are in μm): $n_s = 3.4$, $n_g^f = 3.5$, $n_g^s = 3.6$ and $\chi^{(2)} = 300 \text{ pm/V}$ (the parameters of this waveguide are chosen to match real values [6]). The design of this structure shows that it is strongly guiding at both frequencies although it is a single mode waveguide at the fundamental frequency and a multi-mode waveguide at the second harmonic frequency. Using equal lengths of $8 \mu\text{m}$ for the x and y window sizes and $\Delta x = \Delta y = 0.1$, the computed first guided mode effective indices of both fields, using $\Delta z = 0.025$ and $n_o = n_s$, are $n_{\text{eff}}^f = 3.45299169$ and $n_{\text{eff}}^s = 3.58591145$. Fig. 7.3 shows the normalised field distribution of the first guided mode of the fundamental waveguide. The computed effective area, using Eq. 7.14, of the 3-D fields is $1.92 \mu\text{m}^2$. The parameters of this structure show that it is a non-phase matched waveguide in which Quasi Phase Matching (QPM) techniques could be used to alter the relative phase difference between the two fields [2]. QPM is used for efficient optical SHG and other nonlinear optical processes [2,5-7]. In practical terms, QPM is achieved by introducing structural periodic gratings, in the direction of propagation, of the nonlinear medium. The BPM mismatch wave vector, for QPM, is defined as $\Delta k = k_z^{p,s} - 2k_z^{p,f} - 2\pi/\Lambda$, where $k_z^{p,f}$ and $k_z^{p,s}$ are the parabolic wave numbers of the

fundamental and the second harmonic fields respectively, and Λ is the grating period. The computed grating length, for $\Delta k = 0$ (between the first guided modes of both frequencies), of this waveguide is equal to $5.63269267 \mu\text{m}$. In the following simulation a grating was used in which the second order nonlinear coefficient $\chi^{(2)}$ has either its full value or is zero in alternate half-periods of the grating (see Fig. 7.4).

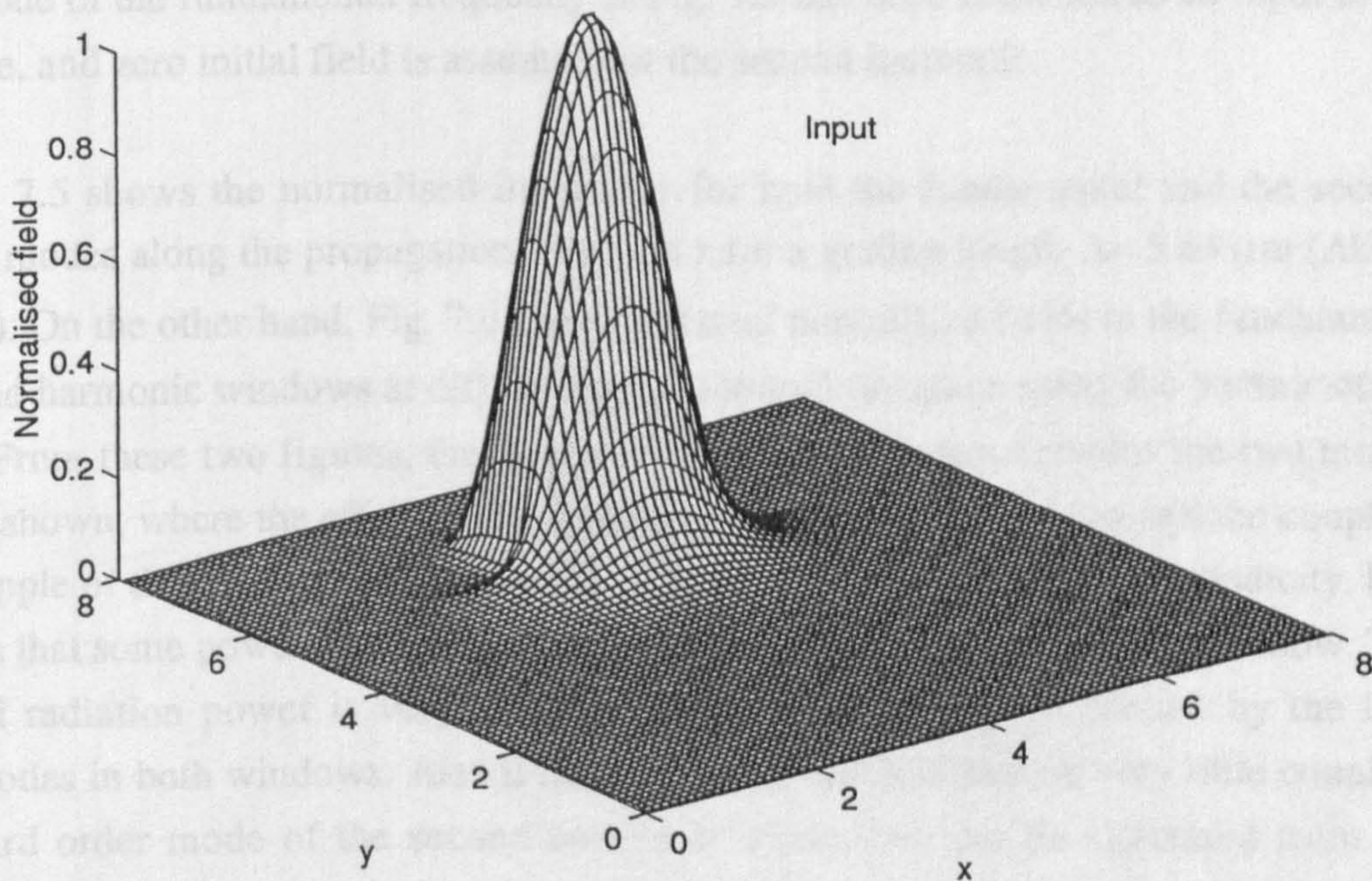


Fig. 7.3 The normalised field distribution of the first guided mode of the fundamental waveguide which has been used as an input for the analysis of QPM case.

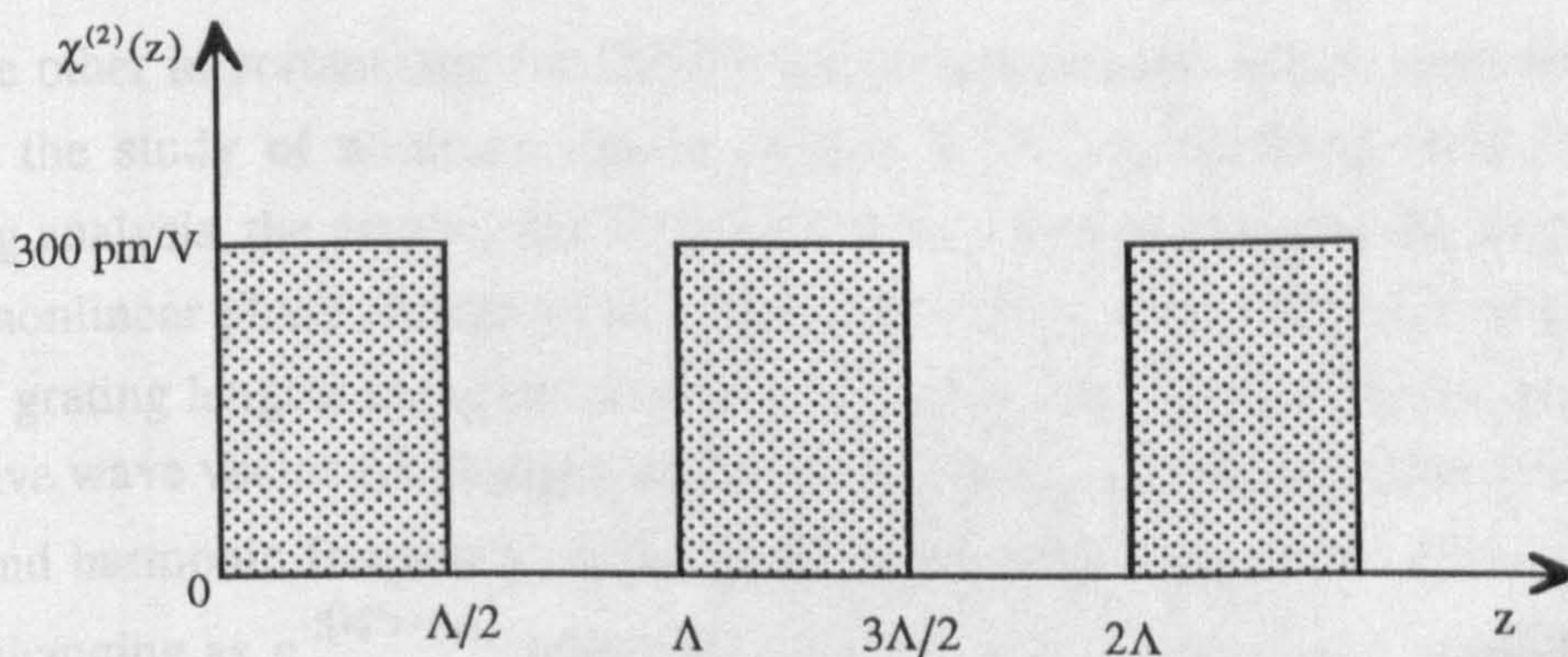


Fig. 7.4 The rectangular periodic grating of the nonlinear medium $\chi^{(2)}$ used for the quasi phased-matched case.

This geometry might be realised in a semiconductor asymmetric quantum-well by selective-area disordering, which periodically annihilates the $\chi^{(2)}$ coefficient produced by the quantum-well breaking the symmetry of the bulk material [20]. It is known that QPM reduces the effectiveness of the second order nonlinear coefficient $\chi^{(2)}$ by a factor $2/\pi$ in the case of domain reversal and by a factor of $1/\pi$ in the case of domain disordering (where $\chi^{(2)}$ is periodically reduced to zero) [2]. In the following computations, the first guided mode of the fundamental frequency of Fig. 7.3 has been launched as an input to the waveguide, and zero initial field is assumed for the second harmonic.

Fig. 7.5 shows the normalised intensities for both the fundamental and the second harmonic modes along the propagation direction z for a grating length $\Lambda = 5.64 \mu\text{m}$ ($\Delta kL = 1.445252$). On the other hand, Fig. 7.6 shows the total normalised fields in the fundamental and second harmonic windows at different propagational distances using the parameters of Fig. 7.5. From these two figures, the power exchange mechanism between the two modes is clearly shown, where the effect of the QPM has been demonstrated through the coupling and the ripple of the intensities shown in Fig. 7.5 (bottom) with the grating periodicity. Fig. 7.6 shows that some power has coupled into radiation in the second harmonic window. The computed radiation power is very small compared with the power carried by the first guided modes in both windows. Also it has been observed that there is very little coupling to the third order mode of the second harmonic where this can be explained from the overlap integral between the first order mode at the fundamental frequency and the modes at the second harmonic frequency which is zero for the second order mode and very small for the third order mode [10,15] (see also S_0 below Eq. 7.9). In addition, the grating length used, in this simulation, results in non-phase-matching for these modes.

The other important case for QPM situation in nonlinear optics, apart from efficient SHG, is the study of nonlinear phase change to the fundamental field [5-7]. In the following analysis, the parallel SHG-EFD-BPM was used to compute the output depletion and the nonlinear phase change of the fundamental field after a distance of $L = 1 \text{ mm}$ for different grating lengths along the direction of propagation. By varying the grating length, the relative wave vector Δk changes while exciting predominantly the first guided mode of the second harmonic frequency. It has been assumed that the phase of the fundamental field is changing as $e^{-j(k_z^F z + \theta_{NL})}$ where θ_{NL} is the nonlinear phase shift of the fundamental field. Figs. 7.7 and 7.8 show the power exchange between the fundamental and the second harmonic and the nonlinear phase shift of the fundamental as a function of the propagation distance for $\Delta kL = \pi$ and 2π respectively.

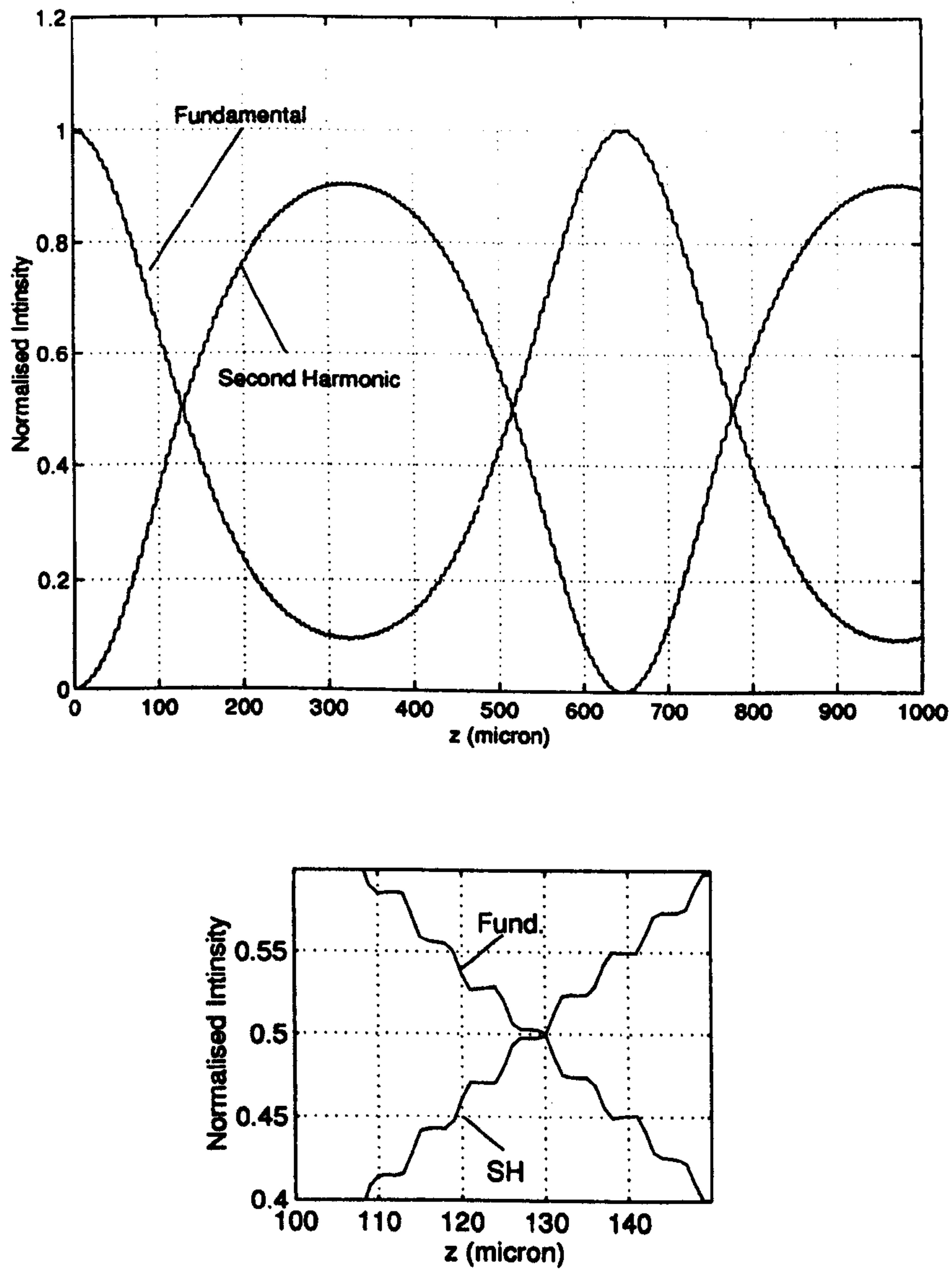


Fig. 7.5 (Top) The normalised intensity for both of the fundamental and the second harmonic fields versus the longitudinal distance $z(\mu\text{m})$, using the parallel SHG-EFD-BPM with an input power = 125.6 W and a grating length $\Lambda = 5.64 \mu\text{m}$. **(Bottom)** A portion from the top figure enlarged to show the ripple caused by the QPM.

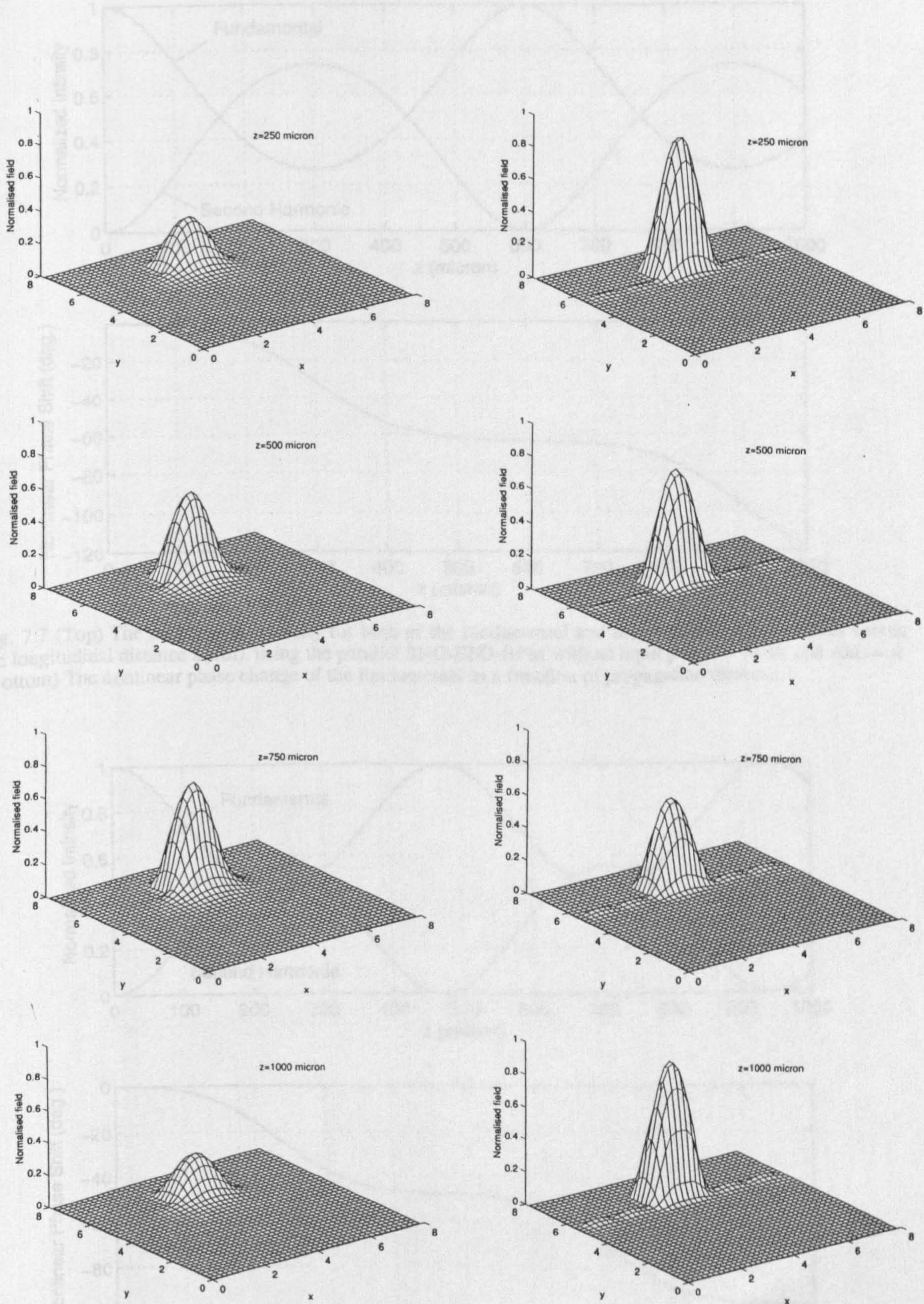


Fig. 7.6 The normalised total fields of the fundamental (left, top to bottom) and the second harmonic (right, top to bottom) at several propagational distances using the parallel SHG-EFD-BPM. All field points are normalised to 21.3×10^7 (see Fig. 7.5 for other details).

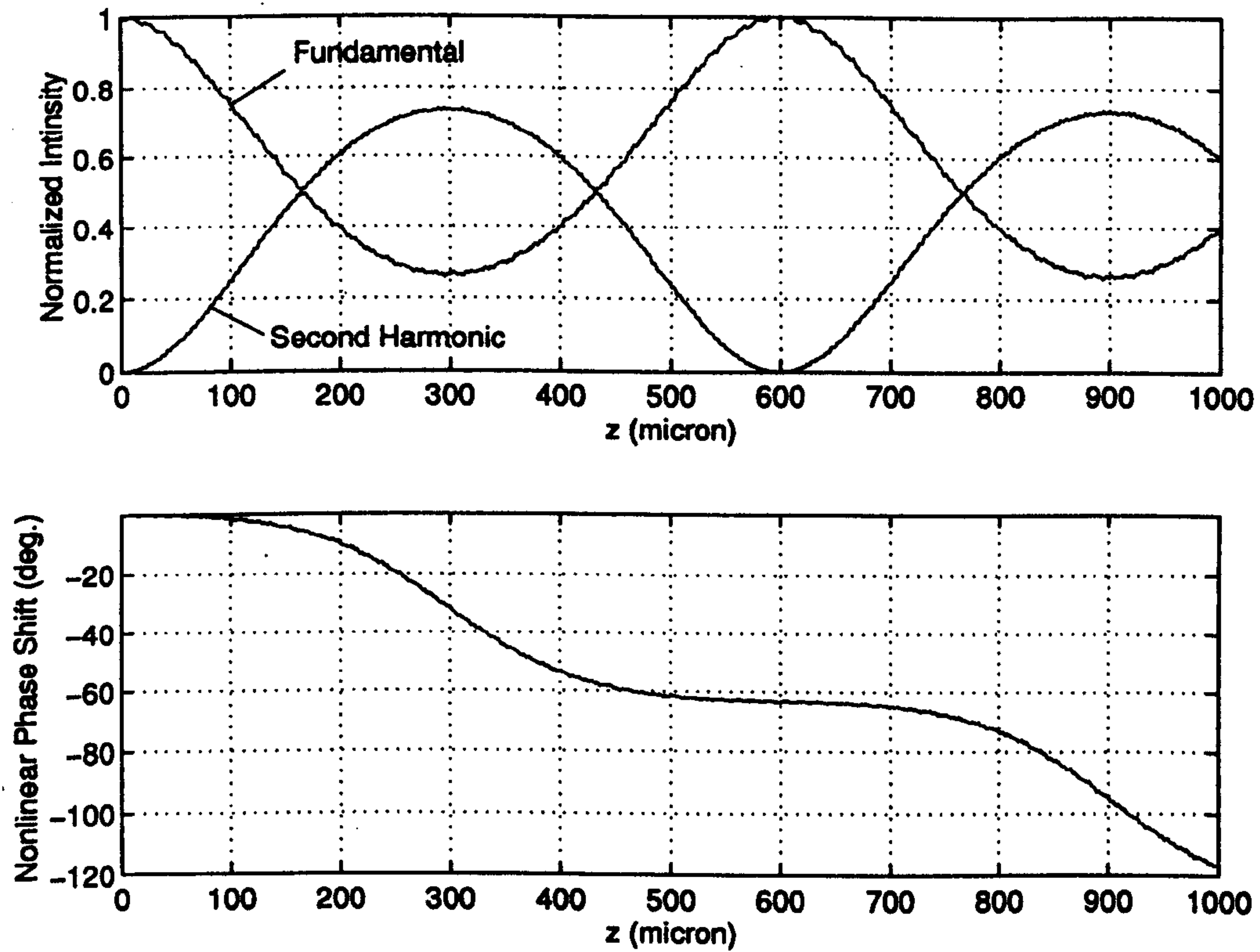


Fig. 7.7 (Top) The normalised intensity for both of the fundamental and the second harmonic fields versus the longitudinal distance $z(\mu\text{m})$, using the parallel SHG-EFD-BPM with an input power = 81 W and $\Delta kL = \pi$ (Bottom) The nonlinear phase change of the fundamental as a function of propagation distance.

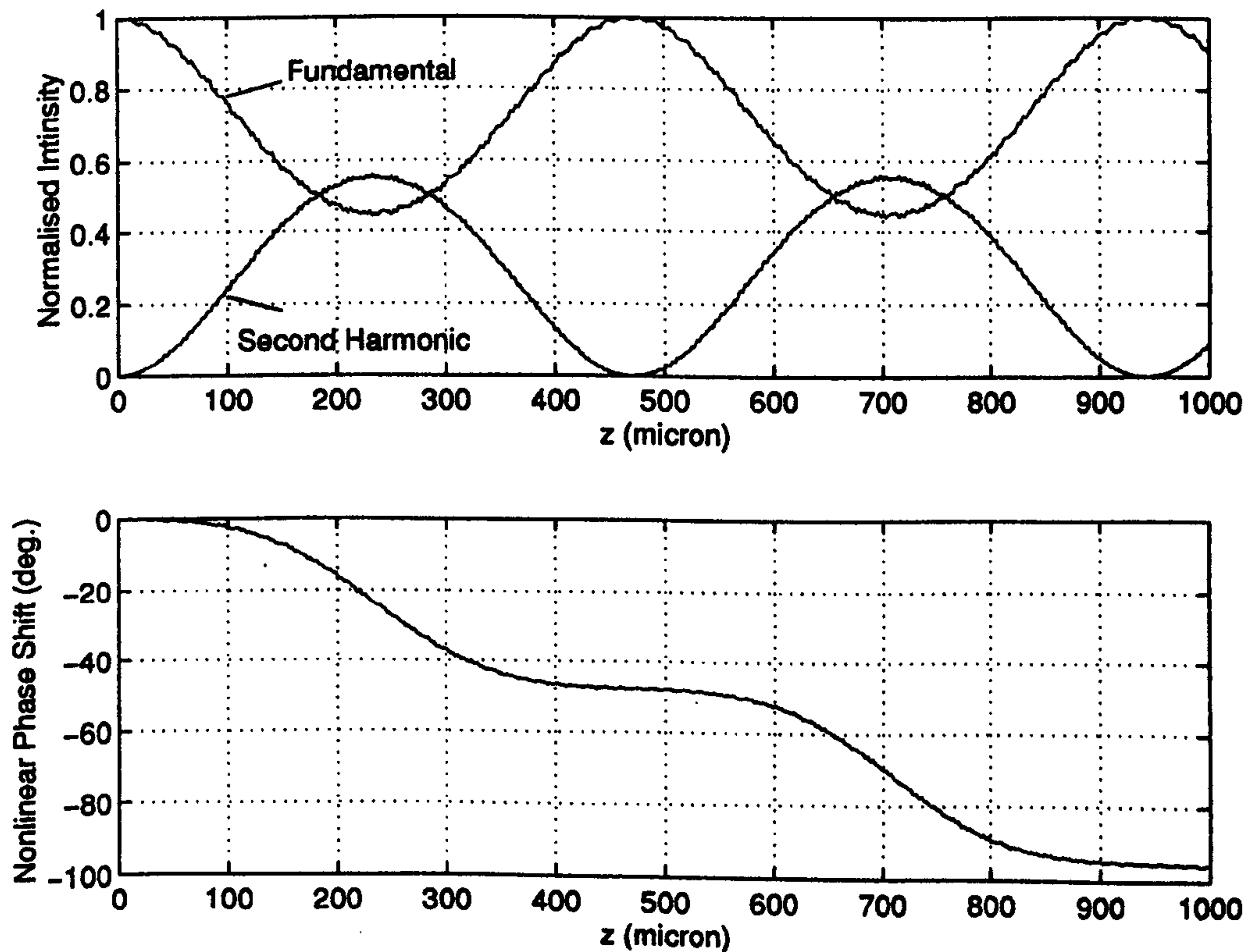


Fig. 7.8 The same as in Fig. 7.7 except $\Delta kL = 2\pi$.

In the following analysis, the depletion and the nonlinear phase shift of the fundamental after $L = 1$ mm will be computed as a function of the input power. In order to check the results of the parallel SHG-EFD-BPM, the solution of the approximate 1-D coupled wave equations (of Eq. 7.12) using the fourth order Runge-Kutta with the same parameters, was used. There are two methods to simulate QPM technique using the solution of Eq. 7.12. The first method is to use exactly the same procedure of the 3-D BPM with actual structural periodic gratings for the second order nonlinear coefficient $\chi^{(2)}$ along the z direction, and the second method is to reduce the value of $\chi^{(2)}$ by the QPM factor $1/\pi$ or $2/\pi$ (without structural grating), while changing explicitly $\Delta k'$ to $\Delta k' = 2k_o(n_{\text{eff}}^s - n_{\text{eff}}^f) - 2\pi/\Lambda$, in Eq. 7.12, to simulate the existence of second-order nonlinear gratings [2,10;14,15]. Both of these methods were tested numerically using the fourth order Runge-Kutta and an excellent agreement between the two results was observed. In the following simulation, the second method was used to compare with the SHG-EFD-BPM results. Fig. 7.9 shows the dependence of the depletion and the nonlinear phase shift θ_{NL} of the fundamental on the input power using the parallel SHG-EFD-BPM and the approximate 1-D Runge-Kutta for $\Delta kL = \pi$ and 2π ($\Delta k'L = \Delta kL$). From this figure we notice that the two methods agree very closely at low power with some little deviation at high power.

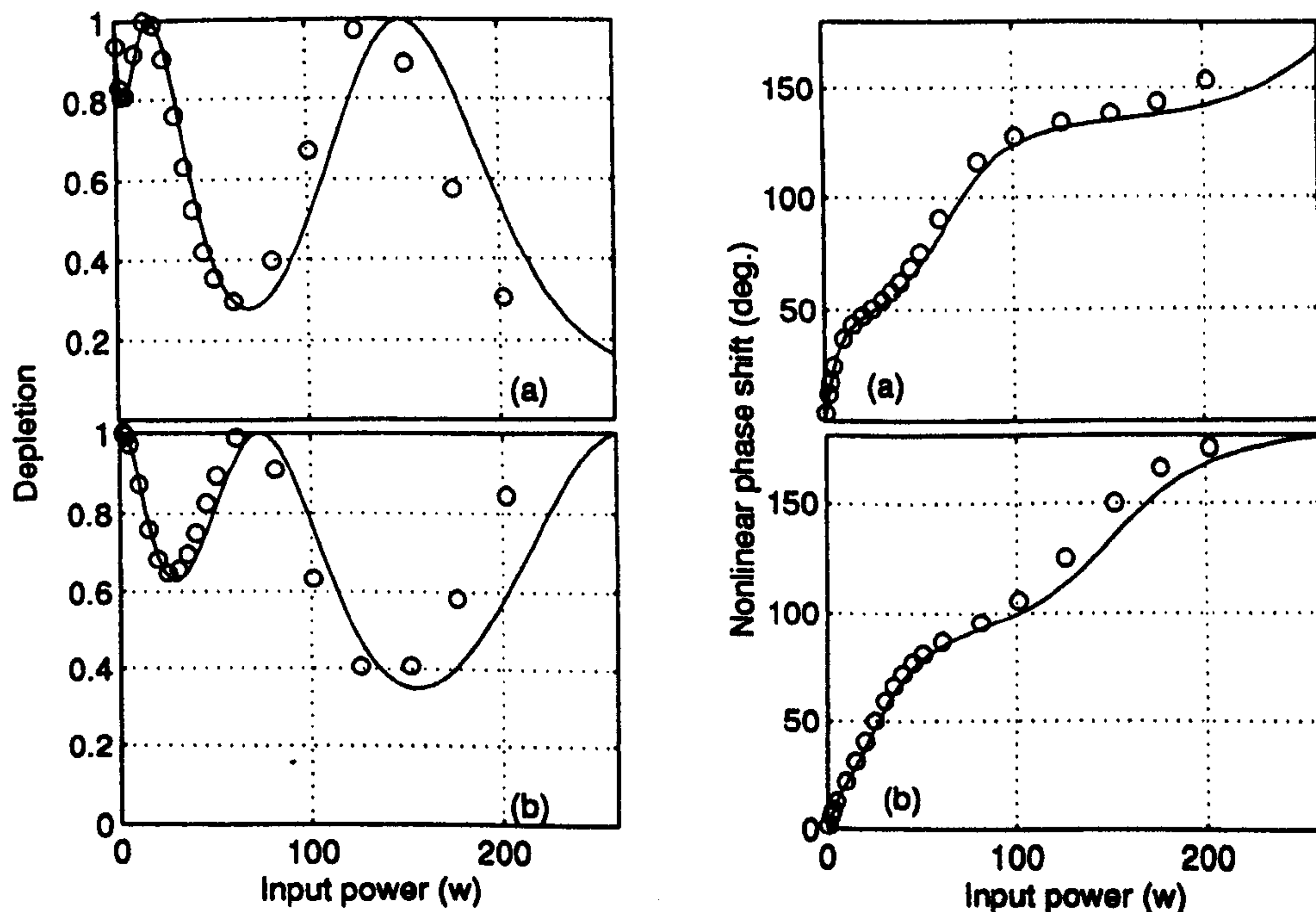


Fig. 7.9 The depletion intensity (left) and the nonlinear phase shift $|\theta_{\text{NL}}|$ (right) of the fundamental as functions of input power using the parallel SHG-EFD-BPM (circles) and the 1-D Runge-Kutta (solid line) for different values of ΔkL ($\Delta k'L = \Delta kL$); (a) $\Delta kL = \pi$, (b) $\Delta kL = 2\pi$.

Fig. 7.10 shows the depletion intensity and the nonlinear phase shift θ_{NL} of the fundamental as functions of ΔkL for both the parallel SHG-EFD-BPM and the approximate 1-D Runge-Kutta ($\Delta k'L = \Delta kL$) for different input powers. The figure shows close agreement between the two results. Also from the figure we can see that the nonlinear phase shift can be positive or negative, depending on the grating period, and this phase shift is increasing when the input power is increased. This effect has been studied theoretically [6], and has been observed experimentally in KTP [5]. The propagation step size Δz used in the last computations (Fig. 7.10) was $0.01\mu\text{m}$, although the SHG-EFD-BPM algorithm is stable for $\Delta z=0.025\mu\text{m}$, but in order to model the exact grating lengths mentioned before using a common step size, Δz had to be reduced.

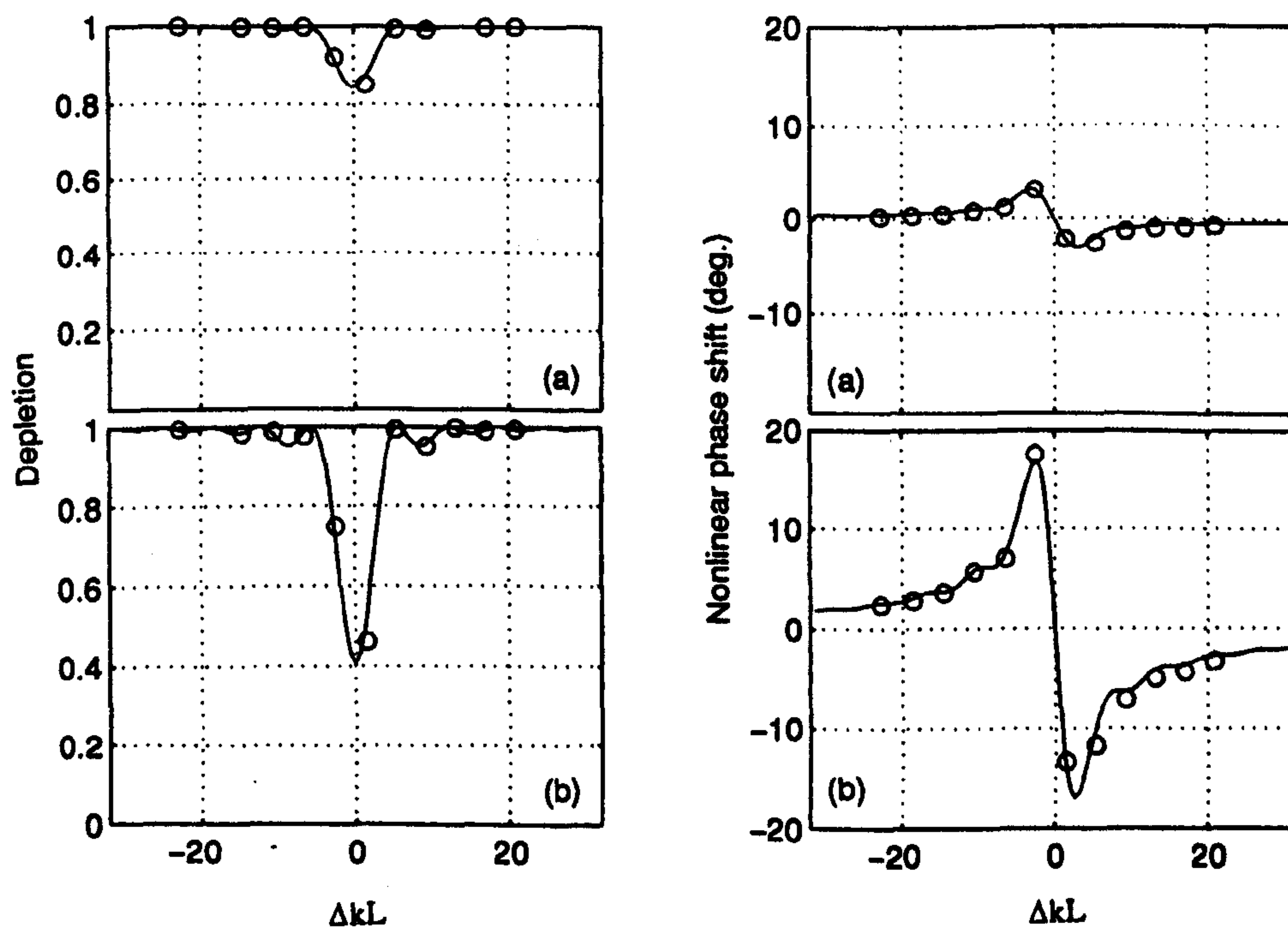


Fig. 7.10 The depletion intensity (left) and the nonlinear phase shift θ_{NL} (right) of the fundamental as functions of ΔkL ($\Delta k'L = \Delta kL$) using the parallel SHG-EFD-BPM (circles) and the 1-D Runge-Kutta (solid line) for (a) input power = 0.51 W and (b) input power = 3.02 W.

Finally, the total execution time for each run of the parallel SHG-EFD-BPM (of Fig. 7.10) is around 10.3 min. on the Connection Machine (CM-200) using 16k processors and around 76.3 min. using 64 processors of the transputer array [14-17]. Since the transputer implementation runs at around 54.2 % efficiency (see chapter 6) [11-13], the serial

execution time on one processor would be 2646.7 min., or 1.84 days. Therefore all the results in Fig. 7.10 consumed 4.12 hours on the CM and 30.52 hours on the 64 transputer array and will take 44.16 days using one processor of the transputer array.

7.5 Conclusion

A numerical method to model three-dimensional optical waveguides containing a second-order nonlinearity has been presented in this chapter. The formulation of the 3-D (SHG-EFD-BPM) and the simplified 1-D nonlinear coupled wave equations, to study second order nonlinear effect, have been shown in detail. The 1-D method was derived to validate some of the 3-D results. The new nonlinear method, which is an extension to the linear parallel EFD-BPM of chapter 6, is also highly parallel. In this chapter, the implementations of the SHG-EFD-BPM on parallel computers have been also shown. Then the parallel SHG-EFD-BPM was used to model semiconductor rib waveguides in the presence of the second order nonlinearity $\chi^{(2)}$ where the source field is allowed to deplete, using the two different supercomputers models (SIMD and MIMD). The Quasi Phase Matching (QPM) technique has been realistically incorporated in the analysis of the waveguide in order to change the relative phase mismatch between the two fields. Excellent agreement between the results of the parallel SHG-EFD-BPM and the solution of approximate 1-D, using the numerical Runge-Kutta method, has been observed. Finally, it has been concluded that the new parallel algorithm is simple, efficient and very useful for modelling large complicated nonlinear optical devices.

REFERENCES

- [1] Yariv, A. and Yeh, P., "Optical Waves in Crystals", John Wiley & Sons, 1984.
- [2] Fejer, M. M., Magel, G. A. Jundt, D. H. and Byer, R. L., "Quasi-Phase Matched Second Harmonic Generation: Tuning and Tolerances", *IEEE J.Q.E.*, **28**, 2631-2654, 1992.
- [3] Rustagi, K. C., Mehendale, S. C. and Meenakshi, S., "Optical Frequency Conversion in Quasi-Phase-Matched Stacks of Nonlinear Crystals", *IEEE J. Q.E.*, **18**, 1029-1041, June 1982.
- [4] Yamamoto, K., Mizuuchi, K. and Taniuchi, T., "Milliwatt-Order Blue-Light Generation in a Periodically Domain-Inverted LiTaO₃ Waveguide", *Opt. Lett.*, **16**, 1156-1158, Aug. 1991.

- [5] DeSalvo, R., Hagan, D. J., Sheikh-Bahae, M., Stegeman, G. and Van Stryland, E. W., "Self-Focusing and Self-Defocusing by Cascaded Second-Order Effects in KTP", *Opt. Lett.*, **17**, 28-30, 1992.
- [6] Ironside, C. N., Aitchison, J. S. and Arnold, J. M., "An All-Optical Switch Employing the Cascaded Second-Order Nonlinearity Effect", *IEEE J.Q.E.*, **29**, 2650-2654, 1993
- [7] Hutchings, D. C., Aitchison, J. S. and Ironside, C. N., "All-Optical Switching Based on Nondegenerate Phase Shifts From a Cascaded Second-Order Nonlinearity", *Opt. Lett.*, **18**, 793-795, May 1993.
- [8] Krijnen, G. J., Hoekstra, H. J. and Lambeck, P. V., "BPM Simulations of Integrated Optic Structure Containing Second Order Nonlinearity", *ECIO (Euro. Conf. Intg. Opt.)*, 5-4 to 5-5, 1993.
- [9] Hermansson, V. and Yevick, D., "A Propagation Beam Method Analysis of Nonlinear Effects in Optical Waveguides", *Opt. Quant. Elect.*, **16**, 525-534, 1984.
- [10] Weitzman, P. S. and Osterberg, U., "A Modified Beam Propagation Method to Model Second Harmonic Generation in Optical Fibers", *IEEE J.Q.E.*, **29**, 1437-1443, 1993.
- [11] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Methods", *IEEE Phot. Tech. Lett.*, **6**, 848-850, July 1994.
- [12] Masoudi, H. M. and Arnold, J. M., "Parallel Three-Dimensional Finite-Difference Beam Propagation Methods", *Int. J. of Num. Mod.*, **8**, 95-107, 1995.
- [13] Masoudi, H. M. and Arnold, J. M., "Parallel-Processing Finite-Difference Beam Propagation Methods", Presented at the *International Symposium on Guided-Wave Optoelectronics: Device Characterization, Analysis and Design*, IEEE and Optical Society of America in Polytechnic University, New York, October 1994.
- [14] Masoudi, H. M. and Arnold, J. M., "Parallel Beam Propagation Method for the Analysis of Second Harmonic Generation", *IEEE Phot. Tech. Lett.*, **7**, 400-402, April 1994.
- [15] Masoudi, H. M. and Arnold, J. M., "Modelling Second-Order Nonlinear Effects in Optical Waveguides Using a Parallel-Processing Beam Propagation Method", submitted to the *IEEE Journal of Quantum Electronics*, October 1994.
- [16] Masoudi, H. M. and Arnold, J. M., "Three-Dimensional Numerical Modelling of the Second-Order Cascaded Effect in a Semiconductor Rib Waveguide Using Parallel-Processing", presented at the *Optical Society of America Topical meeting: Nonlinear Guided Waves and Their Applications*, Dana Point, California, February 1995.

- [17] Masoudi, H. M. and Arnold, J. M., "Numerical Modelling of Second Harmonic Generation in an Optical Waveguide Using Parallel-Processing Finite-Difference Beam Propagation Method", presented at the *URSI Electromagnetic Theory Symposium*, St. Petersburg, Russia, May 1995.
- [18] Chung, Y. and Dagli, N., "Analysis of Z-Invariant and Z-Variant Semiconductor Rib Waveguides by Explicit Finite Difference Beam Propagation Method With Nonuniform Mesh Configuration", *IEEE J.Q.E.*, **27**, 2296-2305, 1991.
- [19] Press, W. H., Teukolsky, S. A. Vetterling, W. T. and Flannery, B. P., "The Art of Scientific Computing: Numerical Recipes in FORTRAN", Cambridge Uni. Press, 1992.
- [20] Kelaidis, C. M., Hutchings, D. C. and Arnold, J. M., "Asymmetric Two-Step GaAlAs Quantum Well for Cascaded Second-Order Processes", Accepted for publication in the *IEEE J. Q. E.*, 1994.

Chapter 8

Summary and Conclusion

In this work the theory of modelling three-dimensional optical devices has been discussed. For z-invariant rectangular waveguides, several numerical techniques can be employed to find the characteristics of the waveguides, ranging from simple and fast to accurate and time consuming. A review of some of these methods has been given in this thesis. On the other hand, modelling 3-D z-variant rectangular waveguides is more difficult than the z-invariant waveguides due to the involvement of radiation modes which are complicated to predict accurately. The Beam Propagation Method (BPM) proved to be an excellent numerical technique to analyse a wide range of optical devices, specially those of longitudinal variations. A review on the latest developments of this method was discussed and the formulation for the most common three BPM techniques, based on the Finite-Difference method (FD), were also shown in detail. The three methods are the Alternating Implicit Direction (ADI), the Real Space (RS) and the Explicit Finite-Difference (EFD) BPMs; the first is an implicit method and the last two are explicit methods. The FD-BPMs proved to be much more efficient than the classical FFT-BPM. As a first step toward implementing the 3-D FD-BPMs, the 2-D versions of the FD-BPMs were first implemented and tested using a normal serial computer to analyse slab geometries. Comparisons between the BPM results and the analytical solution of the slab waveguide have been made. The accuracy analysis of these implementations showed that the accuracies of all the three FD-BPMs improve by reducing the mesh spacing. The comparison between the speed per propagational step of the three FD-BPMs showed that the EFD-BPM is faster than the other two methods and the CN-BPM is faster than the RS-BPM. It has been concluded in the analysis of the 2-D BPM that both the EFD-BPM and the CN-BPM are more efficient than the RS-BPM since the latter converges using a smaller longitudinal step size than the first two, and also the speed of the RS-BPM is slower per propagational step. On the other hand, comparison between the EFD-BPM and the CN-BPM showed that the EFD-BPM is more efficient than the CN-BPM at large mesh spacing, since both algorithms produced similar results using the same longitudinal step size and the EFD-BPM is faster, per propagational step, than the CN-BPM. However as

mesh spacing decreases, the EFD-BPM requires smaller longitudinal step size for stability, which might indicate that the CN-BPM is more efficient for very small mesh spacing. A new technique to the EFD-BPM, to improve the stability limit of the propagational step of the method, has been suggested which requires a slight adjustment to the EFD-BPM numerical equation. This modification has improved the stability of the EFD-BPM, but it also introduced a spurious field to the true solution. Then for 2-D analysis the CN-BPM was used to model practical 2-D devices, where it has been applied to simulate the propagation of guided modes in directional couplers and Y-junctions.

On the other hand, analysing three-dimensional linear optical devices using the BPM requires much more computational effort than the 2-D counterparts. For this reason parallel processing computers have been used to speed up the execution of the BPM. First a review of the field of parallel computation was given, concentrating on the most common parallel machines; namely the transputer array (MIMD) and the Connection Machine (SIMD). Also this thesis discussed the difficulties in developing numerical algorithms for parallel computers. In order to gain maximum efficiency out of parallel implementations, care must be exercised in selecting the appropriate method for parallel implementations. Second, the thesis discussed in detail the implementations of the three-dimensional FD-BPMs, to analyse linear optical devices, on the transputer array and the Connection Machine. It has been demonstrated that a large speed up of the execution of the finite-difference explicit versions of the BPM (the RS-BPM and the EFD-BPM) on parallel computers in comparison to the serial execution can be easily achieved. These methods are very well suited to the parallel environment because they inherit the locality of spatial points, which reduces the communication overhead between parallel processors. On the other hand, the computation of the field at any given spatial point using the implicit ADI-BPM requires information from all parts of the problem, due to the necessity of inverting large matrices, which is very expensive in terms of parallel computing. 90-100 percent efficiency has been observed by implementing the EFD-BPM and the RS-BPM on the transputer array. In addition the implementations of these methods on the Connection Machine have produced even faster parallel computer codes compared to the best performance of the transputer array implementations. All of the parallel-processing performance indicators improve with increasing number of grid points. The comparisons between the two parallel explicit methods showed that the EFD-BPM is several times faster than the RS-BPM per propagational step. In terms of accuracy, the two methods have been used to analyse three-dimensional rib waveguides and directional couplers, and the results have been compared with other serial techniques. It has been concluded that the EFD-BPM is more efficient than the RS-BPM since the latter converges at a similar step size to that of the EFD-BPM.

In the last part of this thesis three-dimensional nonlinear optical waveguides, containing second-order nonlinearity $\chi^{(2)}$, have been analysed using a new numerical technique. The new nonlinear method, which is an extension to the linear parallel EFD-BPM, is also highly parallel. The formulation of the 3-D (SHG-EFD-BPM) coupled wave equations, to study second order nonlinear effect, have been shown in detail. The implementation of the SHG-EFD-BPM on parallel computers was also discussed, where two separate computational windows have been used, one for the fundamental field and the other for the second harmonic field. Then the parallel SHG-EFD-BPM was used to model semiconductor rib waveguides in the presence of the second order nonlinearity $\chi^{(2)}$ where the source field is allowed to deplete, using the two different supercomputers models (SIMD and MIMD). In these analyses, detailed information for the nonlinear phase shift of the fundamental field has been studied. In the simulation, the Quasi Phase Matching (QPM) technique was used in the analysis to change the relative phase mismatch between the two fields. In order to validate some of the 3-D results, the familiar simplified 1-D nonlinear coupled wave equations have been formulated from the 3-D models and have been implemented numerically using the fourth-order Runge-Kutta method. Excellent agreement between the results of the parallel SHG-EFD-BPM and the solution of the approximate 1-D was observed. Finally, the output of this thesis is very important for modelling large complicated linear and nonlinear optical devices due to the simplicity and the high efficiency of the methods used. In addition, this work is also very useful for other mathematical applications where the parabolic equation is used.

Future Work

Many developments to the existing work of this thesis could be studied further, and here are some of these:

1- As new faster parallel computers are emerging, the implementations in this thesis become increasingly useful. For example, the implementations of the EFD-BPM (linear and nonlinear) on the new Cray-T3D* machine can be easily applied with little modifications to the existing computer codes**. This machine consists of 320 processors (MIMD type), with a peak performance of 40 GFLOPS where each processor has 64 Mbytes RAM with a speed of 150 MHz. Data can be transferred between processors with a rate of 100 Mbytes/s.

* This machine has just been installed in the Edinburgh Parallel Processing Centre (EPCC), Edinburgh, UK.

** This work is currently under development.

2- The work in this thesis could be easily used to investigate many new current research problems:

- a) Three-dimensional study of segmented linear rib waveguides.
- b) Three-dimensional study of second-order nonlinearity $\chi^{(2)}$ using coupling between the guided fundamental field with Cerenkov radiation at the second harmonic frequency [1-2]**.
- c) Three-dimensional study of spatial solitons using the second order nonlinearity $\chi^{(2)}$ [3-4]**.
- d) Three-dimensional study of directional couplers containing $\chi^{(2)}$ **.

3- The extension of the three-dimensional EFD-BPM to study nonlinear materials containing $\chi^{(3)}$.

4- The extension of the three-dimensional EFD-BPM to vector fields to study the effect of polarisation dependence both for linear and nonlinear cases.

5- Investigation of Boundary Conditions at the edges of the computational window of the BPM.

6- Extension for the EFD-BPM to the fourth order finite-difference approximation which enables one to use larger propagational step size.

7- Investigation of the modified EFD-BPM (MEFD-BPM), discussed in chapter 5, to improve the stability of the EFD-BPM.

8- Extension for the EFD-BPM to a wide-angle BPM to overcome the paraxial limit.

REFERENCES

- [1] Chikuma, K. and Umegaki, S., "Theory of Optical Second-Harmonic Generation in Crystal-Cored Fibers Based on Phase Matching of Cerenkov-Type Radiation", *J. Opt. Soc. Am. B*, **9**, 1083-1092, 1992.
- [2] Krijnen, G. J. M., Torruellas, W., Stegeman, G. I., Hoekstra, H. J. W. and Lambeck, P. V., "Nonlinear Phase-Shifts by Cascading in the Cerenkov Regime", Presented at the International Symposium on Guided-Wave Optoelectronics: Device Characterization, Analysis and Design, *IEEE and Optical Society of America* in Polytechnic University, New York, October 1994.
- [3] Buryak, A. V. and Kivshar, Y. S., "Spatial Optical Solitons Governed by Quadratic Nonlinearity", *Opt. Lett.*, **19**, 1-3, 1994.
- [4] Torner, L. Menyuk, C. R., Torruellas, W. E. and Stegeman, G. I., "Two-Dimensional Solitons with Second-Order Nonlinearities", *Opt. Lett.*, **20**, 13-15, 1995.