UNI- AND OMNIDIRECTIONAL SIMULATION TOOLS FOR INTEGRATED OPTICS

The research described in this thesis was carried out at the Lightwave Devices Group of the MESA+ Research Institute, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands

This work was financially supported by the Dutch Technology Foundation STW, under grant number TEL 55.3740

Cover:

A 3-D representation of the scattering of a Gaussian beam on a dielectric cylinder, as calculated by the Finite Difference Time Domain method.

ISBN 90-365-1576-9

UNI- AND OMNIDIRECTIONAL SIMULATION TOOLS FOR INTEGRATED OPTICS

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. F.A. van Vught,
volgens besluit van het College voor Promoties
in het openbaar te verdedigen
op donderdag 3 mei 2001 te 15:00 uur.

door

Remco Stoffer

geboren op 28 augustus 1973
te Apeldoorn

Dit proefschrift is goedgekeurd door:

de promotor: Prof. Dr. E. van Groesen
de assistent-promotor: Dr. Ir. H.J.W.M. Hoekstra

Table of Contents

Chapter 1: Introduction

1.1 Integrated optics

Since the demand for data traffic started to rise almost exponentially, the demand for components that can deal with high data rates has also grown at the same rate. Conventional electronics are not capable of operating at the very high frequencies that are needed to process the data; also, conventional copper wires cannot operate at frequencies above some 10 GHz. Therefore, optical fiber technology is becoming the standard for long-distance data communications. An optical fiber is capable of transmitting light having a range of wavelengths (~1530-1570 nm) with little attenuation across large distances. If one were able to utilize this complete wavelength range for the modulation of a signal, the data rate that would be obtained is more than 4 Tb/s. There are several reasons why this cannot be reached directly: One, the electronics that have to modulate the signal cannot operate at this speed; and two, the optical fiber has some amount of wavelength dispersion, which means that short pulses get deformed. The solution to this is to split the wavelength range into independent channels, each of which is modulated by its own electronic modulator. In this way, one may transmit, for example, 64 different channels of information through a single fiber, each of which contains information at a data rate of 10 GHz or more [Koch 1997, Onaka 1996].

At the ends of the optical fiber, the light must be manipulated in order to be able to get access to the different channels. The signal from each source, the lasers and modulators for each wavelength channel, must be combined and inserted into the fiber (multiplexing). On the exit of the fiber, one must be able to split all channels apart (demultiplexing), or to extract one channel and insert new information onto that channel (add-drop filtering). Also, the signal may have to be regenerated (re-amplified, reshaped and retimed). While some of these functions, e.g. amplification, can be done best using bulk optics or fiber-optics, others can be done more efficiently and cheaply by creating small optical chips, able to perform the required functions. This is where Integrated Optics come into play [Tamir 1979].

The term ‘Integrated Optics’ implies that light is manipulated on a small scale, with multiple functions integrated on a single chip. This is usually done by depositing layers of transparent material on a flat substrate like a silicon wafer. These layers are then patterned in order to guide and manipulate the light. The behaviour of these structures can be made wavelength-dependent in order to have wavelength selective functionality.

Telecommunication, however, is not the only application for integrated optics. The layered systems can also be used to make sensors for e.g. relative humidity or the concentration of heavy metal ions in a liquid. A way to do this is by making use of the fact that the wavelength of the light is very small (~10^{-6} m), and therefore, small changes in the speeds at which the light propagates through the structure lead to large phase changes after a short distance (several millimeters or centimeters) of propagation. Changes of the index of refraction of 10^{-8} can be measured in this manner [Heideman et al 1999].

In order to design devices that perform the above-mentioned functions, it is necessary to know how light behaves in these small structures. Since the dimensions are on the order of the wavelength, the wave-like character of the light plays a very important role. For very simple devices, analytical tools are able to predict the behaviour of the light in the structures very well. However, as soon as the complexity increases, one needs to do numerical simulations of the light in the structure. These numerical tools should be as accurate and fast as possible, in order to accelerate the design process of integrated optical components.

1.2 Two-dimensional simulation tools

The structures that must be simulated are 3-dimensional. At the moment of writing, it is not yet feasible to perform complete three-dimensional simulations for most realistic structures, due to the amount of time it would take to do these calculations. Fortunately, in many cases, it is possible to separate one of the spatial variables out of the equation without introducing large errors, by using the so-called effective index method. This effectively reduces the system to a 2-dimensional one. In most applications, these calculations are not completely correct; however, in many cases, trends can be reproduced accurately. This reduction to two dimensions also has the effect of decoupling the two polarizations for isotropic materials; the polarization is either TE (Transverse Electric, with the E-field perpendicular to the calculation plane) or TM (Transverse Magnetic, with the H-field perpendicular to the calculation plane). Of course, this decoupling means that phenomena like polarization conversion cannot be directly modeled.

Even when the component under study cannot be completely reduced to a 2-dimensional one, parts of it can usually be modeled in two dimensions, which may yield enough information to make an efficient design for the component by using e.g. scattering matrices to connect the parts. Also, it may be that interesting phenomena are expected to occur in such a device. One may then first simulate a 2-dimensional structure that exhibits the same or similar phenomena, in order to gain insight in the how and why of these phenomena.

This thesis deals purely with 2-dimensional simulation methods for linear integrated optics. The fields and refractive index only vary in the x- and z-directions, and are assumed to be constant in the y-direction. Although the presented methods are not limited to that, the refractive index is assumed to have a piecewise-constant profile; i.e. the interfaces between materials are abrupt.

In many integrated optical devices, the light travels mainly in one single direction; it enters the component from, say, the left-hand side and travels to the right-hand side. An example of these devices is a Mach-Zehnder interferometer, as used in sensors or add-drop filters. In the simulations, one can make use of this fact, and transform the equations that are to be solved from a boundary value problem into an initial value problem, which means that one can calculate from left to right. Simulation tools that use this approximation are referred to as unidirectional simulation tools; also, the phrase Beam Propagation Method (BPM) is often used to indicate such a method.

Contrarily, there are other types of devices in which the propagation of the light is essentially non-unidirectional; light may travel around a tight bend and propagate perpendicularly to the original direction, or be reflected back in the direction it originally came from. Examples of these structures are cylindrical microcavities and photonic bandgap crystals. For this type of devices, the simulation tools must be able to deal with light traveling in all directions in the plane; these are referred to as omnidirectional simulation tools. One can do this for one single frequency, by solving the Helmholtz equation, or take into account the time dependence of the fields, and thus simulate the full time-dependent Maxwell equations, e.g. using a Finite Difference Time Domain (FDTD) method.

In these simulation methods, two of the most important issues are how one deals with the boundaries of the calculation window, and with interfaces between materials. The first problem is caused by the fact that one cannot simulate an infinite size; the calculations are necessarily performed on a window of finite size. The boundaries of this window should behave as if they were not present; light impinging on the boundaries should be transmitted or absorbed, while not producing any reflection. However, it should simultaneously be possible to define a wave that travels from outside the calculation window into the system. For the unidirectional methods, this is not a problem; since no light travels from right to left, the left-hand boundary does not need to be nonreflecting and any fields that are defined there are automatically incoming fields. However, in omnidirectional methods, the boundary from which the incoming field starts should also be nonreflecting, since light may be scattered from the device toward this boundary.

Due to interfaces between materials with a different index of refraction, the fields at those interfaces are not smooth; some of the vector components of the *E*- and *H*-fields are discontinuous, or have jumps in the first, second or higher order normal derivatives across the interface. This fact must be taken into account in a good way for accurate, rapidly converging simulation methods.

1.3 Maxwell and Helmholtz equations

Since the main part of this thesis deals with methods to solve Maxwell's equations, they are presented here. In linear, isotropic, nonconducting materials without sources, they are:

$$\begin{aligned} \frac{\partial}{\partial t}\varepsilon \vec{E} &= \nabla \times \vec{H} \\ \frac{\partial}{\partial t}\mu \vec{H} &= -\nabla \times \vec{E} \\ \nabla \cdot \left(\varepsilon \vec{E}\right) &= 0 \\ \nabla \cdot \left(\mu \vec{H}\right) &= 0 \end{aligned} \tag{1.1}$$

For timeharmonic solutions with time dependence $\exp(i\omega t)$, they reduce to the Helmholtz equations for TE and TM:

$$TE: \quad \left(\P_{zz} + \P_{xx} + k_0^2 n^2\right) E_y = 0$$
$$TM: \quad \left(\P_z \frac{1}{n^2} \P_z + \P_x \frac{1}{n^2} \P_x + k_0^2\right) H_y = 0 \tag{1.2}$$

1.4 Outline of the thesis

This thesis describes several new methods, or improvements upon older methods, for doing 2-dimensional simulations in integrated optics. The outline is as follows:

- The current chapter gives an introduction to the subject of the thesis.
- In all simulation methods, boundary conditions are a very important issue. Chapter 2 describes some of these boundary conditions, including the Perfectly Matched Layer (PML), which is used in all of the following chapters.
- In Chapter 3, improvements of unidirectional Beam Propagation Methods are presented; the order of convergence is improved by using a higher-order discretization and taking into account the interface conditions. For TE, this is done in two ways: One, by using Taylor expansions around the interface to calculate corrections; and two, by applying a variational approach to the problem. For TM, only the latter approach is used.
- For the omnidirectional problem, a Helmholtz solver with PML's surrounding the whole domain is presented in Chapter 4. Again, a variational approach is used to discretize the equations, and interfaces through grid cells are taken into account. The incoming field is generated on the outside of the PML. The propagation constant of the various modes that should be launched is taken into account. Also, the chapter describes a similar way of launching light into the window for the FDTD method, including modal pulse generation, in which a temporal pulse is generated whose transverse profile corresponds to a mode of the incoming waveguide corresponding to each frequency.
- Chapter 5 presents results of the application of the methods described in Chapter 4 to two kinds of structures: A cylindrical microcavity and photonic crystal structures.
- Finally, Chapter 6 presents a summary of the results of the thesis.

Chapter 2: Nonreflecting boundary conditions

For computations in optical structures the fields are often approximated by a discretization of their values in a necessarily finite window. Light scattering from structures inside the window will eventually reach the boundary of the window, where it should be transmitted (or absorbed) and not reflected back into the window, since reflected light will get back toward the structures and disturb the simulations. So, the boundary should behave as if it were not present; the solution in the interior plus the boundary conditions should be equal to a restriction of the solution on the whole plane to the window.

This chapter will describe two different techniques of 'windowing' for finite difference or finite element methods, to be used in later chapters. The first is based on adding an absorbing layer around the window with special properties, the so-called Perfectly Matched Layer (PML) [Bérenger 1994, Yevick at al 1997]; the second is a (possibly adaptive) Sommerfeld-like boundary condition [Hadley 1991, Moore et al 1988, Vasallo 1996, 1997, Pregla 1995]. Other methods like a Green's function method [Song 1993] are also possible, but seem to be less efficient, and will not be discussed here.

2.1 Absorbing boundary layers

We will investigate how light in absorbing media behaves, and how the medium and the equations should be chosen in order to have reflectionless absorption; this will yield the well-known Perfectly Matched Layers. For these PML's, both electrical and magnetic conductivities will be necessary [Bérenger 1994].

Light propagating in a uniform, isotropic conducting medium decreases in amplitude along the propagation direction. However, in order to have a good absorption of modal fields propagating into the medium, the transverse shape of the mode should be retained. This means that the absorbing medium should only attenuate the light in the direction perpendicular to the boundary; for an index distribution that only varies in the direction parallel to the boundary, this means that the eigenvectors (modal shape) should remain the same, while the eigenvalue (propagation constant) should become complex. If the modal shape in the absorbing medium were different from the one in the in the non-absorbing medium, which is the case when light is also attenuated in the direction parallel to the boundary, there would be reflection on the interface between the two media due to the modal mismatch.

The first two Maxwell's equations with both electrical and magnetic conductivity are:

$$\begin{aligned} \frac{\partial}{\partial t} \varepsilon \vec{E} &= \nabla \times \vec{H} - \sigma_E \vec{E} \\ \frac{\partial}{\partial t} \mu \vec{H} &= -\nabla \times \vec{E} - \sigma_H \vec{H} \end{aligned} \tag{2.1}$$

where s_E and s_H are the electrical respectively magnetic conductivities in the material. In this form, a wave propagating in a conducting medium is attenuated, no matter its propagation direction. In the absorbing boundary layer, light should only be attenuated in a direction normal to the boundary. It turns out that in order to have a directional absorption (i.e. absorbing only light that propagates with a component of the wave vector in the direction perpendicular to the boundary), it is necessary to split each variable into two parts and change to a somewhat different version of (2.1). One way is to go toward complex-valued variables and use tensors with complex diagonal elements for both μ and ε. Another way, which was pioneered by J.P. Bérenger [Bérenger 1994] and will be described here, is to split each variable into two parts and changing the equations; this has the advantage that fewer operations are needed per point than when one goes to a complex-variable approach. The upper equation of (2.1) is changed to:

$$\begin{cases} \dfrac{\partial}{\partial t}\varepsilon E_{xy} = \dfrac{\partial\left(H_{zx}+H_{zy}\right)}{\partial y} - s_{E,y}E_{xy} \\ \dfrac{\partial}{\partial t}\varepsilon E_{xz} = -\dfrac{\partial\left(H_{yx}+H_{yz}\right)}{\partial z} - s_{E,z}E_{xz} \end{cases}$$
$$\begin{cases} \dfrac{\partial}{\partial t}\varepsilon E_{yx} = -\dfrac{\partial\left(H_{zx}+H_{zy}\right)}{\partial x} - s_{E,x}E_{yx} \\ \dfrac{\partial}{\partial t}\varepsilon E_{yz} = \dfrac{\partial\left(H_{xy}+H_{xz}\right)}{\partial z} - s_{E,z}E_{yz} \end{cases} \qquad (2.2)$$
$$\begin{cases} \dfrac{\partial}{\partial t}\varepsilon E_{zx} = \dfrac{\partial\left(H_{yx}+H_{yz}\right)}{\partial x} - s_{E,x}E_{zx} \\ \dfrac{\partial}{\partial t}\varepsilon E_{zy} = -\dfrac{\partial\left(H_{xy}+H_{xz}\right)}{\partial y} - s_{E,y}E_{zy} \end{cases}$$

while the lower equation of (2.1) becomes:

$$
\begin{cases}
\dfrac{\partial}{\partial t}\mu H_{xy} = -\dfrac{\partial\left(E_{zx}+E_{zy}\right)}{\partial y} - \sigma_{H,y} H_{xy} \\
\dfrac{\partial}{\partial t}\mu H_{xz} = \dfrac{\partial\left(E_{yx}+E_{yz}\right)}{\partial z} - \sigma_{H,z} H_{xz}
\end{cases}
\begin{cases}
\dfrac{\partial}{\partial t}\mu H_{yx} = \dfrac{\partial\left(E_{zx}+E_{zy}\right)}{\partial x} - \sigma_{H,x} H_{yx} \\
\dfrac{\partial}{\partial t}\mu H_{yz} = -\dfrac{\partial\left(E_{xy}+E_{xz}\right)}{\partial z} - \sigma_{H,z} H_{yz}
\end{cases}
\begin{cases}
\dfrac{\partial}{\partial t}\mu H_{zx} = -\dfrac{\partial\left(E_{yx}+E_{yz}\right)}{\partial x} - \sigma_{H,x} H_{zx} \\
\dfrac{\partial}{\partial t}\mu H_{zy} = \dfrac{\partial\left(E_{xy}+E_{xz}\right)}{\partial y} - \sigma_{E,y} H_{zy}
\end{cases}
\tag{2.3}
$$

in which $\sigma_{E,x}$ only depends on x, $\sigma_{E,y}$ only on y and $\sigma_{E,z}$ only on z, and similar for $\sigma_{H,*}$.

The actual E and H fields are obtained from the quantities in equations (2.2) and (2.3) as follows:

$$
\begin{aligned}
E_x &= E_{xy} + E_{xz} \\
E_y &= E_{yx} + E_{yz} \\
E_z &= E_{zx} + E_{zy} \\
H_x &= H_{xy} + H_{xz} \\
H_y &= H_{yx} + H_{yz} \\
H_z &= H_{zx} + H_{zy}
\end{aligned}
\tag{2.4}
$$

First, a one-dimensional system will be considered; afterward, the 2-dimensional case is studied, in which it will be made clear why this splitting of the variables gives the intended directional behaviour of the attenuation.

2.1.1 One-dimensional absorbing layers

In a one-dimensional system, i.e. in a system where nothing depends on y or z, equations (2.2) and (2.3), combined with (2.4), reduce to:

$$
\begin{aligned}
\frac{\partial}{\partial t}\varepsilon E_z &= \frac{\partial}{\partial x} H_y - \sigma_{E,x} E_z \\
\frac{\partial}{\partial t}\mu H_y &= \frac{\partial}{\partial x} E_z - \sigma_{H,x} H_y
\end{aligned}
\tag{2.5}
$$

The splitting of the field components is not necessary in the one-dimensional case, since E_{zy} and H_{yz} are equal to zero, so the field can be completely described by E_{zx} and H_{yx}. The one-dimensional system will first be examined, after which perfectly matched layers will be discussed for the 2- or 3-dimensional case as well.

In the next section, the reflectivity on the interface between a non-absorbing material and a layer of material in which only the electrical conductivity is non-zero will be derived. The section after that deals with magnetic conductivity only, followed by a section dealing with both electrical and magnetic conductivities, culminating in a condition on the conductivities that yields a Perfectly Matched Layer, i.e., a layer that causes no reflection on the transition between normal material and PML material.

2.1.1.1 Electrical conductivity only

If the magnetic conductivity equals zero, time-harmonic solutions for E have the following form in areas where the conductivity does not depend on x:

$$\begin{aligned} E_z &= E_z(0)\, e^{\pm i k_0 \hat{n} x} e^{i\omega t} \\ \hat{n}_E &\equiv \hat{n}_{E,re} + i\hat{n}_{E,im} = n\left(\sqrt{\frac{1}{2}\left(\sqrt{1+\frac{\sigma_{E,x}^2}{\omega^2\varepsilon^2}}+1\right)} - i\sqrt{\frac{1}{2}\left(\sqrt{1+\frac{\sigma_{E,x}^2}{\omega^2\varepsilon^2}}-1\right)}\right) \\ k_0 &\equiv \omega\sqrt{\varepsilon_0\mu} \\ n &\equiv \sqrt{\frac{\varepsilon}{\varepsilon_0}} \end{aligned} \tag{2.6}$$

which are solutions that propagate to the left or right, depending on the sign in the exponentials, while decreasing in amplitude. When there is a jump in the conductivity, both E_z and H_y must be continuous. In the CW limit, i.e. when dealing with time-harmonic solutions, equation (2.5) shows that continuity of H_y implies continuity of $\frac{1}{\omega\mu - j\sigma_{H,x}}\frac{\partial}{\partial x}E_z$. If the magnetic conductivity is zero, this means that the normal derivative of E_z is continuous.

Consider a wave coming in from the left in a semi-infinite layer of lossless material with the same dielectric constant and magnetic permeability as the lossy material. At x=0, the material changes to a material with electrical conductivity $\sigma_{E,x}$. In the conducting material, only the right-traveling wave may exist with amplitude t, since there is no source at positive x. In the lossless material, there may be both a left-traveling wave with amplitude r and the incoming right-traveling wave with amplitude 1. The boundary conditions then give the following equations:

$$\begin{aligned} 1 + r &= t \\ -k_0 n + r k_0 n &= -t k_0 \hat{n}_E \end{aligned} \tag{2.7}$$

from which it follows that

$$r = \frac{n - \hat{n}_E}{n + \hat{n}_E} \tag{2.8}$$

2.1.1.2 Magnetic conductivity only

If, on the other hand, the electrical conductivity is zero while the magnetic conductivity is nonzero, the solution becomes:

$$\begin{aligned} E_z &= E_z(0)\, e^{\pm i k_0 \hat{n} x} \\ \hat{n}_H &\equiv \hat{n}_{H,re} + i\hat{n}_{H,im} = \sqrt{\frac{e}{e_0}} \left(\sqrt{\frac{1}{2}\left(\sqrt{1 + \frac{s_{H,x}^2}{w^2 m^2}} + 1 \right)} - i \sqrt{\frac{1}{2}\left(\sqrt{1 + \frac{s_{H,x}^2}{w^2 m^2}} - 1 \right)} \right) \\ k_0 &\equiv w \sqrt{e_0 m} \end{aligned} \tag{2.9}$$

and the interface conditions yield the reflection coefficient for the interface between a region without and a region with magnetic conductivity:

$$r = \frac{n - \dfrac{1}{1 - i\dfrac{s_{H,x}}{m w}} \hat{n}_H}{n + \dfrac{1}{1 - i\dfrac{s_{H,x}}{m w}} \hat{n}_H} \tag{2.10}$$

2.1.1.3 General conductivity

In general, the reflection coefficient from a nonlossy medium to a medium with both electrical and magnetic conductivities can be calculated to be:

$$r = \frac{1 - i\dfrac{s_{H,x}}{w m} - \dfrac{e_0}{e} \hat{n}_E \hat{n}_H}{1 - i\dfrac{s_{H,x}}{w m} + \dfrac{e_0}{e} \hat{n}_E \hat{n}_H} \tag{2.11}$$

This gives one simple relation that must be satisfied in order to have a reflectionless transition:

$$1 - i\frac{s_{H,x}}{w m} = \frac{e_0}{e} \hat{n}_E \hat{n}_H$$

$$\Leftrightarrow \quad \frac{\sigma_{H,x}}{\mu} = \frac{\sigma_{E,x}}{\varepsilon} \equiv s_x \tag{2.12}$$

If this condition is satisfied there is no reflection on the transition from a region without conductivities and a region with these conductivities, or from the transition between regions with different nonzero s_x's. Therefore, no reflection occurs for any distribution of s_x; the general solution for a system in which s_x depends on x is:

$$E = E(0)\, e^{\pm ik_0\sqrt{\frac{\varepsilon}{\varepsilon_0}}x}\, e^{\pm k_0\sqrt{\frac{\varepsilon}{\varepsilon_0}}\int_0^x \frac{\sigma_x}{w}(x')dx'} \tag{2.13}$$

so no matter what s_x is, there are two uncoupled solutions, one of which can be considered left-traveling and one that can be considered right-traveling.

2.1.1.4 Reflectivity

When the calculation window is surrounded by a layer of absorbing material, there will be some reflection back into the window. This section calculates this reflection for the one-dimensional system. The outer boundary, i.e. the real boundary of the calculation window, will have a Dirichlet boundary condition; the field on the boundary is put at zero.

Consider a structure consisting of a semi-infinite layer of a material with complex refractive index n_1 and a layer of thickness d with complex refractive index n_2, bounded by a perfectly conducting mirror. Light impinging normally onto the interface between layer 1 and layer 2 will be partially reflected and partially transmitted. The transmitted part propagates through layer 2 to the mirror, where it is reflected and propagates back toward the interface. Each roundtrip in layer 2 gives a contribution to the field propagating backwards in layer one. The reflection of a continuous wave under normal incidence can thus be calculated. The reflection and transmission coefficients for light going from layer 1 to layer 2 will be defined as r_{12} and t_{12}; from layer 2 to layer 1 they are r_{21} and t_{21}. Then, the total reflectivity becomes:

$$R = \left| r_{12} + t_{12}t_{21}e^{-2k_0n_2d+p}\sum_{m=0}^{\infty}\left(r_{21}e^{-2k_0n_2d+p}\right)^m \right|^2 = \left| r_{12} + \frac{t_{12}t_{21}e^{-2k_0n_2d+p}}{1-r_{21}e^{-2k_0n_2d+p}} \right|^2 \tag{2.14}$$

The extra phase shift of π is due to the reflection at the perfectly conducting mirror.

The reflection and transmission coefficients and the propagation constants for the three types of absorbing layer (electric, magnetic and perfectly matched) can be calculated as shown above. Figure 2.1 shows the reflectivity as a function of the wavelength when the thickness of layer 2 is 1 μm and s_x is chosen such that the reflectivity of the PML is $(1/2500)^2$. The graph shows the reflectivity of the three considered types of absorbing boundary condition: electrical and magnetic absorbers

and PML's, and for comparison a Sommerfeld-like boundary condition with zero reflection for one wavelength (l=1.0 μm).

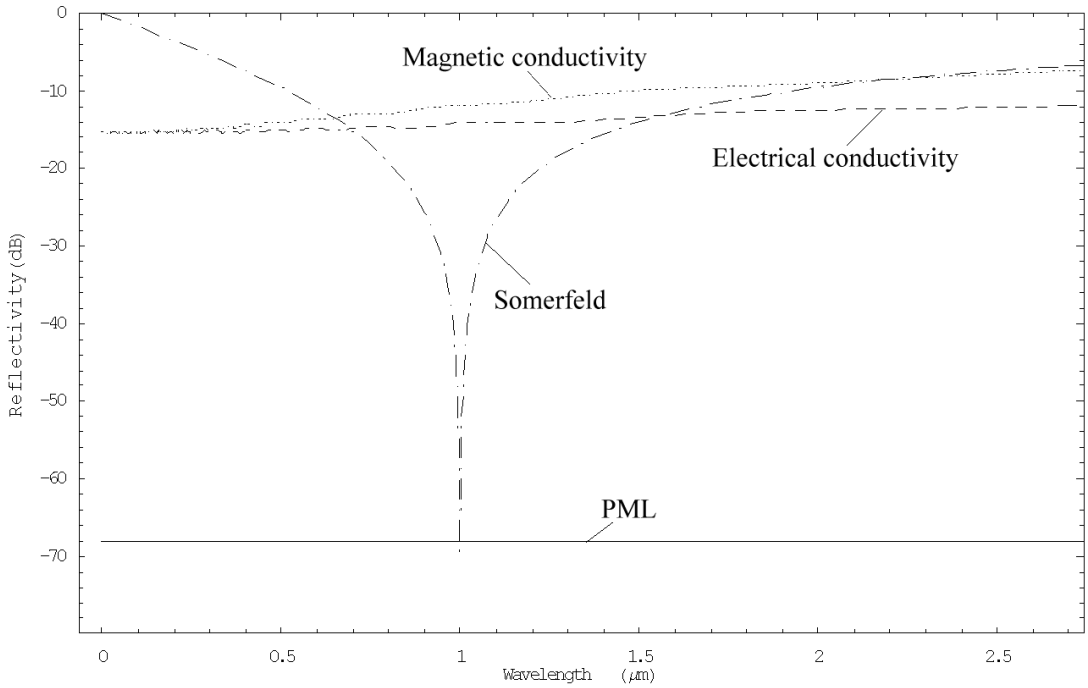


Figure 2.1: Reflectivity from a 1-micron thick absorbing layer, whose strength is normalized such that the PML gives (1/2500)^2 reflection. For comparison, reflection from a Sommerfeld-type boundary is also shown.

2.1.2 Two-dimensional Perfectly Matched Layer

In the spatially two-dimensional case, with the two spatial coordinates *x* and *z*, one can differentiate between the two polarizations TE and TM. For TE, one component of the $\vec{E}$ vector (E_y) and two components of the $\vec{H}$ vector (H_x and H_z) are nonzero; for TM, H_y, E_x and E_z are the only nonzero components. Because TM is very similar, we will first look at TE. We will look at the properties of the PML that needs to absorb light impinging on a boundary that is perpendicular to the x-axis; therefore, only s_x will be nonzero in the analysis.

If one were to use the unsplit equations (2.1) with the PML condition (2.12) for the conductivities, the time-harmonic solution for the equations:

$$\begin{aligned}
\frac{\partial}{\partial t}\varepsilon E_y &= \frac{\partial}{\partial z}H_x - \frac{\partial}{\partial x}H_z - \varepsilon s_x E_y \\
\frac{\partial}{\partial t}\mu H_x &= \frac{\partial}{\partial z}E_y - \mu s_x H_x \\
\frac{\partial}{\partial t}\mu H_z &= -\frac{\partial}{\partial x}E_y - \mu s_x H_z
\end{aligned} \tag{2.15}$$

is:

$$\begin{aligned}
E_y &= e^{i\omega t}e^{-\left(k_x\frac{s_x}{\omega}+ik_x\right)x}e^{-\left(k_z\frac{s_z}{\omega}+ik_z\right)z} \\
H_x &= -\frac{k_z}{\mu\omega}E_y \\
H_z &= \frac{k_x}{\mu\omega}E_y
\end{aligned} \tag{2.16}$$

in which $k_x^2 + k_z^2 = k_0^2 n^2 = \omega^2\varepsilon\mu$, which means that the ratio of the length of the $\vec{E}$-vector to the length of the $\vec{H}$-vector is exactly equal to $\sqrt{\mu/\varepsilon}$, the impedance in the same material without conductivity.

On an interface between a material where the conductivities are zero and one with nonzero s_x, there is no reflection for normal incidence, as shown in the 1-dimensional example. However, for non-normal incidence, there will be reflection. Consider a plane wave coming in from the non-attenuating medium with wave numbers $\hat{k}_x$ and $\hat{k}_z$, in which again $\hat{k}_x^2 + \hat{k}_z^2 = k_0^2 n^2 = \omega^2\varepsilon\mu$. The component of the wave vector parallel to the interface must be conserved across the interface, so k_z in equation (2.16) can be calculated to be:

$$k_z = \frac{\hat{k}_z}{1 - i\frac{s_x}{\omega}}$$

which means that k_x becomes:

$$k_x = \sqrt{\omega^2 \varepsilon \mu - k_z^2} = \sqrt{\omega^2 \varepsilon \mu - \frac{\hat{k}_z^2}{\left(1 - i\frac{\sigma_x}{\omega}\right)^2}}$$

Across the interface, both E_y and H_z must be continuous. From the timeharmonic version of equation (2.15), it shows that continuity of H_z implies continuity of $\frac{1}{1 - i\sigma_x/\omega}\frac{\partial}{\partial x}E_y$. If there were no reflection from the interface, there should be a solution with only the right-traveling wave in both areas, so the interface conditions should be satisfied directly. This would imply that $k_x = \hat{k}_x$, which is obviously not the case. Ergo, there will be reflection from the interface for oblique incidence.

For this reason, the variables are split as shown in equations (2.2), (2.3) and (2.4). Since we look only at the 2-dimensional TE case now, the necessary equation are:

$$\begin{aligned}
\frac{\partial}{\partial t}\varepsilon E_{yx} &= -\frac{\partial}{\partial x}H_z - \varepsilon\sigma_x E_{yx} \\
\frac{\partial}{\partial t}\varepsilon E_{yz} &= \frac{\partial}{\partial z}H_x \\
\frac{\partial}{\partial t}\mu H_x &= \frac{\partial}{\partial z}\left(E_{yx} + E_{yz}\right) \\
\frac{\partial}{\partial t}\mu H_z &= -\frac{\partial}{\partial x}\left(E_{yx} + E_{yz}\right) - \mu\sigma_x H_z
\end{aligned} \tag{2.17}$$

Solutions of these equations are:

$$\begin{aligned}
E_y \equiv E_{yx} + E_{yz} &= e^{i\omega t}e^{-ik_z z}e^{-\left(k_x\frac{\sigma_x}{\omega} + ik_x\right)x} \\
E_{yx} &= \frac{k_x^2}{\omega^2\varepsilon\mu}E_y \\
E_{yz} &= \frac{k_z^2}{\omega^2\varepsilon\mu}E_y \\
H_x &= -\sqrt{\frac{\varepsilon}{\mu}}\frac{k_x}{\omega\sqrt{\varepsilon\mu}}E_y \\
H_z &= \sqrt{\frac{\varepsilon}{\mu}}\frac{k_z}{\omega\sqrt{\varepsilon\mu}}E_y
\end{aligned} \tag{2.18}$$

in which, again, $k_x^2 + k_z^2 = k_0^2 n^2 = \omega^2\varepsilon\mu$.

With these solutions, there is no reflection at all when coming from a nonabsorbing medium into the PML; since $k_z = \hat{k}_z$, k_x is also equal to $\hat{k}_x$, satisfying the interface conditions.

In the time-harmonic case with frequency ω, equations (2.17) can be reduced to a single equation in the following manner:

$$
\begin{aligned}
& i\omega\varepsilon E_{yx} = -\frac{\partial}{\partial x} H_z - \sigma_x E_{yx} \\
& i\omega\varepsilon E_{yz} = \frac{\partial}{\partial z} H_x \\
& i\omega\mu H_x = \frac{\partial}{\partial z} E_y \\
& i\omega\mu H_z = -\frac{\partial}{\partial x} E_y - \mu\sigma_x H_z \\
& \Leftrightarrow \\
& E_{yx} = -\frac{1}{\omega^2\varepsilon\mu} \frac{1}{1 - i\frac{\sigma_x}{\omega}} \frac{\partial}{\partial x} \left(\frac{1}{1 - i\frac{\sigma_x}{\omega}} \frac{\partial}{\partial x} E_y \right) \\
& E_{yz} = -\frac{1}{\omega^2\varepsilon\mu} \frac{\partial^2}{\partial z^2} E_y \\
& \Leftrightarrow \\
& \frac{\partial^2}{\partial z^2} E_y + \frac{1}{1 - i\frac{\sigma_x}{\omega}} \frac{\partial}{\partial x} \left(\frac{1}{1 - i\frac{\sigma_x}{\omega}} \frac{\partial}{\partial x} E_y \right) + \omega^2\varepsilon\mu E_y = 0
\end{aligned}
\tag{2.19}
$$

The latter equation is very much like the standard TE Helmholtz equation, but with the *x*-coordinate stretched into the imaginary axis.

2.1.3 Performance of Perfectly Matched Layer for modes

If the index of refraction is constant in the x-direction, but varies in the *z*-direction, modes may be supported for which

$$E_y(x,z) = a(x)e(z) \tag{2.20}$$

and

$$\partial_{zz} e(z) + \omega^2 \varepsilon(z)\mu_0 e(z) = \beta^2 e(z) \tag{2.21}$$

for any *z*. For the normal Helmholtz equation,

$$a(x) = \exp(\pm ib\, x) \tag{2.22}$$

In the PML, this changes into

$$a(x) = a(0)\exp(\pm ib\, x)\exp\left(\pm b \int_0^x \frac{s_x}{w}\, dx'\right) \tag{2.23}$$

So the transverse profile of the mode does not change in the PML; the only effect of the PML is the fact that it makes the propagation constant complex, effectively attenuating any mode, whether it travels to the right or to the left. However, since each mode has a different b, they will all be attenuated at a different rate.

2.2 Sommerfeld-like boundary conditions

PML-type boundary conditions work well in practice. However, there are also other methods; one of these is a Sommerfeld-type boundary condition. It may be possible to combine PML's with these boundary conditions; in the previous section, there was unitary reflection on the boundary outside the PML. This reflection could be reduced by using a Sommerfeld boundary condition on the outer boundary. This section will describe the properties of Sommerfeld boundary conditions, and show the problems that occur when one tries to improve upon the method by optimizing for multiple wave numbers at once.

The discretized analogue of the (second-order) differential operator always leads to the problem that on the boundaries of the computational window, one has to make some assumptions or estimates of the fields in one or more discretized points just outside the window. Assuming, for example, that the fields in these points are zero, would lead to full reflection. For the rest of this section, we will assume a discretization stencil that uses only one off-central point on either side, so only one boundary field will have to be taken into account in the calculation of the boundary condition. Extension to more complicated stencils is straightforward.

If one assumes that the light impinging on the boundary consists of one or more, say M, plane waves, the field on the boundary can be expressed in the field on M points inside the calculation window. This means that, for computational methods with a correct implementation of the above, there should be no reflection on the boundary for these specific plane waves; the effect for other plane waves will be discussed below. For the expressions below we consider light at a frequency w, propagating in a uniform material with index n. So, the wave number perpendicular to the boundary is $nw\sin(q)/c$, with q the angle of incidence. The light impinges on the boundary at x=0; the boundary is perpendicular to x and the discretization in x is taken uniform with stepsize Dx. We will use the symbol y for the time-independent part of the field. A subscript denotes the field at a grid point; $y_m = y(x = mDx)$.

The field y_0, on the boundary of the 2D calculation window, will be expressed in M fields inside the calculation window, y_1 to y_M. So:

$$y_0 = \sum_{m=1}^{M} a_m y_m \tag{2.24}$$

Below, expressions for the a's will be derived for the case that the field is supposed to consist of M incoming plane waves with transverse wave numbers k_m, so

$$y(x) = \sum_{m=1}^{M} c_m e^{ik_m x} \tag{2.25}$$

The number of plane waves is chosen equal to the number of points that the field on the boundary is expressed in; one might also choose to use more and adopt a least-squares algorithm in the rest of the algorithm. However, we keep the numbers both equal to M.

From this, the field on each discretized point can be calculated in terms of the c_m's and k_m's:

$$y_l = \sum_{m=1}^{M} c_m e^{ik_m l\Delta x} \tag{2.26}$$

This can be written in a matrix-vector equation, in which the vectors have length M and the matrix A is an MxM matrix:

$$\vec{y} = A\vec{c} \tag{2.27}$$

where the elements of the matrix A are given by:

$$A_{lm} = \left(e^{ik_m \Delta x}\right)^l$$

and $\vec{y}$ is

$$\vec{y} = (y_1, \ldots, y_M)^T$$

This type of matrix is known as a Vandermonde matrix. If the matrix is not singular, which will only be the case if two wave numbers are equal or an integer number times $2\pi/\Delta x$ apart, the inverse of this matrix can be rewritten as a polynomial expansion. The vector $\vec{c}$ can thus be expressed explicitly as a function of the field on the points 1 to M.

Since y_0, expressed in the plane wave expansion (2.26), is equal to $\sum_{m=1}^{M} c_m$, this inverse Vandermonde matrix directly gives the desired coefficients α:

$$a_m = \sum_{l=1}^{M} \left(A^{-1} \right)_{lm} \tag{2.28}$$

Of course, it is not absolutely necessary to calculate the inverse of A; it is sufficient to find a's for which $A^* a = \mathbf{1}$.

It is possible to calculate the reflectivity from the boundary, given a set of a's. For an incoming plane wave with a wave number in the x-direction belonging to one of the chosen values k_m, the reflectivity will be zero, but also for other wave numbers, the reflectivity can be calculated. This is done by assuming the field inside the calculation window consists of an incoming plane wave with amplitude 1, and a reflected wave with amplitude r:

$$y(x) = e^{ik_x x} + re^{-ik_x x} \tag{2.29}$$

Now a value of r is sought that complies with the boundary condition (2.24), into which equation (2.29) is plugged for each y_m, so:

$$\begin{aligned} 1 + r &= \sum_{m=1}^{M} a_m \left(e^{imk_x \Delta x} + re^{-imk_x \Delta x} \right) \\ \Leftrightarrow r &= -\frac{1 - \sum_{m=1}^{M} a_m e^{imk_x \Delta x}}{1 - \sum_{m=1}^{M} a_m e^{-imk_x \Delta x}} \end{aligned} \tag{2.30}$$

In a discretized system, the boundary condition described above can either be directly implemented into the matrix equation that results from e.g. the discretization of the full Helmholtz equation, or be used to calculate the field on the boundary in unidirectional methods in space (unidirectional BPM) or time (FDTD). However, there is a problem in this algorithm when one tries to implement it for larger M (M>1); the scheme becomes very sensitive to small changes in the values of a_m. As an example, we calculate the reflectivity ($|r|^2$) for M=4 for an implementation into the unidirectional beam propagation method, in which the lateral stepsize is 0.05 μm and the wavelength is 1.32 μm. Figure 2.2 shows the analytical reflectivity for a's corresponding to zero-reflection for four incoming plane waves at angles of 5, 10, 15 and 20 degrees.

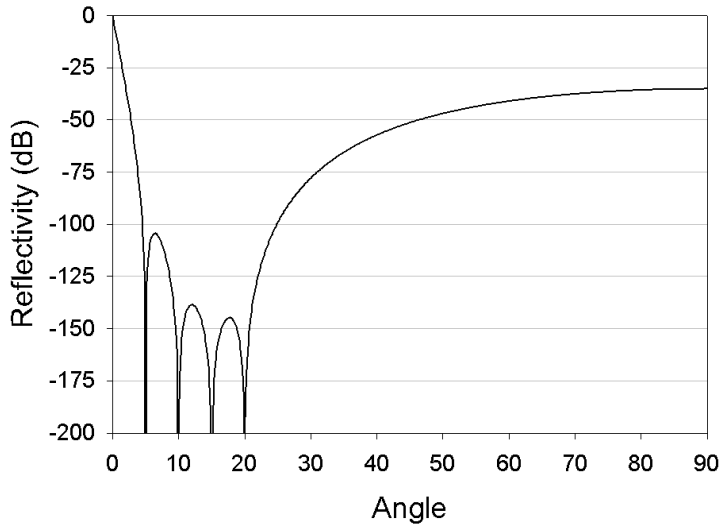


Figure 2.2: Reflection from a Somerfeld-like boundary condition with coefficients such that the reflection is zero for the four angles 5, 10, 15 and 20^0.

When the coefficient α_1 is increased by 0.001%, the resulting reflectivity graph becomes as shown in Figure 2.3:

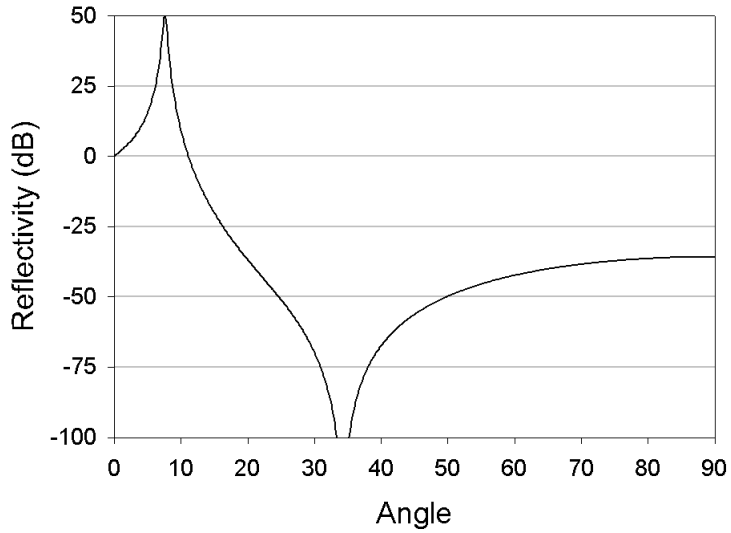


Figure 2.3: The same picture as in Figure 2.1, but with one coefficient in the calculations (a_1) increased by 0.001%

This shows the extreme dependence of the reflectivity on the coefficients a when the field on the boundary is expressed in a larger number of plane waves; when they are wrong by even a small amount, the reflectivity becomes higher than one. This effect

even occurs for M=2; a configuration can be found where a 1% change in a_1 makes the field at the boundary blow up. This type of boundary condition needs further study; if it can be made stable, it is a very promising method, especially in combination with the Perfectly Matched Layers that will be described in the next section.

This type of boundary condition can in principle be made adaptive, making the coefficients k_m depend on the field inside the calculation window. When using only one incoming plane wave (M=1), this is known as Hadley's Transparent Boundary Condition (TBC). In a practical implementation for a unidirectional BPM, FDTD, or omnidirectional Helmholtz solver, k_1 is determined from the argument of the ratio of y_1 and y_2 on a previous propagation or iteration step:

$$\frac{y_1}{y_2} = \frac{e^{ik_1\Delta x}}{e^{2ik_1\Delta x}} = e^{-ik_1\Delta x}$$

$$\Leftrightarrow k_1 = \frac{i}{\Delta x}\ln\left(\frac{y_1}{y_2}\right)$$

If this k_1 is positive, it represents an outgoing wave, which is acceptable; however, a negative k_1 represents an incoming wave. When this happens, k_1 is set equal to zero in order not to have any influx through the boundary. In implementations of the BPM algorithm, this turns out to work well as long as the boundary is not struck simultaneously by plane waves with a wide range of angles.

For the unidirectional methods, this adaptive method poses no great problems; the field on the previous propagation step can be used to calculate k_1. In an iterative solver, like the Helmholtz solver that will be discussed in Chapter 4, the matrices involved in the calculation would have to be adjusted every time k_1 changes. This means that many iterative methods, e.g. conjugate gradient, would have to be re-initialized, since their search directions depend on the matrix and the previous search directions. This re-initialization would have a negative influence on the performance of the algorithm.

Hadley's boundary condition works well in many cases, and has proved to be stable. However, since the version of Sommerfeld with the boundary field expressed in more than one plane wave is very sensitive to small errors, it is unlikely that an adaptive version would be stable.

2.3 Conclusions

Using conducting layers as boundary conditions is a robust way to implement nonreflecting boundary conditions. One needs a special combination of electrical and magnetic conductivities in order to have a reflectionless transition from the non-conducting material to the conducting material. This special combination is called the Perfectly Matched Layer (PML). For the two- or three-dimensional PML, the time-dependent equations must be split in order to have directional attenuation and to have zero reflection at the interface for both normal and oblique incidence and for modal fields. In the time-harmonic case (Helmholtz equation), the PML corresponds to the addition of some extra coefficients into the equation. For an incoming modal field, the lateral field distribution does not change in the PML; however, the amplitude of the modal field decreases as the mode traverses the PML. A property of modal propagation in the PML which will turn out to be very important in Chapter 4 is the fact that the attenuation of a modal field depends on its propagation constant b.

It may be possible to use a combination of Sommerfeld-like boundary conditions with PML's; in the method described here, the field on the outside of the PML is set at zero, thus causing complete reflection. This reflection may be reduced by means of a Sommerfeld-like outer boundary. Some extra research will be necessary; inside the PML, the field can not longer be considered to be composed of plane waves, since all waves decay normal to the boundary. This will have to be taken into account in the implementation of this combined boundary condition. Also, more work will be necessary if one wants to implement a stable Sommerfeld-like boundary condition that is optimized for more than one incoming plane wave.

Chapter 3: Unidirectional Beam Propagation Methods

3.1 Introduction

Beam propagation methods are being used widely in integrated optics [Hoekstra 1997, and references therein]. Several methods are known, e.g. Fourier Transform (FT) BPM [Thylen 1983], Finite Difference (FD) BPM [Yevick at al 1984], the Method of Lines (MoL) [Pregla 1995], and others. This chapter will describe improvements of the FD BPM.

In 2-dimensional beam propagation methods, one wants to solve the Helmholtz equation for either TE or TM:

$$\begin{aligned} TE:&\quad \left(\partial_{zz} + \partial_{xx} + k_0^2 n^2\right) E_y = 0 \\ TM:&\quad \left(\partial_z \frac{1}{n^2} \partial_z + \partial_x \frac{1}{n^2} \partial_x + k_0^2\right) H_y = 0 \end{aligned} \tag{3.1}$$

In systems in which the index of refraction is piecewise constant, the TM equation reduces to the TE equation everywhere except on interfaces between materials. This means that only the interface conditions for the fields are different for TE and TM.

In the unidirectional BPM methods that are considered here, the lateral index distribution is presumed to be invariant in the propagation direction during each propagation step. This means that interfaces between materials that are not parallel to the propagation direction are approximated by a staircase. It also means that the derivative of the index with respect to the propagation direction is zero along the propagation step, so the first term in the TM Helmholtz equation becomes identical to that in the TE equation.

In order to create a unidirectional BPM, the Slowly Varying Envelope Approximation (SVEA) must be made [Hoekstra 1997]. This way, one gets rid of the second derivative in the propagation direction, z. A constant reference index n_0 is introduced, which should be close to the effective index of the light propagating in the structure. The following substitution for the field is made:

$$E_y(x,z) = \psi(x,z)\, e^{-ik_0 n_0 z} \tag{3.2}$$

and similar for TM. If one assumes that the light mainly propagates with wave number $k_0 n_0$, $\partial_{zz}\psi \ll 2k_0 n_0 \partial_z \psi$ and can be neglected. Then the TE part of Eq. (3.1) becomes:

$$-2ik_0 n_0 \partial_z \psi + \partial_{xx}\psi + k_0^2 \left(n^2 - n_0^2\right)\psi = 0 \tag{3.3}$$

For TM, the same equation may be used everywhere except on interfaces between materials.

If the reference index n_0 is not very close to the actual effective index of a mode, there will be considerable phase errors. These phase errors can be reduced by going to the so-called wide-angle approximations [Hoekstra et al 1993, Hoekstra et al 1992[b]], which use a more accurate approximation for the second-order derivative with respect to *z*. In this chapter, this wide-angle approach will not be taken, but it is relatively simple to adapt the presented algorithms to make use of this approach.

The simulation window is discretized into a rectangular grid, i.e. a grid with constant stepsizes Δx and Δz along the x- and z-directions. In the rest of this chapter, the value of a quantity *f* on a grid point will be written as $f_m^l = f\left(m\Delta x, l\Delta z\right)$; the vector of values with the same z-position $l\Delta z$ is $\mathbf{f}^l = \left(f_0^l, f_1^l, ..., f_{N-1}^l, f_N^l \right)^T$. A matrix on $z=l\Delta z$ is written as M^l.

In all beam propagation methods described here, the differentiation z is approximated as:

$$\partial_z y_m^{l+\frac{1}{2}} = \frac{y_m^{l+1} - y_m^l}{\Delta z} + O\left(\Delta z^2\right) \tag{3.4}$$

The other terms in Eq. (3.3) can be collected into a matrix-vector product:

$$\partial_{xx} y_m^l + k_0^2 \left(n_m^{l\,2} - n_0^2 \right) y_m^l = \sum_j M_{mj}^l y_j^l \tag{3.5}$$

where the matrix M^l must be chosen symmetric for reasons given below. However, the matrix product $M^l \mathbf{y}^l$ is evaluated at $z=l\Delta z$, while Eq. (3.4) evaluates at $z=(l+1/2)\Delta z$. So, the more accurate way to use *M* is by taking $M^l (\mathbf{y}^l + \mathbf{y}^{l+1})/2$, which does evaluate halfway between discretization lines; this yields the standard Crank-Nicolson integration scheme. Note that the since the index distribution is constant in the propagation direction during a calculation step, the same matrix operates on the vectors $\mathbf{y}^l$ and $\mathbf{y}^{l+1}$. This fact means that when interfaces between materials are not parallel to the propagation direction, they are approximated by a staircase. In order to perform accurate simulations, one may have to adjust the stepsize in *z* according to the angle of the interfaces. So the complete discretized algorithm becomes:

$$\frac{4ik_0 n_0}{\Delta z} \mathbf{y}^{l+1} - M^l \mathbf{y}^{l+1} = \frac{4ik_0 n_0}{\Delta z} \mathbf{y}^l + M^l \mathbf{y}^l \tag{3.6}$$

It will prove necessary to rewrite the first term on the left and right side as a matrix-vector product, which gives:

$$\left(A^l - M^l \right) \mathbf{y}^{l+1} = \left(A^l + M^l \right) \mathbf{y}^l \tag{3.7}$$

Disregarding any boundary conditions, the matrices A and M are symmetric and have purely imaginary and purely real elements, respectively, which also means that their eigenvalues are imaginary and real. Using the notation $\langle j^{\nu} | A | y^{\nu} \rangle = \left(j^{\nu}\right)^{*} A y^{\nu}$, the following derivation:

$$\langle y^{\nu+1} | A^l - M^l | y^{\nu+1} \rangle = \langle y^{\nu+1} | A^l + M^l | y^{\nu} \rangle = \langle -\left(A^l - M^l\right) y^{\nu+1} | y^{\nu} \rangle =$$
$$\langle -\left(A^l + M^l\right) y^{\nu} | y^{\nu} \rangle = \langle y^{\nu} | A^l - M^l | y^{\nu} \rangle$$
$$\Leftrightarrow$$
$$\langle y^{\nu+1} | A^l | y^{\nu+1} \rangle - \langle y^{\nu+1} | M^l | y^{\nu+1} \rangle = \langle y^{\nu} | A^l | y^{\nu} \rangle - \langle y^{\nu} | M^l | y^{\nu} \rangle$$

shows that both $\langle y^{\nu} | A | y^{\nu} \rangle$ and $\langle y^{\nu} | M | y^{\nu} \rangle$ are conserved quantities, since the terms containing A are purely imaginary and the terms containing M are purely real, for real values of n^2.

In the paraxial approximation, the integral across a cross-section of the component of the Poynting vector in the direction of propagation should be conserved, since all energy is assumed to flow in the direction of propagation. The Poynting vector is:

$$\mathbf{S} = \frac{1}{2}\mathrm{Re}\left(\mathbf{E} \times \mathbf{H}^{*}\right) \tag{3.8}$$

So the component of the Poynting vector in the propagation direction for a propagating mode with propagation constant $k_0 n_{eff}$ is for the two polarizations:

$$\begin{aligned} TE: \quad S_z &= \frac{1}{2}\mathrm{Re}\left(E_y \cdot H_x^{*}\right) = \frac{1}{2}\mathrm{Re}\left(-E_y \cdot \frac{i}{\omega \mu_0}\partial_z E_y^{*}\right) = \frac{n_{eff}}{\mu_0 c}\left|E_y\right|^2 \\ TM: \quad S_z &= \frac{1}{2}\mathrm{Re}\left(E_x \cdot H_y^{*}\right) = \frac{1}{2}\mathrm{Re}\left(-\frac{i}{\omega \varepsilon_0 n^2}\partial_z H_y \cdot H_y^{*}\right) = \frac{n_{eff}}{\varepsilon_0 n^2 c}\left|H_y\right|^2 \end{aligned} \tag{3.9}$$

So for TE, we want the integral of $|E_y|^2$ along x to be conserved, while for TM, the integral of $|H_y|^2/n^2$ should be conserved. A discretized version of these integrals is the dot product $\langle E_y^{\nu} | E_y^{\nu} \rangle$ for TE and $\langle H_y^{\nu}/n | H_y^{\nu}/n \rangle$ for TM. If the simulation scheme as in equation (3.6) is used, the quantity $\langle y^{\nu} | y^{\nu} \rangle$ is conserved through a propagation step. So, if this scheme is used and the standard dot product is used to calculate e.g. overlap integrals, it is necessary to use H_y/n for y in order to assure conservation of the norm $\langle H_y^{\nu}/n | H_y^{\nu}/n \rangle$.

The focus of this chapter will be on obtaining the matrices A and M. The second order derivative with respect to x should be discretized with a certain order of accuracy in the lateral stepsize Δx. The problems are located on the points close to interfaces, since the field is not smooth at the interface; for TE, there are discontinuities in the second and higher order derivatives, while for TM, the jumps already occur in the first

derivative. Since the errors that occur near interfaces happen on a fixed number of points, irrespective of the lateral stepsize, the accuracy on these points does not need to be of the same order in the lateral stepsize as the accuracy on other points; one order of accuracy may be lost, since the relative contribution of this error to the global error decreases as the stepsize decreases (and thus, the number of discretization points increases). By 'global error' we mean the error in integrated quantities, since one is usually interested in the total power carried or the power carried by specific waveguide modes, and the propagation constants of these modes, rather than in the exact fields at discrete points.

First, a method based upon a finite difference scheme with Taylor expansions of the fields around interfaces is investigated for second and fourth order accuracy in the stepsize for the propagation constant. Since we there is no easy road to obtain explicit formulas for fourth order accuracy in TM polarization with this method, a different method based on a variational approach for the lateral direction will be investigated.

3.2 Finite difference: Taylor expansions around interfaces

We consider an interface between two materials. Without loss of generality, the interface may be assumed to be at the x-position (½ + γ)Δx, so $\gamma \in [-½, ½]$ denotes the position of the interface within the grid cell, as shown in Figure 1.

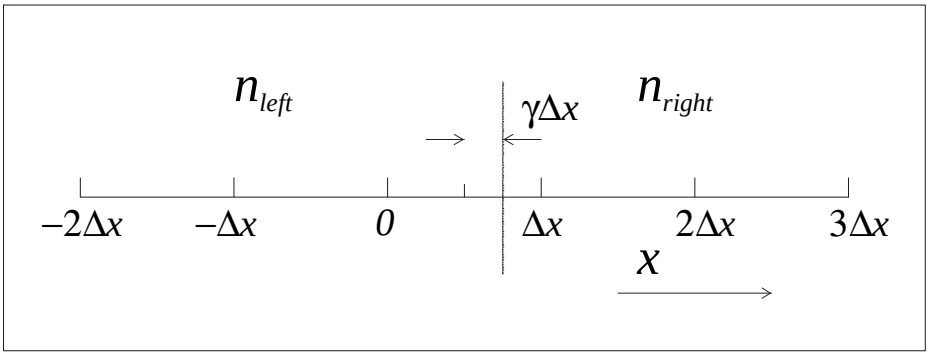


Figure 3.1: Position of interface between materials

At the interface, the field is not smooth. There are discontinuities in first (TM) or second (TE) and higher order derivatives. The field and its derivatives on the right side of the interface ($E_{left} \equiv \lim_{x \downarrow (g+\frac{1}{2})\Delta x} E_y$, etc.) can be expressed in their values on the left side ($E_{right} \equiv \lim_{x \uparrow (g+\frac{1}{2})\Delta x} E_y$, etc.) and average values on the interface using the following relationships:

$$\text{TE:}\begin{cases} E_{right} = E_{left} \\ \partial_x E_{right} = \partial_x E_{left} \\ \left(\partial_{xx} + k_0^2 n_{right}^2\right) E_{right} = \left(\partial_{xx} + k_0^2 n_{left}^2\right) E_{left} \\ \partial_x \left(\partial_{xx} + k_0^2 n_{right}^2\right) E_{right} = \partial_x \left(\partial_{xx} + k_0^2 n_{left}^2\right) E_{left} \\ \left(\partial_{xx} + k_0^2 n_{right}^2\right)^2 E_{right} = \left(\partial_{xx} + k_0^2 n_{left}^2\right)^2 E_{left} \end{cases}$$

$$\text{TM:}\begin{cases} H_{right} = H_{left} \\ \frac{1}{n_{right}^2}\partial_x H_{right} = \frac{1}{n_{left}^2}\partial_x H_{left} \\ \left(\partial_{xx} + k_0^2 n_{right}^2\right) H_{right} = \left(\partial_{xx} + k_0^2 n_{left}^2\right) H_{left} \\ \frac{1}{n_{right}^2}\partial_x \left(\partial_{xx} + k_0^2 n_{right}^2\right) H_{right} = \frac{1}{n_{left}^2}\partial_x \left(\partial_{xx} + k_0^2 n_{left}^2\right) H_{left} \\ \left(\partial_{xx} + k_0^2 n_{right}^2\right)^2 H_{right} = \left(\partial_{xx} + k_0^2 n_{left}^2\right)^2 H_{left} \end{cases}$$

For these conditions, we made use of the last two Maxwell's equations, combined with the Helmholtz equations. The last two conditions of both TE and TM are derived from the modal field equation:

$$\partial_{xx} y^{\,i} + k_0^2 n^2 y^{\,i} = b^{\,i^2} y^{\,i} \tag{3.10}$$

where i is the mode number and $y = E_y$ and H_y for TM. Since in the uniform part either left or right of the interface, this equation holds for both TE and TM, the third condition can be directly derived for each mode; and when it holds for each mode, it will also hold for the complete field. The fourth and fifth condition for both polarizations can be derived in a similar fashion by taking the derivative of Equation (3.10) with respect to x, respectively by plugging both sides of (3.10) into the Helmholtz equation. From these conditions, we can calculate the fields and their derivatives on the right side of the interface in terms of the same fields and derivatives on the left side, and the average value of fields and derivatives on the interface. The average value of a quantity on the interface is defined as the average of the value tot the left and the value to the right of the interface, and denoted by a bar over the symbol. For the sake of simplicity, the vacuum wave number and permittivity contrast will be lumped into the parameter a:

$$a \equiv k_0^2 \left(n_{left}^2 - n_{right}^2\right) \tag{3.11}$$

The above leads to:

$$\text{TE:}\begin{cases} E_{right} = E_{left} \\ \partial_x E_{right} = \partial_x E_{left} \\ \partial_{xx} E_{right} = \partial_{xx} E_{left} + a\bar{E} \\ \partial_{xxx} E_{right} = \partial_{xxx} E_{left} + a\bar{E}' \\ \partial_{xxxx} E_{right} = \partial_{xxxx} E_{left} + 2a\bar{E}'' \end{cases} \tag{3.12}$$

and:

$$\text{TM:}\begin{cases} H_{right} = H_{left} \\ \partial_x H_{right} = \dfrac{n_{left}^2}{n_{right}^2}\partial_x H_{left} \\ \partial_{xx} H_{right} = \partial_{xx} H_{left} + a\bar{H} \\ \partial_{xxx} H_{right} = \dfrac{n_{left}^2}{n_{right}^2}\partial_{xxx} H_{left} + \dfrac{2n_{right}^2}{n_{right}^2 + n_{left}^2} a\bar{H}' \\ \partial_{xxxx} H_{right} = \partial_{xxxx} H_{left} + 2a\bar{H}'' \end{cases} \tag{3.13}$$

If we use $\psi = \frac{1}{n} H_y$, the interface conditions (3.13) change into:

$$\begin{cases} \psi_{right} = \dfrac{n_{left}}{n_{right}} \psi_{left} \\ \partial_x \psi_{right} = \dfrac{n_{right}}{n_{left}} \partial_x \psi_{left} \\ \partial_{xx} \psi_{right} = \dfrac{n_{left}}{n_{right}} \partial_{xx} \psi_{left} + \dfrac{2n_{left}}{\left(n_{left} + n_{right}\right)} \alpha\psi^- \\ \partial_{xxx} \psi_{right} = \dfrac{n_{right}}{n_{left}} \partial_{xxx} \psi_{left} + \dfrac{2n_{right}}{\left(n_{left} + n_{right}\right)} \alpha\psi^{-\prime} \\ \partial_{xxxx} \psi_{right} = \dfrac{n_{left}}{n_{right}} \partial_{xxxx} \psi_{left} + \dfrac{4n_{left}}{\left(n_{left} + n_{right}\right)} \alpha\psi^{-\prime\prime} + \dfrac{2n_{left}\left(n_{left} - n_{right}\right)^2}{\left(n_{left} + n_{right}\right)} \alpha\psi^- \end{cases} \tag{3.14}$$

3.2.1 Second-order accuracy

As an example, we calculate the interface corrections for a second-order TE scheme. If there were no interface, with the whole area composed of the left material, we could calculate Taylor expansions of the fields at points (-1, 0, 1) around the interface, all expressed in the fields and derivatives on the left-hand side (lhs) of the interface, e.g. up to order Δx^4. Using the notation for e.g. the point at $x=-D\,x$ for the field expressed in this way as $E_{-1,left}$, this becomes:

$$
\begin{aligned}
E_{-1,left} &= E_{left} - (1\tfrac{1}{2}+g)\Delta x\partial_x E_{left} + \frac{1}{2}(1\tfrac{1}{2}+g)^2\Delta x^2\partial_{xx}E_{left} \\
&\quad - \frac{1}{6}(1\tfrac{1}{2}+g)^3\Delta x^3\partial_{xxx}E_{left} + O\left(\Delta x^4\right) \\
E_{0,left} &= E_{left} - (\tfrac{1}{2}+g)\Delta x\partial_x E_{left} + \frac{1}{2}(\tfrac{1}{2}+g)^2\Delta x^2\partial_{xx}E_{left} \\
&\quad - \frac{1}{6}(\tfrac{1}{2}+g)^3\Delta x^3\partial_{xxx}E_{left} + O\left(\Delta x^4\right) \\
E_{1,left} &= E_{left} + (\tfrac{1}{2}-g)\Delta x\partial_x E_{left} + \frac{1}{2}(\tfrac{1}{2}-g)^2\Delta x^2\partial_{xx}E_{left} \\
&\quad + \frac{1}{6}(\tfrac{1}{2}-g)^3\Delta x^3\partial_{xxx}E_{left} + O\left(\Delta x^4\right)
\end{aligned}
$$

The same can be done using only the field, derivatives and refractive index on the right-hand side (rhs) of the interface, for points (0, 1, 2):

$$
\begin{aligned}
E_{0,right} &= E_{left} - (\tfrac{1}{2}+g)\Delta x\partial_x E_{left} + \frac{1}{2}(\tfrac{1}{2}+g)^2\Delta x^2\left(\partial_{xx}E_{left} + a\bar{E}\right) \\
&\quad - \frac{1}{6}(\tfrac{1}{2}+g)^3\Delta x^3\left(\partial_{xxx}E_{left} + a\bar{E}'\right) + O\left(\Delta x^4\right) \\
E_{1,right} &= E_{left} + (\tfrac{1}{2}-g)\Delta x\partial_x E_{left} + \frac{1}{2}(\tfrac{1}{2}-g)^2\Delta x^2\left(\partial_{xx}E_{left} + a\bar{E}\right) \\
&\quad + \frac{1}{6}(\tfrac{1}{2}-g)^3\Delta x^3\left(\partial_{xxx}E_{left} + a\bar{E}'\right) + O\left(\Delta x^4\right) \\
E_{2,right} &= E_{left} + (1\tfrac{1}{2}-g)\Delta x\partial_x E_{left} + \frac{1}{2}(1\tfrac{1}{2}-g)^2\Delta x^2\left(\partial_{xx}E_{left} + a\bar{E}\right) \\
&\quad + \frac{1}{6}(1\tfrac{1}{2}-g)^3\Delta x^3\left(\partial_{xxx}E_{left} + a\bar{E}'\right) + O\left(\Delta x^4\right)
\end{aligned}
$$

in the derivation of which equation (3.12) was used.

When using only the expansions from the lhs of the interface, the second order derivative on point 0 would be approximated by the standard finite difference operator:

$$\hat{\partial}_{xx}E_0 = \frac{E_{-1,left} - 2E_{0,left} + E_{1,left}}{\Delta x^2} = \partial_{xx}E_0 + O\left(\Delta x^2\right) \qquad (3.15)$$

A hat (^) on an operator denotes its discretized analogue.

However, due to the interface, the field E_1 should be expressed in an expansion from the rhs, instead of the lhs; so, $E_{1.right}$ should be used. Doing so, using the above, one gets:

$$\frac{E_{-1,left} - 2E_{0,left} + E_{1,right}}{\Delta x^2} = \frac{E_{-1,left} - 2E_{0,left} + E_{1,left}}{\Delta x^2} + \frac{1}{2}\left(\tfrac{1}{2} - g\right)^2 a\bar{E} + \frac{1}{6}\left(\tfrac{1}{2} - g\right)^3 \Delta x a\bar{E}'$$

So,

$$\partial_{xx}E_0 = \frac{E_{-1,left} - 2E_{0,left} + E_{1,right}}{\Delta x^2} - \frac{1}{2}\left(\tfrac{1}{2} - g\right)^2 a\bar{E} - \frac{1}{6}\left(\tfrac{1}{2} - g\right)^3 \Delta x a\bar{E}' + O\left(\Delta x^2\right)$$

Since the accuracy on these interface points needs only be $O(\Delta x)$ when the accuracy on other points is $O(\Delta x^2)$, only the term containing the average value of E on the interface needs be included, and only up to $O(\Delta x)$. The total discretization of the second derivative on point 0 then becomes:

$$\hat{\partial}_{xx}E_0 = \frac{E_{-1,left} - \left(2 + \frac{1}{2}\left(\tfrac{1}{2} - g\right)^2 a\right)E_{0,left} + E_{1,right}}{\Delta x^2} = \partial_{xx}E_0 + O\left(\Delta x\right) \qquad (3.16)$$

On the other side of the interface, at point 1, the same formula holds if one inverts both the sign of a and γ. An interpretation of this formula is that the interface is taken into account in a weighted fashion; when the interface is exactly central between two points, no correction is necessary, and if the interface moves away from this, the corrections on the two points adjacent to the interface increase, with opposite signs.

So, the global error of the scheme can be made $O(\Delta x^2)$ by only adjusting the diagonal elements of the matrix M, which allows the implemented computer algorithm to be fast.

3.2.2 Fourth-order accuracy

In regions away from interfaces, the second order derivative can be discretized up to fourth order in Δx as follows:

$$\hat{\partial}_{xx}y_m = \frac{-y_{m-2} + 16y_{m-1} - 30y_m + 16y_{m+1} - y_{m+2}}{12\Delta x^2} = \partial_{xx}y_m + O\left(\Delta x^4\right) \qquad (3.17)$$

The same procedure as above can be followed to calculate the error that is made in this standard finite difference operator when it is applied to points next to interfaces,

again expressed in terms of the average field and derivatives on the interface. Only the four points −1 to 2 are affected, since the discretization of the second derivative around the other points does not cross the interface. For TE ($y \equiv E_y$), this error, defined as $f_m = \left(\hat{\P}_{xx} E\right)_m - \P_{xx} E_m$, is:

$$
\begin{aligned}
f_{-1} &= -\frac{1}{24}\left(\tfrac{1}{2}-g\right)^2 a\bar{E}_{\text{int}} - \frac{1}{72}\left(\tfrac{1}{2}-g\right)^3 a\Delta x\bar{E}'_{\text{int}} - \frac{1}{144}\left(\tfrac{1}{2}-g\right)^4 a\Delta x^2\bar{E}''_{\text{int}} + O\left(\Delta x^3\right) \\
f_0 &= \left(\frac{2}{3}\left(\tfrac{1}{2}-g\right)^2 - \frac{1}{24}\left(\tfrac{3}{2}-g\right)^2\right)a\bar{E}_{\text{int}} + \left(-\frac{2}{9}\left(\tfrac{1}{2}-g\right)^3 - \frac{1}{72}\left(\tfrac{3}{2}-g\right)^3\right)a\Delta x\bar{E}'_{\text{int}} + \\
&\quad + \left(-\frac{1}{9}\left(\tfrac{1}{2}-g\right)^4 - \frac{1}{144}\left(\tfrac{3}{2}-g\right)^4\right)a\Delta x^2\bar{E}''_{\text{int}} + O\left(\Delta x^3\right) \qquad (3.18) \\
f_1 &= \tilde{f}_0 \\
f_2 &= \tilde{f}_{-1}
\end{aligned}
$$

where the ~ transformation means a, g and Δx should be replaced by their negative value. Note that $\tilde{\tilde{f}} = f$. The errors are only calculated up to order Δx^3, since this local error may be one degree lower than the desired global error.

These error terms must again be eliminated by adjusting the matrix that represents the finite difference second order derivative operator. There are a couple of conditions we will impose on these corrections on the matrix. These are:

- The resulting matrix must remain symmetrical
- The resulting matrix must remain pentadiagonal

The reason for the latter condition is ease of implementation and speed/memory efficiency. The first condition is necessary for ‘energy’ conservation, as discussed on page 23 near Equation (3.9). So, we can write down the part of the discretization matrix for the second derivative for the four affected points with corrections on those places where this is possible:

$$
\begin{aligned}
\hat{\P}_{xx} &= \left(\hat{\P}_{xx}\right)_{uniform} + C = \\
&= \frac{1}{12\Delta x^2}\begin{bmatrix}
-1 & 16 & -30+c_1 & 16+c_5 & -1+c_8 & 0 & 0 & 0 \\
0 & -1 & 16+c_5 & -30+c_2 & 16+c_6 & -1+c_9 & 0 & 0 \\
0 & 0 & -1+c_8 & 16+c_6 & -30+c_3 & 16+c_7 & -1 & 0 \\
0 & 0 & 0 & -1+c_9 & 16+c_7 & -30+c_4 & 16 & -1
\end{bmatrix} \qquad (3.19)
\end{aligned}
$$

Each of the terms on the right-hand side of equation (3.18) can be expressed in terms of the field on the four surrounding points. On each on of the points −1 to 2, there are three of these error terms. It might be possible to eliminate these error terms by adjusting three elements of the matrix on each line. This gives twelve more or less independent error terms that must be eliminated in total, and therefore twelve corrections in the matrix. However, due to the symmetry restriction, only nine of these

correction terms in the correction matrix C are independent. It turns out to be impossible to eliminate all error terms with these corrections. So, we are unable to get the *local error* in each point down to $O(\Delta x^3)$.

Therefore, we must approach the problem in a slightly different manner, by going toward a weak formulation of the problem. Instead of trying to get the error on each points near the interface to the right order:

$$C|E\rangle = -f + O\left(\Delta x^3\right) \tag{3.20}$$

we take the dot product of both the error f and the correction $C|E\rangle$ with the field y :

$$\langle E|C|E\rangle = -\langle E|f\rangle + O\left(\Delta x^3\right) \tag{3.21}$$

This dot product is defined as:

$$\langle E|f\rangle \equiv \sum_{m=-1}^{2} E_m f_m \tag{3.22}$$

This condition is sufficient to ensure global $O(\Delta x^4)$ convergence of the quantities that are generally interesting, since they are integrated quantities.

Since f is expressed in terms of the average value of E and the averaged derivatives on the interface, we write E on all points concerned as a Taylor expansion around the interface, taking into account the discontinuities at the interface. This gives the following equations:

$$\begin{aligned}
E_{-1} &= \left(1-\tfrac{1}{4}\Delta x^2 a\left(\tfrac{3}{2}+g\right)^2\right)\bar{E} - \Delta x\left(\tfrac{3}{2}+g\right)\bar{E}' + \tfrac{1}{2}\Delta x^2\left(\tfrac{3}{2}+g\right)^2\bar{E}'' + O\left(\Delta x^3\right)\\
E_{0} &= \left(1-\tfrac{1}{4}\Delta x^2 a\left(\tfrac{1}{2}+g\right)^2\right)\bar{E} - \Delta x\left(\tfrac{1}{2}+g\right)\bar{E}' + \tfrac{1}{2}\Delta x^2\left(\tfrac{1}{2}+g\right)^2\bar{E}'' + O\left(\Delta x^3\right)\\
E_{1} &= \text{[illegible]}\bar{E}_0\\
E_{2} &= \text{[illegible]}\bar{E}_{-1}
\end{aligned} \tag{3.23}$$

If we substitute (3.18) and (3.23) into (3.21), both sides of the resulting equation can then be written as the sum of four terms, yielding the following equation:

$$a\,\bar{E}^2 + b\,\bar{E}\bar{E}' + d\,\bar{E}'^2 + e\,\bar{E}\bar{E}'' = r\bar{E}^2 + s\bar{E}\bar{E}' + t\bar{E}'^2 + u\bar{E}\bar{E}'' \tag{3.24}$$

with:

$$a = \left\{ c_1 + c_2 + c_3 + c_4 + 2c_5 - \frac{a}{8}\Delta x^2 \left(\begin{array}{l} (2g+3)^2 c_1 + (2g+1)^2 c_2 - (2g-1)^2 c_3 - \\ (2g-3)^2 c_4 + 8g c_5 \end{array} \right) \right\} / 12\Delta x^2$$

$$b = -\frac{1}{12\Delta x}\left((2g+3)c_1 + (2g+1)c_2 + (2g-1)c_3 + (2g-3)c_4 + 4g c_5\right)$$

$$d = \frac{1}{48}\left((2g+3)^2 c_1 + (2g+1)^2 c_2 + (2g-1)^2 c_3 + (2g-3)^2 c_4 + 2\left(4g^2+1\right)c_5\right)$$

$$e = \frac{1}{48}\left((2g+3)^2 c_1 + (2g+1)^2 c_2 + (2g-1)^2 c_3 + (2g-3)^2 c_4 + 2\left(4g^2-1\right)c_5\right)$$

and

$$r = -ga - \frac{a^2\Delta x^2}{384}\left(112g^4 - 56g^2 - 1\right)$$

$$s = \frac{a\Delta x}{12}\left(12g^2 - 1\right)$$

$$t = -\frac{ga\Delta x^2}{12}\left(4g^2 - 1\right)$$

$$u = -\frac{ga\Delta x^2}{12}\left(4g^2 - 1\right)$$

in which we have taken c_6 to c_9 equal to zero, since they turn out to be unnecessary, and it is desirable to have as few off-diagonal correction elements, for efficiency of implementation.

When we equate $a=r$, $b=s$, $d=t$ and $e=u$, Equation (3.24) holds. With these five unknowns, the system is obviously underdetermined with four equations. Since it would be beneficial for the computer code to only have corrections on the interfaces, c_5 is also set to zero. Then, the last two equations ($d=t$ and $e=u$) are equal, so we are still left with an underdetermined system. From here, one can make several choices. The one we found to yield the most simple formulas for the correction terms is the following: The first equation ($a=r$), can be considered to consist of two parts: one that does not contain Δx, and one that contains Δx^2. These parts can be treated as separate equations, so the system again consists of four equations and the four unknowns c_1 to c_4. This is solvable and yields the following correction terms:

$$
\begin{aligned}
c_1 &= \frac{a\left(64g^5 - 16g^4 - 96g^3 + 24g^2 + 36g - 5\right)\Delta x^2}{16\left(4g^2 - 3\right)} \\
c_2 &= -\frac{a\left(64g^5 + 336g^4 - 480g^3 - 184g^2 + 324g - 39\right)\Delta x^2}{16\left(4g^2 - 3\right)} \\
c_3 &= -\frac{a\left(64g^5 - 336g^4 - 480g^3 + 184g^2 + 324g + 39\right)\Delta x^2}{16\left(4g^2 - 3\right)} \\
c_4 &= \frac{a\left(64g^5 + 16g^4 - 96g^3 - 24g^2 + 36g + 5\right)\Delta x^2}{16\left(4g^2 - 3\right)} \\
c_i &= 0, \quad i = 5..9
\end{aligned}
\tag{3.25}
$$

Three important things may be noted about this solution:

- As could be expected, c_4 transforms into c_1 and c_3 into c_2 when we simultaneously replace a and g by $-a$ and $-g$. So, $\tilde{c}_4 = c_1$ and $\tilde{c}_3 = c_2$. This should be the case, as it is dictated by the symmetry of the problem.
- The correction terms are linear in the permittivity contrast, which is contained in the parameter a. The absence of terms of a^2 is beneficial for the case of two interfaces very close together. We have not calculated the correction terms of these very thin layers; this would require lengthy calculations. However, if there were terms containing a^2, the corrections on interfaces that are infinitely close together would add up. With this linear dependence on a, these corrections cancel out, as they should. We have investigated the effect of two interfaces in close proximity numerically (see below, the thin waveguide (Figure 5)).
- The denominator never becomes zero, since $g \in \left[-\frac{1}{2}, \frac{1}{2}\right]$.

Since the only correction terms are again on the diagonal of the matrix, the implemented algorithm can be fast; the off-diagonal elements are equal on each line of the matrix, so these numbers can be hard-coded into the computer program.

For TM polarization (using $y \equiv \frac{1}{n} H_y$ instead of E_y), the same procedure may be followed. Due to the more complicated interface conditions, the error vector f contains more terms; the average values of the zeroth up to the fourth order derivative on the interface are included. The formulas for this error will not be written here; they can be easily calculated, but yield long, complicated expressions. An expression similar to equation (3.24) can also be formulated. However, this expression has nine terms instead of the four that were obtained for TE polarization; the products $\bar{y}^2, \bar{y}\bar{y}', \bar{y}\bar{y}'', \bar{y}\bar{y}''', \bar{y}\bar{y}'''', \bar{y}'^2, \bar{y}'\bar{y}'', \bar{y}'\bar{y}'''$ and $\bar{y}''^2$ are included. Since there are also nine possible correction elements, this system may be solvable. The equation reduces to a matrix equation of the type $Ax=b$. It turns out that the matrix A is singular, which means that the system is either unsolvable or underdetermined. Due to the extremely complicated formulas that make up the elements of the matrix A, an explicit solution

was not found. However, when actual numbers are substituted for the index of refraction on either side of the interface, the equation turns out to be solvable, but underdetermined, so one has some freedom in choosing the correction elements. However, as will be shown below, there is a more straightforward way to tackle the TM case. So, no further effort was put into treating this case along the lines given above.

3.2.3 Results of 4th-order scheme: TE using Taylor expansions

The finite difference operator for the second order derivative has been implemented into the BPM program Menufast, which was developed in previous projects at the Lightwave Devices Group by G.J.M. Krijnen and H.J.W.M. Hoekstra. For the boundaries, simple Hadley transparent boundary conditions were used [Hadley 1991].

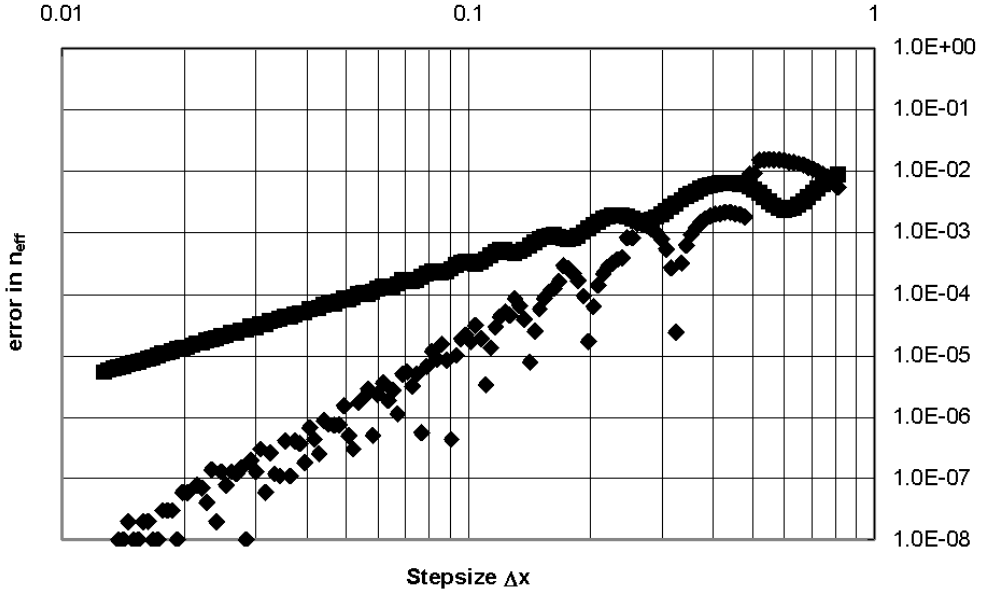


Figure 3.2 : Log-log plot of the error in effective index after 30 μm of propagation through a straight waveguide, for both the fourth-order accurate and the second-order accurate scheme.

A simple test was performed to check the order of convergence. The fundamental TE mode of a 1-μm-wide waveguide was launched. The film refractive index is 2.0, that of the cladding is 1.8 and the substrate is 1.9. The used wavelength is 1.0 μm. The reference refractive index n_0, as used in the SVEA, was chosen equal to the analytical value of the effective index of the mode, i.e. 1.97035688. The stepsize in the propagation direction was chosen to be 1.0 μm. Figure 3.2 shows the error in the effective index of the field after 30 μm of propagation as a function of the stepsize perpendicular to propagation (Δx). This was done for both the 5-point method and the 3-point method [Hoekstra et al 1992[a]]. This effective index is calculated by considering the overlap of the field around this propagation distance with the analytical mode of the waveguide, and numerically determining the derivative of its phase with respect to z. It is clear that with the new method n_{eff} indeed converges with

$O(\Delta x^4)$. Notice that the worst case forms a line of $O(\Delta x^4)$; all points are below this line, except for small errors due to numerical noise.

The 5-point method is, of course, slower than the 3-point method. However, in our current implementation, the speed is only 20% lower than that of the 3-point method. The improved accuracy largely compensates this loss of speed. For example, if we want accuracy of 5 decimals in the effective index for the shown waveguide, we can see from Figure 2 that we need about 5 times less points in the 5-point implementation than in the 3-points implementation. Therefore, the simulation takes 4 times less time (5 times fewer points, 20 % slower algorithm) in the 5-point case than in the 3-point case. A slight increase in computation speed might be possible by using the Douglas scheme [Yamauchi et al 1997, Yevick 1994]. We remark that the presented corrections for TE, i.e., corrections only on the diagonal, may be quite suitable for the Douglas scheme.

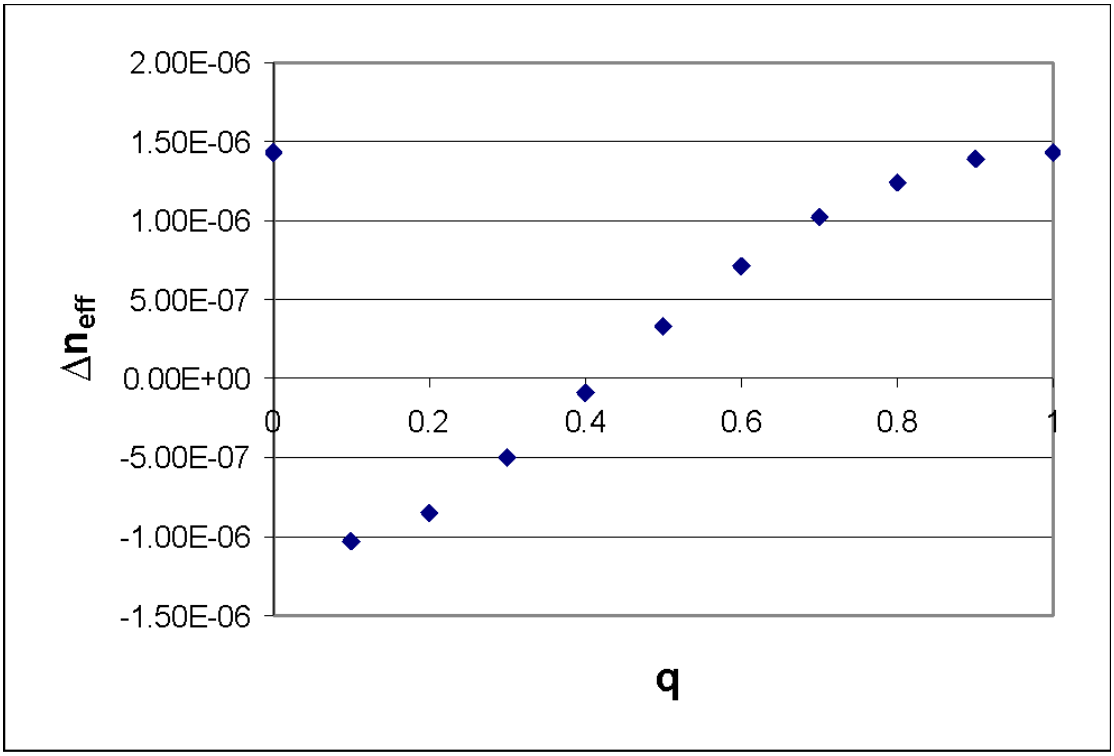


Figure 3.3 : Error in effective index as a function of q, the lateral shift of the structure. q=1 means a shift of one stepsize Δx. The lateral stepsize is 0.05 μm.

To verify the effect of the exact place of the interfaces between two points, the same structure was shifted along the x-direction. The shift was varied between 0 and 50 nm. The stepsize Δx was 50 nm, so the structure was shifted between 0 and 1 stepsize. This shift factor is called q; the shift is 50*q nm. When q is 0 or 1, both interfaces are exactly on a discretization point. It can be seen in Figure 3.3 that the error does not vary much with the place of the interface. The best results can be obtained when the interface is approximately in the middle between two discretization points. This result explains why the convergence curve is not steady; the error passes through zero when an interface is near the middle of two discretization points. Since the absolute value of the error is taken in Figure 3.2, this means that the curve has steep dips.

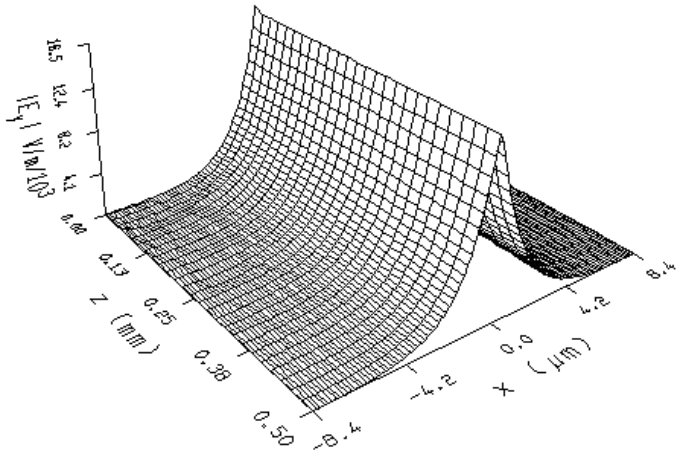


Figure 3.4: Propagation through a waveguide with no discretization points in the waveguide.

Lastly, propagation in a waveguide was calculated, where the discretization is so coarse that there are no points inside the guiding layer. The guiding layer width is 0.1μm, the film refractive index is 2.0, that of the substrate and cladding is 1.9. The discretization stepsize Dx is 0.26 μm, and the entire waveguiding layer is placed between two points. The wavelength is 1.0 μm. Many beam propagation methods cannot simulate propagation in such a case, since the refractive index on the discretization points is equal on each point. Since the exact placement of the interfaces is taken into account in the present method, it is possible to accurately simulate the behaviour of such a structure. The 'exact' modal field was discretized and launched, then propagated for 500 μm with 1 μm steps. As can be seen in Figure 3.4, the shape of the field is preserved quite well. The calculated overlap with the starting field is 99.94%. So, even such a coarse grid still allows for relatively accurate propagation.

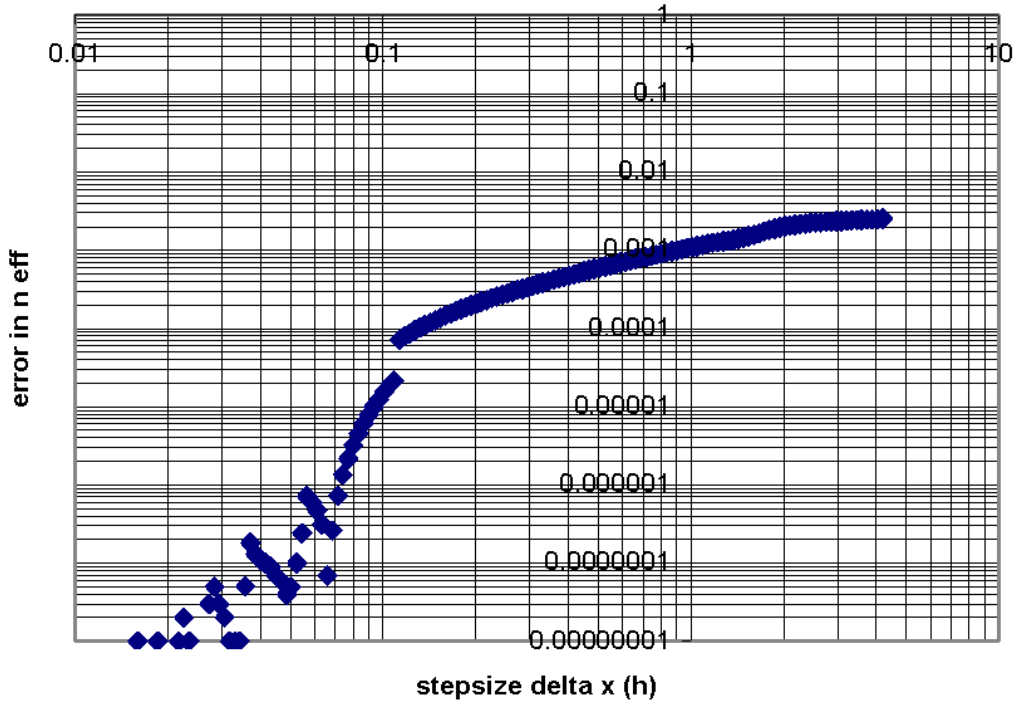


Figure 3.5: Log-log plot of the error in effective index after 500 μm of propagation through the narrow channel. Above Δx=0.11, there are no points in the guiding layer. Note that one can observe new points entering the guiding layer periodically.

Figure 3.5 shows the dependence of the error of the effective index on the stepsize Δx for this particular waveguide structure. Above Δx=0.11, there are no points inside the guiding layer. We see that the error does not increase rapidly with the stepsize; even when the stepsize is 40 times the guiding layer width, the error is still only about 0.002. This means that the error is no longer proportional to Δx^4 when there are no points left in a layer. In this region the error seems to be $O(\Delta x)$, but the error is very small. The latter holds for this structure, but for other structures of this type, i.e. with zero discretization points inside a waveguiding layer, similar results were found.

3.3 Variational formalism

In this section, another method to obtain fourth order accuracy in the unidirectional beam propagation method will be investigated. It is based on a variational formalism for the discretization in the lateral direction, while the propagation step is still performed by the same finite difference (FD) Crank-Nicolson method as described before.

Again, the goal is to simulate the TE and TM Helmholtz equations (3.1). Since the index of refraction is assumed to have a staircase shape, i.e., it is invariant in the propagation direction during each propagation step, the TM equation may be written as:

$$\left(\frac{1}{n^2} \partial_{zz} + \partial_x \frac{1}{n^2} \partial_x + k_0^2 \right) H_y = 0 \tag{3.26}$$

As said before, the main reason why the TM equation is more difficult to solve than the TE equation is the fact that there is a derivative of the (discontinuous) refractive index in the equation, which is related to the more difficult interface conditions on the field. We will show that with the variational formalism one mat get rid of the derivative of the index.

Since the formalism that will be developed here is extremely similar for both TE and TM, we will use the symbol j for both E_y and H_y for the respective polarizations. It will be clear from the text which polarization, and thus which interpretation of j , is meant.

In the transverse direction, a uniform-grid finite element method will be used. This method is based on the minimization of a functional. Since we only use this finite element method in the transverse direction, the functional is an integral over x only. The second order derivative in the propagation direction is treated as a function independently from the field. It is known that for interfaces that are parallel to the propagation direction, the field and its second derivative with respect to z are be subjected to the same interface conditions, and can therefore be treated similarly. In the staircase approximation, the interfaces are parallel to the propagation direction during the propagation step, so this is a valid assumption. So, when approximations of the field are discussed in the following, $\P_{zz} j$ must be treated in the same way. The discretization of this second order derivative utilizing the SVEA and Crank-Nicolson is done after the discretization in the lateral direction has been calculated, so for now, the $\P_{zz} j$ term will be kept in the equations as an independent function.

The following expressions:

$$\begin{aligned} F_{TE} &= \int j \, \partial_{zz} j^{*} - \frac{1}{2} \left(\partial_x j \right)^2 + \frac{1}{2} k_0^2 n^2 j^2 \, dx \\ F_{TM} &= \int \frac{1}{n^2} j \, \partial_{zz} j^{*} - \frac{1}{2n^2} \left(\partial_x j \right)^2 + \frac{1}{2} k_0^2 j^2 \, dx \end{aligned} \tag{3.27}$$

are considered to be functionals of $j(x,z)$, treating $\partial_{zz}j$ as a given independent function. The variational derivative of these functionals is:

$$\begin{aligned} d_j F_{TE} &= \left(\partial_{zz} + \partial_{xx} + k_0^2 n^2\right) j \\ d_j F_{TM} &= \left(\frac{1}{n^2}\partial_{zz} + \partial_x \frac{1}{n^2}\partial_x + k_0^2\right) j \end{aligned} \tag{3.28}$$

So when this variational derivative vanishes, the field j satisfies the correct equations. In the functionals, there is no derivative of the index of refraction anymore. If the integration is done over an interval where the refractive index is constant, the TE and TM functionals differ only by a factor of n^2. This means that, in this approach, TE and TM are very similar: when the TE scheme has been evaluated, going to TM will require only a minor adaptation of the algorithm.

Using the values of the field j on the grid points, $\vec{j}$, and choosing specific spline functions to approximate the function in-between, the functional $F(j)$ can be approximated by a function $\tilde{F}$ such that $F(j) \cong \tilde{F}(\vec{j})$. The condition $d_j F = 0$ then corresponds to:

$$\nabla \tilde{F}(\vec{j}) = 0 \tag{3.29}$$

where ∇ denotes the gradient with respect to the components of $\vec{j}$. $\tilde{F}$ will have the following form:

$$\tilde{F}(\vec{j}) = \sum_{l,m} a_{lm} j_l j_m \tag{3.30}$$

so Equation (3.29) will have this form:

$$\left(\nabla \tilde{F}(\vec{j})\right)_l = \sum_m \left(1 + d_{lm}\right) a_{lm} j_m = 0 \tag{3.31}$$

where $d_{lm}=0$ for $l \neq m$ and 1 for $l=m$. The d_{lm} appears due to the derivative of the square of j_l, i.e. for $l=m$.

3.3.1 Second order approximation

The functionals F must be approximated. Both the field j and its second derivative in the propagation direction are approximated in the same manner. When these approximations are done with an accuracy of a certain order in the transverse stepsize, the accuracy of the resulting scheme will be of the same order (in the correct norm, which will be discussed later). The easiest approximation is by simple linear interpolation between grid points (corresponding to a triangle-shaped basis function).

This approximation is $O(\Delta x^2)$ everywhere. This approximation is shown in Figure 1. From here on, we will use the notation $F_{a,b}^{lin}$ for the integration between points $x=a$ and $x=b$ of the functional using linear approximations. The error that is made in this integration is $O(\Delta x^3)$, but since the total number of integration intervals is $O(\Delta x^{-1})$, the error in the total functional is $O(\Delta x^2)$, the same order as the error in the approximation.

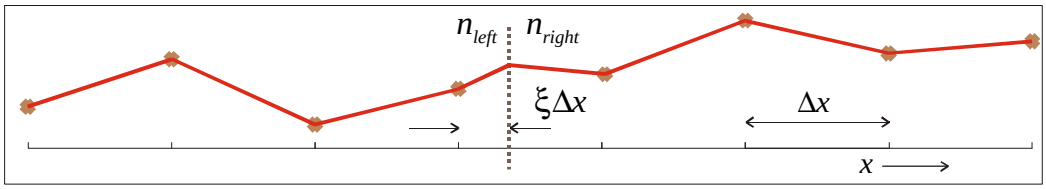


Figure 3.6: Linear interpolation between grid points. At an interface, the integration interval is split in two parts.

When an interface lies between two grid points, the integration over that cell is split into two parts. As shown in Figure 3.6, in both parts a linear interpolation is used. The value at the interface can be determined from the interface conditions (3.12) and (3.13), by making sure the interpolations comply with these interface conditions.

For this case of linear approximation we need only the first two interface conditions. It is simple to construct a value of j and $\partial_{zz} j$ that makes the linear interpolations comply with the interface conditions. Assuming the interface is between grid points $x=0$ and $x=\Delta x$, at the position $x=\xi\Delta x$, the formulas are:

$$\begin{aligned} TE: \quad & j_\xi = j_0(1-\xi) + j_1\xi \\ TM: \quad & j_\xi = \frac{j_0 n_{right}^2(1-\xi) + j_1 n_{left}^2 \xi}{n_{right}^2(1-\xi) + n_{left}^2 \xi} \end{aligned} \tag{3.32}$$

and similarly for $\partial_{zz} j$. In the TE case, the approximation of the field remains a straight line across the interface, since the field must be both continuous and differentiable. For TM, there is a jump in the derivative, as is prescribed by the interface conditions.

By substituting the linear interpolations and equation (3.32) into the functional (3.27) , we are able to approximate the functional up to $O(\Delta x^2)$. The matrix equation that results from (3.29) (before discretization in the propagation direction) is:

$$A^l \partial_{zz} \mathbf{j} = B^l \mathbf{j} \tag{3.33}$$

After this treatment of the discretization in the x-direction, the discretization in the z-direction will be discussed. We apply the SVEA and a Crank-Nicolson scheme. Again, inserting the ansatz $j(x,z) = \psi(x,z) e^{-ik_0 n_0 z}$ into (3.33) and assuming ψ varies slowly with z (i.e. assuming $\partial_{zz}\psi$ can be neglected), one can derive the following propagation equation:

$$\left(A^l - M^l\right)y^{\mathbf{v}_{+1}} = \left(A^l + M^l\right)y^{\mathbf{v}}$$
$$M^l \equiv \frac{i\Delta z}{4k_0 n_0}\left(B^l - k_0^2 n_0^2 A^l\right) \tag{3.34}$$

As proven in Section 3.1, this propagation algorithm conserves the quantity $\left\langle y^{\mathbf{v}} \mid A^l \mid y^{\mathbf{v}} \right\rangle$. This matrix A^l is no longer a constant times the identity matrix, so it must be included in the dot product, and this is the dot product that must be used in the calculation of e.g. modal overlap integrals. Since the matrix A is basically just a distribution of the diagonal of the unity matrix over three diagonals in the TE case, or a distribution of $1/n^2$ in the TM case, this is a good norm for the integrated Poynting vector, which (in the SVEA approximation) is proportional to $\int y^{2} dx$ for TE and $\int \frac{1}{n^2} y^{2} dx$ for TM. In uniform parts of the structure, the matrix A has the following form for the two polarizations:

$$TE: \quad A = \begin{bmatrix} .. & . & . & . & .. \\ . & \frac{2}{3} & \frac{1}{6} & 0 & . \\ . & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & . \\ . & 0 & \frac{1}{6} & \frac{2}{3} & . \\ .. & . & . & . & .. \end{bmatrix}$$
$$TM: \quad A = \begin{bmatrix} .. & . & . & . & .. \\ . & \frac{2}{3n^2} & \frac{1}{6n^2} & 0 & . \\ . & \frac{1}{6n^2} & \frac{2}{3n^2} & \frac{1}{6n^2} & . \\ . & 0 & \frac{1}{6n^2} & \frac{2}{3n^2} & . \\ .. & . & . & . & .. \end{bmatrix} \tag{3.35}$$

We only show the matrices in uniform parts because the matrices near interfaces are much more complicated, and the thing we want to show is only in what way they differ from the unity matrix.

3.3.2 Fourth order approximation

When we want to develop a fourth-order accurate scheme, the way that first comes to mind is by approximating the fields up to fourth order accuracy. This is possible by constructing a cubic polynomial through four grid points and using this polynomial to calculate the integral over one grid cell. However, this procedure causes the resulting matrices A and B to be heptadiagonal (seven non-zero diagonals), since the cubic polynomial is built using four grid points that are thus linked in the scheme; each point is linked to three points on its left-hand side and three points on its right-hand side. It is well-known that there are fourth-order accurate discretizations for e.g. the

second derivative which yield pentadiagonal matrices. So, in order to reduce computation time, we will look for a different way to approximate the functional up to fourth order accuracy in the lateral stepsize, which will yield nearly pentadiagonal matrices. It might be possible to use quadratic interpolating functions, but for simplicity, we choose to utilize combinations of linear interpolations.

3.3.2.1 Uniform refractive index

We will first describe a way to eliminate the $O(\Delta x^3)$ terms in the functional, due to integration over intervals with uniform refractive indices. Interfaces will be treated in Section 3.3.2.2. The error that is made in the three terms of the TE and TM functionals when the field is approximated through linear interpolation is the following:

$$\begin{aligned}
\int_a^b j\,\partial_{zz} j\,dx - \int_a^b j^{lin}\partial_{zz} j^{lin}dx &= -\frac{1}{24}(b-a)^3\left\{j\left(\tfrac{a+b}{2}\right)\partial_{xx}\left(\partial_{zz} j\left(\tfrac{a+b}{2}\right)\right)+\right.\\
&\left.+\partial_{xx} j\left(\tfrac{a+b}{2}\right)\partial_{zz} j\left(\tfrac{a+b}{2}\right)\right\}+\mathrm{O}\left((b-a)^5\right)\\
\int_a^b -\frac{1}{2}\left(\partial_x j\right)^2 dx - \int_a^b -\frac{1}{2}\left(\partial_x j^{lin}\right)^2 dx &= -\frac{1}{24}(b-a)^3\left(\partial_{xx} j\left(\tfrac{a+b}{2}\right)\right)^2\\
&+\mathrm{O}\left((b-a)^5\right)\\
\int_a^b \frac{1}{2} j^2 dx - \int_a^b \frac{1}{2} j^{lin^2} dx &= -\frac{1}{24}(b-a)^3 j\left(\tfrac{a+b}{2}\right)\partial_{xx} j\left(\tfrac{a+b}{2}\right)\\
&+\mathrm{O}\left((b-a)^5\right)
\end{aligned} \tag{3.36}$$

So the total integral between $x=a$ and $x=b$ can be written as:

$$F_{a,b} = F_{a,b}^{lin} + (b-a)^3\, err\left(\tfrac{a+b}{2}\right) + \mathrm{O}\left((b-a)^5\right) \tag{3.37}$$

where $err(x)$ is a function of the fields j and $\partial_{zz} j$ and their second order derivatives with respect to x.

The expression for the error terms proportional to $(b-a)^3$ can be used in Richardson-like extrapolation. This is a technique that combines results from calculations on different grid sizes. The following three integration intervals can be used to describe the process:

$$\begin{aligned}
F_{-\Delta x,0}^{lin} &= F_{-\Delta x,0} - \Delta x^3 err\left(-\frac{\Delta x}{2}\right) + \mathrm{O}\left(\Delta x^5\right) = F_{-\Delta x,0} - \frac{1}{2}\Delta x^3 err(-\Delta x) - \frac{1}{2}\Delta x^3 err(0) + \mathrm{O}\left(\Delta x^5\right)\\
F_{0,\Delta x}^{lin} &= F_{0,\Delta x} - \Delta x^3 err\left(\frac{\Delta x}{2}\right) + \mathrm{O}\left(\Delta x^5\right) = F_{0,\Delta x} - \frac{1}{2}\Delta x^3 err(0) - \frac{1}{2}\Delta x^3 err(\Delta x) + \mathrm{O}\left(\Delta x^5\right)\\
F_{-\Delta x,\Delta x}^{lin} &= F_{-\Delta x,\Delta x} - 8\Delta x^3 err(0) + \mathrm{O}\left(\Delta x^5\right)
\end{aligned} \tag{3.38}$$

The error terms corresponding to integration over the two shorter intervals have been split into two, in such a way that the extra error is $O(\Delta x^5)$ and is thus contained in the higher order terms. The following combination:

$$F^{lin}_{-\Delta x,0} + F^{lin}_{0,\Delta x} - \frac{1}{8} F^{lin}_{-\Delta x,\Delta x} = \frac{7}{8} F_{-\Delta x,\Delta x} - \frac{1}{2}\Delta x^3 \left(err(-\Delta x) + err(\Delta x) \right) + \mathrm{O}\left(\Delta x^5\right) \qquad (3.39)$$

exactly eliminates the error proportional to Δx^3 at $x=0$. This same procedure can be repeated on all points of the fine grid, thereby eliminating all $O(\Delta x^3)$ error terms. It is possible to eliminate the $O(\Delta x^3)$ terms using only two grids, e.g. the grids $[\ldots, -\Delta x, 0, \Delta x, 2\Delta x, 3\Delta x, \ldots]$ and $[\ldots, -\Delta x, \Delta x, 3\Delta x, 5\Delta x, \ldots]$. However, this will cause the resulting matrices to have three non-zero elements on even-numbered rows and five non-zero elements on odd-numbered rows. Therefore, the grid $[\ldots, -2\Delta x, 0, 2\Delta x, 4\Delta x, \ldots]$ will be used as well, so that each row of the matrices will be of the same form. The following combination is the one that is used:

$$F = \frac{4}{3}\sum_i F^{lin}_{i\Delta x,(i+1)\Delta x} - \frac{1}{6}\sum_l F^{lin}_{2l\Delta x,(2l+2)\Delta x} - \frac{1}{6}\sum_l F^{lin}_{(2l-1)\Delta x,(2l+1)\Delta x} + \mathrm{O}\left(\Delta x^4\right) \qquad (3.40)$$

Again, the total error in the functional is $O(\Delta x^4)$, since the error in each integration interval is $O(\Delta x^5)$, and the total number of intervals is $O(\Delta x^{-1})$.

The above only holds in uniform parts of the structure. Near interfaces, more care must be taken with the discretization. Due to the jump in the refractive index, it is not possible to use a Richardson extrapolation scheme near interfaces. Therefore, we will use a different approach.

3.3.2.2 Around interfaces

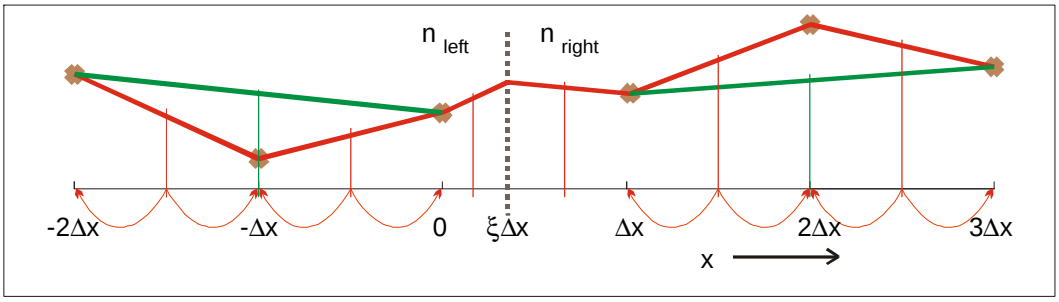


Figure 3.7: Interpolations around an interface for fourth-order accurate discretization. Straight vertical lines denote the 'positions' of err-terms, which may be split, according to (3.38), toward grid points, as denoted by the curved arrows.

In Figure 3.7, the area around an interface is shown. We will again assume that the interface is between $x=0$ and $x=\Delta x$, at $x=\xi\Delta x$. We will have to consider the interval between $-2\Delta x$ and $+3\Delta x$. The following combination of integration intervals is used for the approximation of the functional around the interface:

$$\frac{4}{3}F^{lin}_{-2\Delta x,-\Delta x}+\frac{7}{6}F^{lin}_{-\Delta x,0}-\frac{1}{6}F^{lin}_{-2\Delta x,0}+F^{lin}_{0,\xi\Delta x}+F^{lin}_{\xi\Delta x,\Delta x}-\frac{1}{6}F^{lin}_{\Delta x,3\Delta x}+\frac{7}{6}F^{lin}_{\Delta x,2\Delta x}+\frac{4}{3}F^{lin}_{2\Delta x,3\Delta x}=$$
$$F_{-2\Delta x,3\Delta x}-\Delta x^3\left(\frac{7}{6}err\left(-\tfrac{1}{2}\Delta x\right)+\frac{2}{3}err\left(0\right)+\xi^3 err\left(\tfrac{1}{2}\xi\Delta x\right)+\left(1-\xi\right)^3 err\left(\tfrac{1}{2}\left(1+\xi\right)\Delta x\right)+\right.$$
$$\left.+\frac{2}{3}err\left(\Delta x\right)+\frac{7}{6}err\left(\tfrac{3}{2}\Delta x\right)\right)+\left(\frac{1}{6}F_{-2\Delta x,-\Delta x}+\frac{1}{6}F_{2\Delta x,3\Delta x}-\frac{2}{3}\Delta x^3 err\left(-2\Delta x\right)-\right.$$
$$\left.-\frac{2}{3}\Delta x^3 err\left(3\Delta x\right)\right)+O\left(\Delta x^5\right) \tag{3.41}$$

The terms between the braces on the third line of this equation are eliminated by Richardson extrapolation left and right of the interval that is treated here. The equation shows that there are six additional error terms that must be eliminated in order to approximate the functional over this area up to the correct order. Although it might be enough to approximate the *err* terms up to $O(\Delta x)$ (since there are $O(1)$ interfaces, and $O(\Delta x^{-1})$ grid cells), we choose to approximate those terms up to order $O(\Delta x^2)$ in order to get the same order of accuracy near interfaces as in uniform parts of the structure. Since the *err* terms contain second order derivatives, it is necessary to employ a cubic polynomial to do this. So, cubic polynomials are constructed for the fields j and $\P_{zz}j$.

The field on the left-hand and right-hand side of the interface is approximated by cubic polynomials, whose coefficients are different, but related to one another by the interface conditions. The polynomials should be equal to the field on the four grid points surrounding the interface. The easiest way to construct these polynomials is by translating the x-axis such that the interface lies at the origin x=0. Say the polynomials on the left-hand and right-hand side are:

$$\begin{aligned} j_{left}(x) &= a_l x^3 + b_l x^2 + c_l x + d_l \\ j_{right}(x) &= a_r x^3 + b_r x^2 + c_r x + d_r \end{aligned} \tag{3.42}$$

One can then use the interface conditions (3.12) to relate the coefficients on the right-hand side to those on the left-hand side. As an example, for TE, we can derive:

$$\begin{aligned} & j_{left}(0) = j_{right}(0) \quad \Leftrightarrow \quad d_r = d_l \\ & j'_{left}(0) = j'_{right}(0) \quad \Leftrightarrow \quad c_r = c_l \\ & j''_{left}(0) + k_0^2 n_{left}^2 j_{left}(0) = j''_{right}(0) + k_0^2 n_{right}^2 j_{right}(0) \quad \Leftrightarrow \quad b_r = b_l + \tfrac{1}{2} k_0^2 \left(n_{left}^2 - n_{right}^2 \right) d_l \\ & j'''_{left}(0) + k_0^2 n_{left}^2 j'_{left}(0) = j'''_{right}(0) + k_0^2 n_{right}^2 j'_{right}(0) \quad \Leftrightarrow \quad a_r = a_l + \tfrac{1}{6} k_0^2 \left(n_{left}^2 - n_{right}^2 \right) c_l \end{aligned} \tag{3.43}$$

and similar for TM. In this way, the field on the four grid points surrounding the interface can be expressed in the four coefficients a_l to d_l. This system of equations can be inverted to express a_l to d_l in terms of the field on these four grid points. Then, the value of the polynom and its derivatives is known on each x-position, expressed in the field on the four grid points. The polynoms are used for both the $err(x)$ terms in equation (3.41), and for the determination of the fields on the interface, since they are needed in the integration intervals directly adjacent to the interface. Thus, all terms of the approximated functional are expressed in products of the field on grid points, just like the rest of the approximated functional. The only difference is that while in uniform space, only products of fields at points located at most $2\Delta x$ are present, near the interface, there are also terms containing the product of fields at points $3\Delta x$ apart. This leads to two extra nonzero elements per interface in the matrix equations, located outside of the normally pentadiagonal band.

So, since the coefficients of the polynomial contain the field on each of the four points around the interface, the inclusion of the *err* terms causes the matrices that result from equation (3.29) to have nonzero elements in a block of 4x4 matrix elements around the interface; this means that two extra elements of the matrices just outside the normal pentadiagonal band to become non-zero. These two elements occur at transposed positions and are identical, not disturbing the symmetry of the matrices. We still can utilize a standard pentadiagonal matrix equation solver if these extra elements are dealt with separately, by means of a simple transformation before and after the solving step, which involves simple row- and column operations. Therefore, the fact that the matrices are not completely pentadiagonal does not cause the calculation time to increase significantly.

When the polynomial is constructed and the extra error terms are eliminated, the total error around an interface is of the same order as the error that is made in uniform parts of the structure. So, the total functional is approximated up to order $O(\Delta x^4)$. In the correct norm, the results of the scheme (integrated quantities) will thus converge as $O(\Delta x^4)$, as was intended.

The matrix equation that results from the scheme has the same form as equation (3.34). The A and B matrices still have the same meaning as in the second-order case. Since there are no integration intervals across an interface, TM propagation is almost

identical to TE propagation. The only things that are different between TE and TM polarization are the $F_{a,b}^{lin}$ terms, which must be divided by n^2, and the cubic polynomial, which must comply with different interface conditions.

3.3.2.3 Boundary conditions

Two types of boundary conditions were implemented: Hadley's standard Transparent Boundary Condition (TBC) [Hadley 1991] and Perfectly Matched Layers (PML's) [Bérenger 1994], the basics of which were described in the previous chapter. Both are very easy to implement in this variational method.

By the PML's, light reaching the boundaries along x of the computational domain should be absorbed. Therefore, only s_x needs to be nonzero in the PML's. The Helmholtz equation for TE changes into:

$$\frac{1}{1-is_x(x)}\partial_x\frac{1}{1-is_x(x)}\partial_x y + \partial_{zz} y + k_0^2 n^2 y = 0 \tag{3.44}$$

The corresponding functional is:

$$F = \int \left(1-is_x(x)\right) y\,\partial_{zz} y - \frac{1}{2\left(1-is_x(x)\right)}\left(\partial_x y\right)^2 + \frac{\left(1-is_x(x)\right)}{2} k_0^2 n^2 y^2 dx \tag{3.45}$$

And similar for TM.

As shown in Chapter 2, a solution to Equation (3.44) in areas where s_x equals zero is the same as a solution to the normal Helmholtz equation; in the layer where s_x has a non-zero value, both incoming and outgoing waves are attenuated at a rate determined by s_x. For analytical solutions, there is no reflection on the perfectly matched layer; the only reflections that occur analytically are reflections on the outer boundary (where we take Dirichlet boundary conditions). However, there are numerical reflections when s_x changes abruptly. Therefore, s_x must increase smoothly in the PML. A good choice, which has been used here, is a quadratic shape [Vasallo 1996]. Since PML's are non-adaptive, the reflection coefficient of a plane wave does not change when other plane waves are also incident on the boundary; it is a linear method. Therefore, this method is more robust than Hadley's TBC.

If s_x is kept constant over an integration interval, the functional over that interval can very easily be calculated. So, if a step function with a quadratic envelope is used, PML's are easy to implement. It might be simple to not use a step function for s_x but use its analytical form in the calculation of the functionals, but we have not investigated this further.

3.3.3 Results

In this section, some results of the unidirectional BPM with variational derivation of the discretization will be presented. Only results for the TM algorithm are shown.

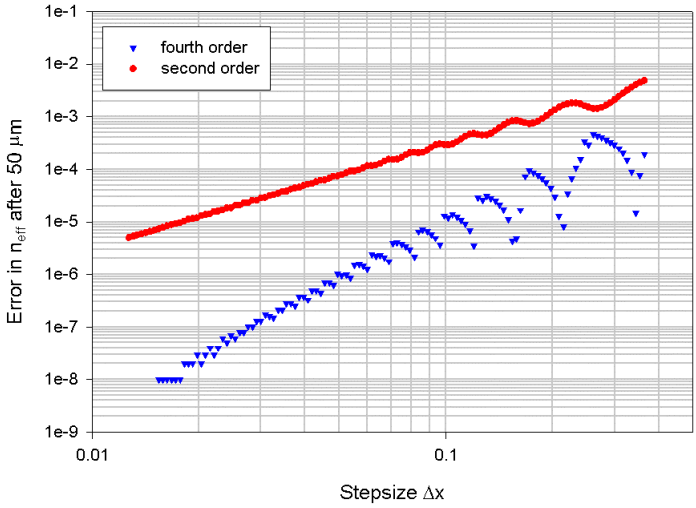


Figure 3.8: Log-log plot of the error in effective index after 30 µm of propagation through a straight waveguide, for both the fourth-order accurate and the second-order accurate variational scheme.

First, we consider propagation through a straight waveguide. The wavelength of the light is 1.0 µm, the width of the guide is 1.0 µm and the refractive indices of cladding, core and substrate are resp. 1.8, 2.0 and 1.9. A TM_0 mode is launched into this waveguide and propagated for 50 µm, with 1-µm steps in the propagation direction. The reference refractive index that is used in the SVEA approximation is chosen equal to the effective index of the TM_0 mode. After these 50 µm of propagation, the error that is made in the effective index of the zeroth-order mode is calculated and plotted against the lateral stepsize in Figure 3. Both the second-order and fourth-order schemes are used and plotted in the same graph. It is clear that the order of convergence is indeed as expected. The steep dips in the fourth-order plot are caused by the fact that the absolute value of the error is plotted, and, similar to what was observed in Section 3.2.3, the actual error may cross zero as the positions of the interfaces move through a grid cell – though this only seems to occur at higher stepsizes.

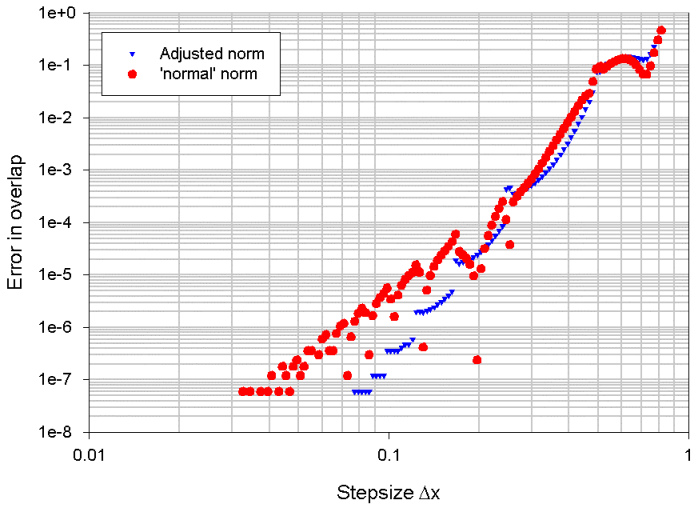


Figure 3.9: Overlap with analytical TM_0 mode, using both the original norm (from non-variational scheme) and the adjusted, correct norm.

When a certain scheme is used, the norm used to calculate energies and overlaps should correspond to the same scheme. If the scheme is correct up to fourth order, the effective index of a mode should behave as $O(\Delta x^4)$, while at the same time, the overlap calculated with the norm that follows from the scheme should be $O(\Delta x^8)$. This can be seen from the following reasoning, which will be done for a system in which one mode beside the correct one is excited; it can be easily extended to any number of modes.

A mode of the normalized numerical system f_0^n is different from the analytical mode f_0, due to the numerical errors. The error in the modes will be of the same order as the scheme, in this case $O(\Delta x^4)$. Say the error is purely in the first order numerical mode, so the analytical mode f_0 can be written as $f_0 = a f_0^n + b\Delta x^4 f_1^n$. In the numerical scheme, the correct norm $\langle f|f\rangle$ is conserved in these straight waveguides. Also, the numerical modes are orthogonal in this norm, so $\langle f_0^n|f_1^n\rangle = 0$. So, when the analytical mode f_0 is launched at the beginning of the window, the overlap (defined as $|\langle f|f_0^n\rangle|^2$) will be a^2 everywhere in the window. When this mode is normalized such that $\langle f_0|f_0\rangle$ is equal to 1, it can be seen that:

$$\langle\phi_0|\phi_0\rangle = a^2\langle\phi_0^n|\phi_0^n\rangle + b^2\Delta x^8\langle\phi_1^n|\phi_1^n\rangle + ab\Delta x^4(\langle\phi_0^n|\phi_1^n\rangle + \langle\phi_1^n|\phi_0^n\rangle) = 1$$

which, due to the orthogonality of modes and the fact that the numerical modes are normalized, means that $a^2 = 1 - \Delta x^8 b^2$. So, the error in the overlap will be $O(\Delta x^8)$.

This overlap is investigated in Figure 3.9. Both the correct norm that follows from the variational scheme, $\langle f \mid A^l \mid f \rangle$, and the norm that follows from the non-variational scheme, $\langle f/n \mid f/n \rangle$, are calculated. This overlap should converge to 1, so we plot the quantities (1-overlap). The correct norm gives $O(\Delta x^8)$ convergence, as expected.

However, the other norm clearly becomes much worse when the stepsize becomes small. The order of convergence for the 'old' norm is found to be approximately $O(\Delta x^4)$.

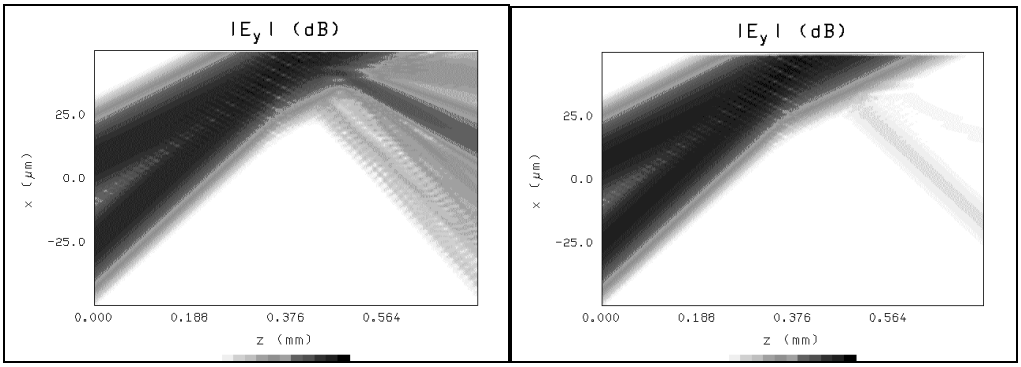


Figure 3.10: Reflection from the boundary of the calculation window with Hadley's TBC (left) and PML's (right), for two incoming Gaussian beams. The power reflectivity is 1.7% for the TBC, and 0.0014% for the PML.

Finally, the boundary conditions are tested. In a uniform structure with refractive index 3.3 we launch two 15-μm wide Gaussian beams at angles of 5.7 and 11.5° [Song 1993]. The wavelength is 0.828 μm. At *z=0*, the centers of the beams are located at resp. *x=-7* and *x=+25* μm. The beams are propagated over 750 μm, using the fourth order TE scheme. In Figure 3.10, the fields are plotted on a logarithmic scale, in order to make the reflections more visible. In the left-hand picture, Hadley's TBC is used; in the right-hand one, PML's. In this particular example, the PML performs much better than TBC; for TBC, the power reflectance is approximately 1.7%, while it is only 0.0014% for the PML.

3.4 Conclusions

Two different methods have been developed to obtain fourth-order accurate unidirectional beam propagation methods. Interfaces between materials at arbitrary positions within grid cells are taken into account, so there are no restrictions on the discretization grid.

The first method applies Taylor decompositions of the fields around interfaces to obtain corrections on the discretization of the SVEA Helmholtz equation. It yields very simple formulas for TE polarization, in which the only correction terms are on the diagonal of the matrices, which allows for the implementation in a computer program to work very fast. For TM polarization, the formulas for the corrections become very complex; they have not been explicitly shown, since the second method gives a much more elegant way to derive the correction terms.

In the second method, a finite difference scheme is used in the propagation direction and a finite element method in the transverse direction. This finite element approach causes the implementation for TE and TM to be very similar. Both second-order accurate and fourth-order accurate methods are shown. For the calculation of integrated quantities, which is what one is generally interested in, a different norm from the standard dot product is needed in order to achieve modal power conservation.

The error made in the effective index converges as $O(\Delta x^2)$ or $O(\Delta x^4)$ for the 3- and 5-point operator, respectively. The increased accuracy of the fourth-order scheme decreases the amount of grid points that are needed, so the total computation time and memory usage can be much smaller at a given accuracy. For the boundary of the computation window, either Hadley Transparent Boundary Conditions (TBC) or Perfectly Matched Layers (PML's) are used. As expected from literature [Vasallo 1996], Perfectly Matched Layers perform better. The speed of the fourth order method is approximately 50% of a standard, second order method [Hoekstra et al 1992]. The memory consumption is approximately 70% higher.

We expect that similar techniques can be applied to 3D BPM methods. Of course, a vectorial approach would be needed, with a vectorial functional. Due to the fact that this functional is used, singularities near corners of waveguides will not cause problems, since the integral over an area containing the singularity should still be finite for physical reasons.

Chapter 4: Omnidirectional simulation tools

In the previous chapter, highly accurate simulation methods for unidirectional propagation were developed. In these methods, some assumptions were made in order to transform the Helmholtz equation, which is a boundary value problem, into an initial value problem. This has the great advantage that the simulations can be performed fast, since no information travels in a direction opposite to the main propagation direction. This means that with these algorithms one cannot do calculations on a structure in which the fact that the the forward and backward-traveling beams are coupled may not be neglected. Also, as a consequence of the assumptions made, light traveling at a large angle with respect to the main propagation direction is, in general, not treated properly. However, there are many real-life problems in which there are considerable reflections of light back in the reverse propagation direction, or straight up or down; examples are a cylindrical microresonator or any photonic crystal structure; see Chapter 5 for these examples. These structures are used more and more extensively in photonic integrated circuits for e.g. wavelength division multiplexing or passband flattening, so one needs a simulation tool that can take into account this propagation in all directions: omnidirectional simulation [Boisvert 1987, Yee 1966, Schmidt 1998, Taflove 1995].

This chapter will first describe a new solver for the 2-dimensional Helmholtz equation, based on a complete variational formalism, which utilizes the Perfectly Matched Layer (PML) boundary condition [Bérenger 1994] that was described in Chapter 2 for the generation of incoming fields. Next, it will describe a Finite Difference Time Domain method [Taflove 1995] with the same type of boundary condition and startfield generation. The startfield generation will be generalized to the ability to launch temporal modal pulses; i.e. pulses in a waveguide for which each frequency has its own modal field profile.

4.1 Helmholtz solver

This section will describe a two-dimensional solver for the Helmholtz equation. First, the constituent equations will be shown and the boundary conditions discussed; next, the equations will be discretized using the variational formalism on a rectangular grid. Lastly, the solver for the matrix equation that results from the discretization will be shown, as well as the results from some simple examples.

4.1.1 Constituent equations

The Helmholtz equation for TE or TM:

$$\begin{aligned} TE:&\quad \left(\partial_{xx}+\partial_{zz}+k_0^2 n^2\right)E_y = 0 \\ TM:&\quad \left(\partial_x\left(\frac{1}{n^2}\partial_x\right)+\partial_z\left(\frac{1}{n^2}\partial_z\right)+k_0^2\right)H_y = 0 \end{aligned} \tag{4.1}$$

should be solved on a rectangular domain in the *x*-*z* plane. The angular frequency of the light ω is fixed and lumped into the parameter $k_0=\omega/c$, with *c* the speed of light; a time dependence of $\exp(i\omega t)$ is assumed. The index of refraction *n* is a piecewise constant function of *x* and *z*. The boundary should behave as if it were not there, so e.g. waveguides that cross the boundary should guide a mode out of the system without reflections.

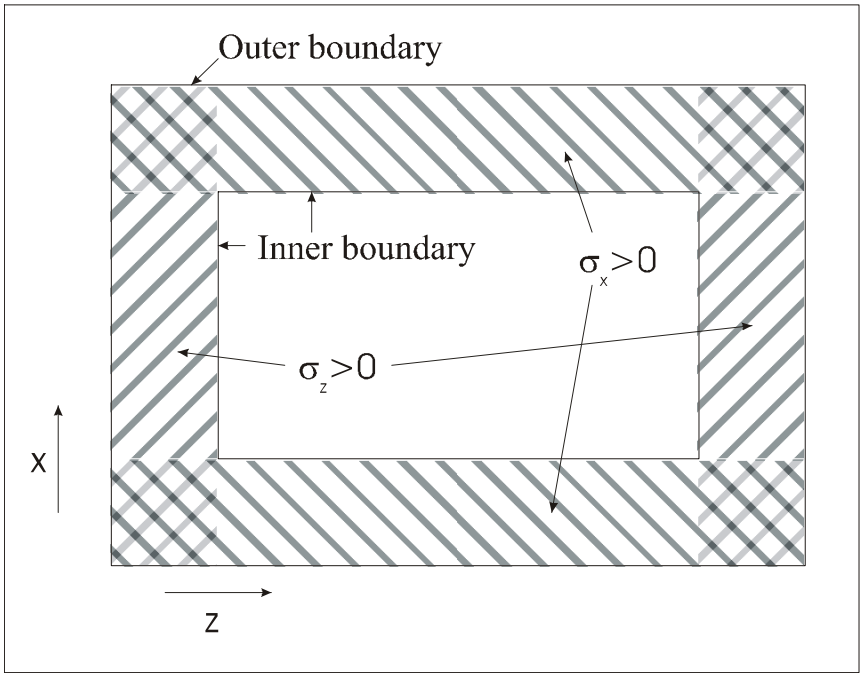


Figure 4.1: The calculation window, surrounded by Perfectly Matched Layers (PML). The outer boundary is the actual boundary of the window; the inner boundary is the place where the PML's begin.

Since it was shown in the previous chapter that the variational formalism can be a very useful method, in which it is relatively simple to accurately take into account the effect of interfaces between materials, the same formalism will be used here. However, while the variational formalism was only used in the direction perpendicular to the propagation direction in the previous chapter, we will use it for the complete discretization in this section. Since light must be able to propagate in any direction, no distinction between a 'propagation direction' and a 'transverse direction' may be made. So, in the same way as in the previous chapter, a functional can be constructed whose extremal points correspond to solutions of the Helmholtz equation:

$$\begin{aligned} F_{TE}\left(E_y\right) &= \iint\left(-\frac{1}{2}\left(\partial_x E_y\right)^2-\frac{1}{2}\left(\partial_z E_y\right)^2+\frac{1}{2}k_0^2 n^2 E_y^2\right)dxdz \\ F_{TM}\left(H_y\right) &= \iint\left(-\frac{1}{2n^2}\left(\partial_x H_y\right)^2-\frac{1}{2n^2}\left(\partial_z H_y\right)^2+\frac{1}{2}k_0^2 H_y^2\right)dxdz \end{aligned} \tag{4.2}$$

since

$$
\begin{aligned}
d_{E_y} F_{TE} &= \left(\partial_{xx} + \partial_{zz} + k_0^2 n^2\right) E_y \\
d_{H_y} F_{TM} &= \left(\partial_x \left(\frac{1}{n^2}\partial_x\right) + \partial_z \left(\frac{1}{n^2}\partial_z\right) + k_0^2\right) H_y
\end{aligned}
\tag{4.3}
$$

At the edges, outgoing light must again be transmitted or absorbed by the boundary. The Perfectly Matched Layer was very easy to include in the variational part of the previous chapter, so the same is expected for the omnidirectional solver. A layer of PML material is included around the whole window, as shown in Figure 4.1. On each boundary the light should be absorbed in the direction perpendicular to the boundary, so s_x is nonzero near the upper and lower boundary, and s_z is nonzero near the left and right boundary. Near the corners, they necessarily overlap. In Chapter 2, the PML for absorption in one direction was derived, yielding equation (2.23). With PML's for both *x*- and *z*- direction, the Helmholtz equations become:

$$
\begin{aligned}
TE: &\quad \left(\frac{1}{1-is_x}\partial_x\left(\frac{1}{1-is_x}\partial_x\right) + \frac{1}{1-is_z}\partial_z\left(\frac{1}{1-is_z}\partial_z\right) + k_0^2 n^2\right) E_y = 0 \\
TM: &\quad \left(\frac{1}{1-is_x}\partial_x\left(\frac{1}{n^2(1-is_x)}\partial_x\right) + \frac{1}{1-is_z}\partial_z\left(\frac{1}{n^2(1-is_z)}\partial_z\right) + k_0^2\right) H_y = 0
\end{aligned}
\tag{4.4}
$$

The functionals (4.2) then change into:

$$
\begin{aligned}
F_{TE}(E_y) = \int\Bigg(&-\frac{(1-is_z)}{2(1-is_x)}\left(\partial_x E_y\right)^2 - \frac{(1-is_x)}{2(1-is_z)}\left(\partial_z E_y\right)^2 + \\
&+\frac{1}{2}k_0^2 n^2 (1-is_x)(1-is_z) E_y^2\Bigg) dxdz \\
F_{TM}(H_y) = \int\Bigg(&-\frac{(1-is_z)}{2n^2(1-is_x)}\left(\partial_x H_y\right)^2 - \frac{(1-is_x)}{2n^2(1-is_z)}\left(\partial_z H_y\right)^2 + \\
&+\frac{1}{2}k_0^2 (1-is_x)(1-is_z) H_y^2\Bigg) dxdz
\end{aligned}
\tag{4.5}
$$

4.1.2 Boundary conditions

Since the Helmholtz equation is a boundary value problem, the only way to prescribe information about incoming fields is by setting the field on the outer boundary of the calculation window. However, from just the field on the boundary it is not possible to distinguish between incoming and outgoing fields. To solve this problem, PML's again turn out to be very useful. We will explain this by considering the analytical solution of a one-dimensional example.

Take the one-dimensional version of the TE part of equation (4.4):

$$\left(\frac{1}{1-is_x}\partial_x\left(\frac{1}{1-is_x}\partial_x\right)+k_0^2n^2\right)E_y=0 \tag{4.6}$$

The solutions of this equation is:

$$E_y=E_0^+e^{i\omega t}e^{+ik_0nx}e^{+k_0n\int_0^x s_x dx'}+E_0^-e^{i\omega t}e^{-ik_0nx}e^{-k_0n\int_0^x s_x dx'} \tag{4.7}$$

corresponding to a combination of a wave traveling to the right (minus sign in the solution) and to the left (plus sign in the solution), and whose amplitudes decay in the propagation direction if s_x is positive.

Figure 4.2: One-dimensional calculation window. PML's are located in the layers of thickness d on the left-hand and right-hand side of the window.

Now consider a finite window of length L, $x\in[0,L]$, as shown in Figure 4.2. The index of refraction is constant throughout the window. On both the left-hand and right-hand side of the window, there is a layer of width d of PML. In this layer, s_x is chosen to be equal to $\ln(50)/(k_0nd)$, which means that the amplitude of light traversing the PML is decreased by a factor of 50. For $x\in[d,L-d]$, s_x is zero, which means that in this region, the solutions behave as normal traveling waves.

On the outer boundary, $x=0$ and $x=L$, the field is set at a certain value; these are Dirichlet boundary conditions. Suppose the desired field is a right-traveling incoming wave coming from the left with amplitude 1, and no left-traveling wave. Then, the (time-independent part of the) field at $x=0$ must be 50, and at $x=L$ must be $\exp(-ik_0nL)/50$, since these boundary values are consistent with only a right-traveling wave in the whole window. Since the value on the right-hand boundary is relatively small, it turns out not to matter much if it is put to zero. When the field is 50 at $x=0$ and 0 at $x=L$, the coefficients of the left- and right-traveling waves can be calculated:

$$E_0^+ + E_0^- = 50$$

$$2500E_0^+e^{ik_0nL}+\frac{1}{2500}E_0^-e^{-ik_0nL}=0$$

$$\Leftrightarrow$$

$$E_0^+=\frac{50}{1+2500^2e^{2ik_0nL}}$$

$$E_0^-=\frac{50}{1+2500^{-2}e^{-2ik_0nL}}$$

As a consequence of setting the field on the right-hand boundary to zero, the left-traveling wave is excited by a small amount; the ratio of the amplitudes of the right-traveling and the left-traveling wave in the interior of the window is about 2500. This means that setting the value of the field on the left-hand boundary very accurately creates a right-traveling wave. Similarly, the field on the right-hand boundary defines the left-traveling wave.

This principle can be extended to 2 dimensions. The window is surrounded by PML's as was shown in Figure 4.1. On all the boundary points where one does not intend to excite incoming waves, the field should be set to zero. This will act similarly to the right-hand boundary of the one-dimensional example; waves impinging on the PML will generate a negligible amount of reflection back into the window. The incoming field can be defined on the corresponding boundary points, taking into account the effect of the attenuation in the PML's in a correct way.

In most applications in integrated optics, one is most interested in the behaviour of modes coming in through a waveguide. As shown in Chapter 2, a mode of a waveguide that traverses a layer of PML material will be attenuated along the waveguide without its shape being distorted. So, launching a mode into a waveguide is simple; one sets the shape of the field on the outer boundary to the desired modal field. The amplitude of the field on the outer boundary must be amplified by the known amount of attenuation that the mode will experience during its journey through the PML, in order to have the desired amplitude at the inner boundary. For example, in the 1-dimensional system described above, the field on the left boundary had to be 50 in order to have an amplitude of 1 on the inner boundary. In general, one knows the propagation constant, b, of the mode that one wants to excite, so the amplification constant is known from equation (2.23).

Often, one wants to not launch just one single mode, but more than one or a larger set of modes; it may be that the two lowest-order modes of a waveguide must be launched simultaneously, or that a Gaussian beam, which is a good approximation of the field at the exit of an optical fiber, should be launched into the window. In the case of a Gaussian beam, the field should be decomposed into the complete set of modes of the structure; so this can also be considered as a set of modes that should be excited. The computational scheme is completely linear, so each of these modes can be considered separately. Each mode has its own propagation constant b, and therefore its own decay constant in the PML. Therefore, each mode must be amplified by a different factor. For example, if we have two different modes with effective indices $n_{eff,1}$=1.9 and $n_{eff,2}$=1.5 for light with a vacuum wave number of k_0, the amplification factors due to a PML of width d and 'strength' s =ln(50)/$k_0 d$ should be:

$$A_1 = e^{k_0 n_{eff,1} s\, d} = e^{n_{eff,1} \ln(50)} = 50^{n_{eff,1}} = 1691$$
$$A_2 = 50^{n_{eff,2}} = 354$$

In the results shown in Section 4.1.5, it will be shown that it is indeed imperative to take these different amplification factors into account, especially when multiple modes with widely differing propagation constants are involved.

4.1.3 Discretization

In order to solve the Helmholtz equation for an arbitrary refractive index distribution, the field, the equations must be discretized. A grid is generated. On the grid points, the field is calculated. Since the variational approach uses integrals over the whole of space (or at least, over the whole calculation window), the field at all positions within grid cells must also be known. This field is interpolated from the grid points surrounding it. In this way, the field everywhere can be expressed in the field on the grid points. So, we have a vector of fields on grid points $\vec{y}$, from which the discretized analogue of the functional F (for either TE or TM), denoted by $\tilde{F}$, can be calculated using the interpolations. So $\tilde{F}$ may be written as:

$$\tilde{F} = \int g(x,z)\,dxdz \tag{4.8}$$

Here, $g(x,z)$ is defined everywhere in terms of the discretized field and the index distribution.

Since solving the equation corresponds to finding extremal points of the functional F, solving the discretized equation corresponds to finding extremal points of $\tilde{F}$, which means that the gradient of $\tilde{F}$ should vanish:

$$\nabla_{\vec{y}} \tilde{F}(\vec{y}) = 0 \tag{4.9}$$

in which the gradient means the vector of derivatives of $\tilde{F}$ with respect to each non-boundary element of $\vec{y}$, since the value on the boundary of the calculation window is not variable but fixed, and the numeric value at boundary points should be substituted in the equations.. We will call the vector of non-boundary (or internal) points $\vec{y}_{\text{int}}$. This yields a system of linear equations in the elements of $\vec{y}_{\text{int}}$. Then, equation (4.9) reduces to a matrix equation of the type

$$\mathbf{A}\vec{y}_{\text{int}} = \vec{b} \tag{4.10}$$

The matrix **A** contains the discretized operators of the Helmholtz equation, and the vector $\vec{b}$ contains information about the incoming fields.

Up to here, no assumptions on the actual grid were made. One may make several choices for the grid; e.g. a uniform, rectangular grid that is easy to set up, or a nonuniform grid that is relatively difficult to generate, but that can be adapted to the index distribution; for example, it can be set up such that the index, which is presumed to be piecewise constant on the whole domain, is constant within each grid cell. We choose to use the former method. Then, it is still possible to take into account the index distribution on a finer scale than just the grid itself, as will be shown below.

The functional $\tilde{F}$ must be calculated using the field on the grid points $\vec{y}$ and the index distribution. We choose to use a rectangular grid, in which each grid cell is also rectangular. Another choice is to use triangular cells, while taking care to not

introduce dependence on direction. However, the book-keeping involved with that is somewhat more complex.

In each grid cell for which the index of refraction is constant, $\hat{F}$ is calculated using bilinear interpolation for the fields; cells with interfaces between materials inside them will be discussed later. Say we have a rectangle with the field at the four corner points y_{00}, y_{01}, y_{10} and y_{11}, where y_{ij} is the field at the position $x=i\Delta x$ and $z=j\Delta z$. The formula for bilinear interpolation is:

$$y(x,z) = ax + bz + cxz + d \tag{4.11}$$

This bilinear interpolation must be equal to the field at the corner points, so the coefficients *a, b, c* and *d* can be expressed in the fields y_{ij} as follows:

$$\begin{aligned} a &= \frac{y_{10} - y_{00}}{\Delta x} \\ b &= \frac{y_{01} - y_{00}}{\Delta z} \\ c &= \frac{y_{00} + y_{11} - y_{10} - y_{01}}{\Delta x \Delta z} \\ d &= y_{00} \end{aligned} \tag{4.12}$$

From this interpolation, the derivatives in the functionals can also be defined in each grid cell. With these coefficients substituted into equation (4.11), the integral that must be calculated to obtain $\hat{F}$ can easily be written out explicitly.

To obtain the matrix **A**, the derivative of the approximated functional with respect to the field on each grid point must be taken. When a point is surrounded by grid cells that all have the same index of refraction, only the surrounding four grid cells will contribute to the derivative with respect to that point. By calculating these four integrals, one can calculate all non-zero elements of one row of the matrix **A**. A more understandable way of describing a row of the matrix is by not writing down the row itself, but the so-called stencil, which reflects the layout of the grid points: Each neighbouring point has a weight in the derivative, as follows:

$$\begin{aligned} \frac{\partial \hat{F}(\mathbf{y})}{\partial y_{ij}} = \; & k y_{(i-1)(j-1)} + l y_{(i-1)j} + m y_{(i-1)(j+1)} + \\ & n y_{i(j-1)} + p y_{ij} + q y_{i(j+1)} + \\ & r y_{(i-1)(j-1)} + s y_{(i-1)j} + t y_{(i-1)(j+1)} \end{aligned}$$

$$\Rightarrow stencil: \quad \begin{vmatrix} k & l & m \\ n & p & q \\ r & s & t \end{vmatrix}$$

The stencil for the TE and TM equations (4.1) in uniform space with these bilinear interpolations can be built from three parts, which are directly related to the three terms in the functionals. The stencils are:

$$TE: \frac{1}{6\Delta x^2}\begin{bmatrix}1 & 4 & 1\\ -2 & -8 & -2\\ 1 & 4 & 1\end{bmatrix} + \frac{1}{6\Delta z^2}\begin{bmatrix}1 & -2 & 1\\ 4 & -8 & 4\\ 1 & -2 & 1\end{bmatrix} + \frac{k_0^2 n^2}{36}\begin{bmatrix}1 & 4 & 1\\ 4 & 16 & 4\\ 1 & 4 & 1\end{bmatrix} \qquad (4.13)$$

and

$$TM: \frac{1}{6n^2\Delta x^2}\begin{bmatrix}1 & 4 & 1\\ -2 & -8 & -2\\ 1 & 4 & 1\end{bmatrix} + \frac{1}{6n^2\Delta z^2}\begin{bmatrix}1 & -2 & 1\\ 4 & -8 & 4\\ 1 & -2 & 1\end{bmatrix} + \frac{k_0^2}{36}\begin{bmatrix}1 & 4 & 1\\ 4 & 16 & 4\\ 1 & 4 & 1\end{bmatrix} \qquad (4.14)$$

If one or more of the adjoining grid cells has a different index of refraction, the affected quadrant of the stencil changes to reflect this.

If an interface between materials crosses a grid cell, this can easily be taken into account for TE polarization, as will be shown here. We will distinguish between two different cases. Either any interfaces crossing the cell are straight and enter the cell from one boundary and leave it through another, without crossing; or, the cell contains one or more corner points, at which interfaces meet. The two cases are distinguished because for TM polarization, singularities may occur at corner points. Examples of these two cases are shown in Figure 4.3.

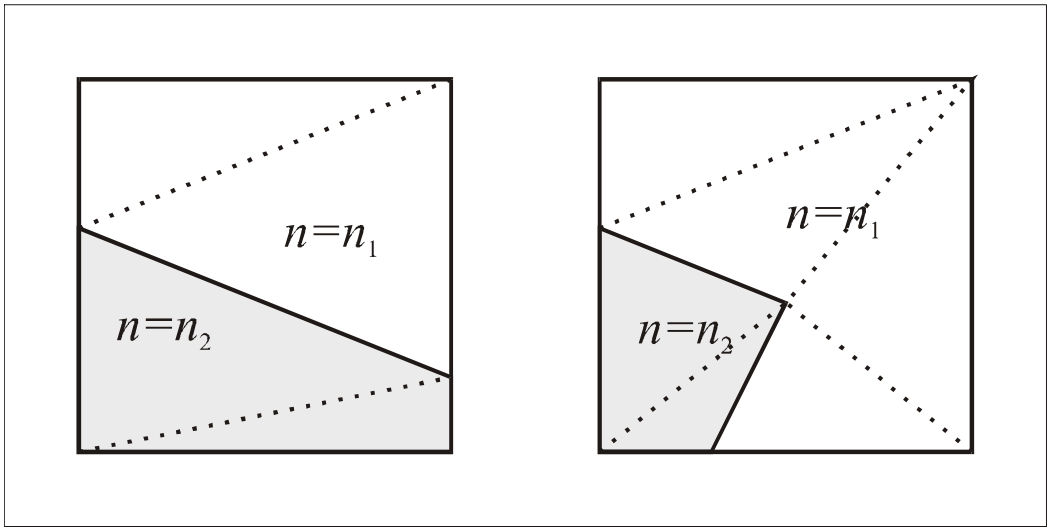


Figure 4.3: The two distinguished cases for an interface between materials crossing a grid cell. The dotted lines show how the split rectangle is divided into triangles. Left: One straight interface. Right: Two straight interfaces and a corner.

Since for TE polarization both the field E_y is and its spatial derivatives are continuous across an interface, one may use the same bilinear interpolation for the field; the interpolation complies with the interface conditions up to the same order as the interpolations themselves. However, this results in very cumbersome expressions as the integrals are performed over the trapezoids or triangles that result from the breaking-up of the original rectangle. An easier method is to:

1. Chop the rectangle into triangles, the angles of which coincide with the interfaces (see Figure 4.3);
2. Use the original bilinear interpolation to determine the field on the corner points of the triangles in terms of the field on the grid points;
3. Once the field on the corner points of the triangle is known in terms of the field on the grid points, the field inside the triangle is again interpolated using a bilinear interpolation similar to equation (4.11), except with c=0, since there are only three corner points and therefore three unknowns (a, b and d) are necessary.
4. This interpolated field can be used to calculate the integral across the triangle. When the contributions from all triangles are added together, the total integral across the rectangle is again know in terms of the field on the grid points.
5. In the same way as before, the new stencil for the affected points can be derived.

Since for sufficiently smooth fields, the interpolation of the fields on the triangles is of the same order in the stepsize as the bilinear interpolation on the rectangle, the order of convergence is not decreased for these sufficiently smooth fields.

For TE polarization, this method works for both distinguished cases (straight interface or interface with corner), since the original bilinear interpolation complies with the interface conditions. However, for TM, the normal derivative of the field H_y is not continuous across an interface, as was also seen in Chapter 2. For the case of a straight interface, it is possible to create triangles in a similar way as for TE, in which the normal derivative across the interface has the correct jump. For the corner inside the cell, this will not be possible; a singularity in the fields may occur at the corner. If one were to use the same interpolation as for TE polarization, the error in a cell containing a corner point will therefore be order 1, while the error in uniform grid cells is of the order of the stepsize squared. However, since the number of cells containing corners is constant, no matter the stepsize, the influence of this error on the total functional is also of the stepsize squared; decreasing the spatial stepsize increases the total number of cells, so the relative influence of the error on the cells containing corners decreases accordingly. However, this was not checked numerically; in the implemented algorithm, corner cells for TM polarization are replaced by similar cells with straight interfaces.

In the described way, the matrix equation $\mathbf{A}\vec{y}_{\text{int}} = \vec{b}$ is derived. The vector $\vec{y}_{\text{int}}$ contains the internal points of the grid; they are ordered by stepping through the grid from the top left-hand corner of the window to the bottom right-hand corner, one column after the other. Since the stencil is 3x3 points in size, the matrix **A** contains 9 diagonals: one on either side of the main diagonal, and 3 on both sides at a distance equal to the number of points in the x-direction away from the center. The matrix equation could be solved directly by some Gaussian elimination process, but the amount of memory that is needed for such a process is very large. During the process,

all diagonals that were originally filled with zeroes between the center and the diagonal that is furthest away from the center will become non-zero, causing a much larger memory requirement. Therefore, the matrix equation needs to be solved in a more memory-efficient manner.

4.1.4 Matrix solver

For simplicity, we will discuss some properties of a solver of the 1-dimensional TE problem. The matrix equation can easily be derived in the same way as shown above by using linear interpolations between grid points. As an example, we will look at a window of uniform material, n=1, in which light with a wavelength l=1.0 μm. The window is 3.0 μm long; the stepsize Dx is 0.1 μm. On the boundaries, Dirichlet boundary conditions are applied. Since the eigenvalue distribution of the matrix **A** is important for the applicability and efficiency of iterative solvers, these eigenvalues are calculated. For the resulting matrix without PML's, the eigenvalues can be calculated to be real with values between approximately –40 and +4, as shown in Figure 4.4a. When a PML with a width of 0.5 μm and s_x=1 is added on both sides of the window, the eigenvalues all become complex with a negative imaginary part, as shown in Figure 4.4b.

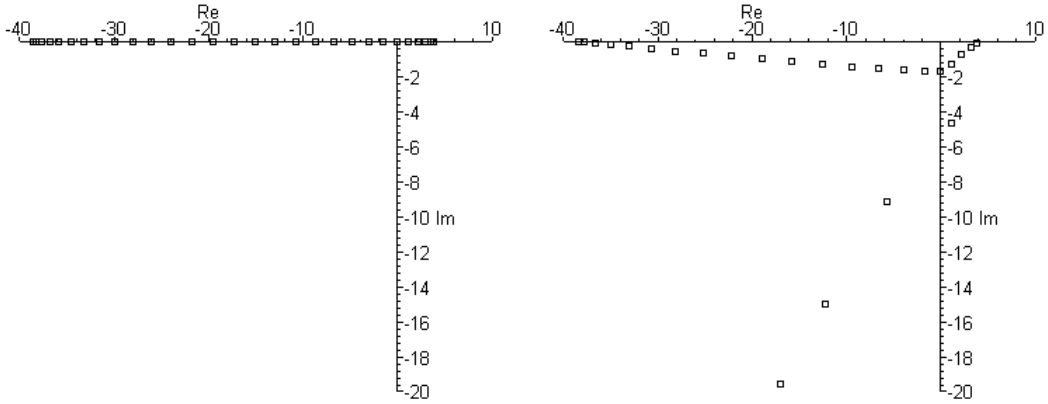


Figure 4.4: Eigenvalues of the matrix **A** *resulting from a 3-*μm *wide window, without (left) and with (right) PML's*

The system without PML's may have an eigenvalue that is very small; a zero eigenvalue means that an integral number of wavelengths fit in the calculation window. The main effect of the PML's on the eigenvalues is to make the eigenvalues complex; however, there are several values in the graph that have a much larger imaginary part than the other eigenvalues. These are due to the fact that the PML has the same strength everywhere; it is switched on abruptly, which has an influence on the numerical performance of the PML. This is seen as extra reflection of the PML. When the PML has a strength that is switched on slowly from the inner boundary toward the outer boundary, these eigenvalues becomes much closer to the other ones, and the numerical reflection from the PML is greatly reduced.

We have chosen to use a Conjugate Gradient solver [Press et al 1989] for the matrix equation, because it is an elegant method that should be able to deal well with any positive-definite matrices (including those with large ratios between their highest and lowest eigenvalue, i.e. with high condition numbers), and is theoretically guaranteed

to converge to the solution in a finite number of steps. However, it, and many other methods with it, demands that the matrix involved has only positive eigenvalues. This is obviously not the case with the matrices involved here. Therefore, instead of equation (4.10), the following equation is solved:

$$\mathbf{A}^{*}\mathbf{A}\vec{y}_{\text{int}} = \mathbf{A}^{*}\vec{b} \tag{4.15}$$

in which $\mathbf{A}^{*}$ is the adjoint of $\mathbf{A}$, which for these symmetric matrices is just its complex conjugate. The multiplication of $\mathbf{A}^{*}$ with $\mathbf{A}$ gives a matrix which has only positive and real eigenvalues, even for the case with PML's; see Figure 4.5. The eigenvalues range from approximately 0.7 to 1480.

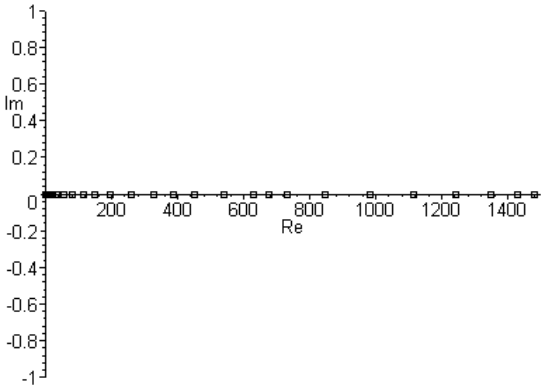


Figure 4.5: Eigenvalues of the matrix $\boldsymbol{A}^{*}\boldsymbol{A}$

This matrix is already pretty badly conditioned with a condition number (ratio of highest to lowest eigenvalue) of about 2000. It gets worse as the window size is increased; for a window width of 5 μm, the condition number becomes 7000, and worse for larger sizes. This is due to the smallest eigenvalues; the highest ones stay near 1500, but the lowest go ever closer to zero. A high condition number is disadvantageous to the solver, because when the solver is optimizing in directions that correspond to very low eigenvalues, the residue of the matrix equation decreases only very slowly.

Theoretically, the Conjugate Gradient solver is guaranteed to reach the solution of the equation within *N* iterations, where *N* is the number of unknowns in the equation. This is because it always chooses search directions that are 'conjugate' to all precious search directions; these conjugate directions are independent of each other. There are only *N* different conjugate directions, so when the algorithm has optimized in all these directions, it has reached the solution. However, due to numerical noise the directions will not be exactly conjugate, which means that in the numerical implementation, it may take more than *N* steps for the system to converge to within an acceptable tolerance.

The one-dimensional case can easily be solved by just using Gaussian elimination, but in order to test the conjugate gradient method in this type of problem, it was also implemented for the one-dimensional case. We look at the speed of convergence of the method for a 1024-cells wide window. The wavelength is 1.0 μm, the stepsize Δx

is 0.05 μm, the PML is 0.4 μm wide, has a quadratic envelope and is chosen such that one traversal through it decreases the amplitude of the light by a factor 25. On the left-hand boundary, the field is set to 25; the right-hand boundary field value is zero. In Figure 4.6, the norm of the residue $\vec{b} - \mathbf{A}\vec{y}_{\text{int}}$ is shown versus the number of iterations. The norm that was used is defined as:

$$|\vec{x}| = \frac{\sqrt{\sum_i |x_i|^2}}{n}$$

with n the number of elements of the vector $\vec{x}$.

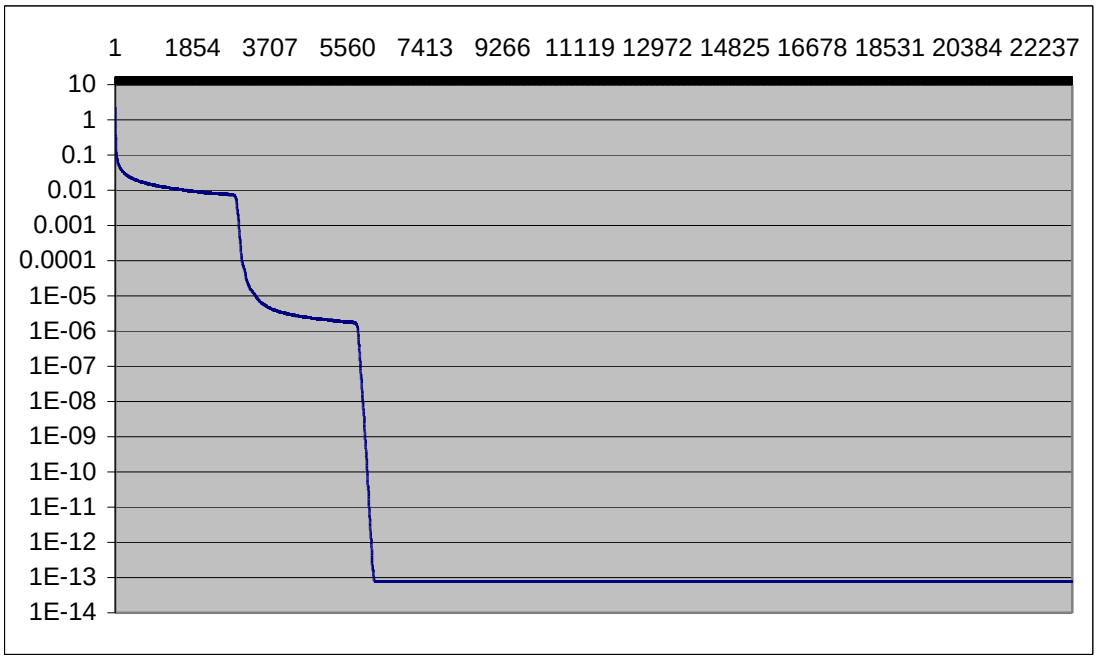


Figure 4.6: Convergence of conjugate gradient method for a 1024-cell one-dimensional system with PML's

Although the conjugate gradient method should theoretically have converged after 1024 steps, this is not the case. This is due to numerical errors; the search directions of the conjugate gradient method should all be conjugate, but due to the limited precision of the computer, they are not exactly conjugate. The graph becomes constant after about 6000 iterations; this is again due to the accuracy of the computer. However, the quantity $\mathbf{A}^*\vec{b} - \mathbf{A}^*\mathbf{A}\vec{y}_{\text{int}}$ is calculated inside the algorithm with higher accuracy, and this shows a continued periodic behaviour like this: First, the residue drops quickly, after which the convergence slows down dramatically; the error stays nearly constant for a large number of iteration steps, after which it starts the cycle again. This effect is likely related to the bad conditioning of the matrix; the system spends a lot of time optimizing in directions corresponding to eigenvalues that are nearly zero, which has very little influence on the residue.

Exactly the same algorithm has been implemented for the two-dimensional system. Similar behaviour is observed: First, the residue drops quickly but it gets slower and

slower, until it speeds up again. As said, this effect is likely due to the bad conditioning of the matrix. Therefore, it can be very beneficial to look for good preconditioners for the matrix, which has the effect of improving the effective condition number of the matrix.

The theoretical calculation time necessary for this method is proportional to the total number of points squared, since the theoretical number of iterations is equal to the number of points, and the time spent in each iteration step is proportional to the number of points. However, since the condition number increases when the stepsize is increased, the actual speed is worse than this; a worse condition number means longer calculation times.

4.1.5 Results of simple tests

This section contains two parts: First, the boundary conditions and incoming field generation are tested; second, interfaces crossing grid cells are investigated. All the results shown are for the TE method.

In order to investigate the boundary conditions, two different things are done: First, the effect of the adjustment of the amplitude of modes for a Gaussian beam is considered; then, the reflectivity of the boundaries and the modal field generation is investigated.

A calculation window of 9x9 μm is used, in which the stepsize Δx and Δz are 0.1 μm. The index of refraction of the whole window is 1.0, and the wavelength of the light used is 1.0 μm. A PML with a quadratic envelope (i.e. its strength increases quadratically from the inner boundary to the outer boundary) surrounds the window, which attenuates light traversing it at normal incidence by a factor of 50. A Gaussian with a beam waist (the full width where the intensity of the beam is 1/e) of 0.5 μm is launched into the window from the left. Figure 4.7 shows for two cases the resulting field after the iterations have converged. In the left-hand picture, all modes have the same amplification factor, so the field profile at the outer boundary is just the Gaussian, and in the right-hand picture, each mode in the structure (in this case, each plane wave, since there is no index contrast) has its own amplification constant, corresponding to its propagation constant normal to the boundary. The graphs show a logarithmic contour plot of the intensity.

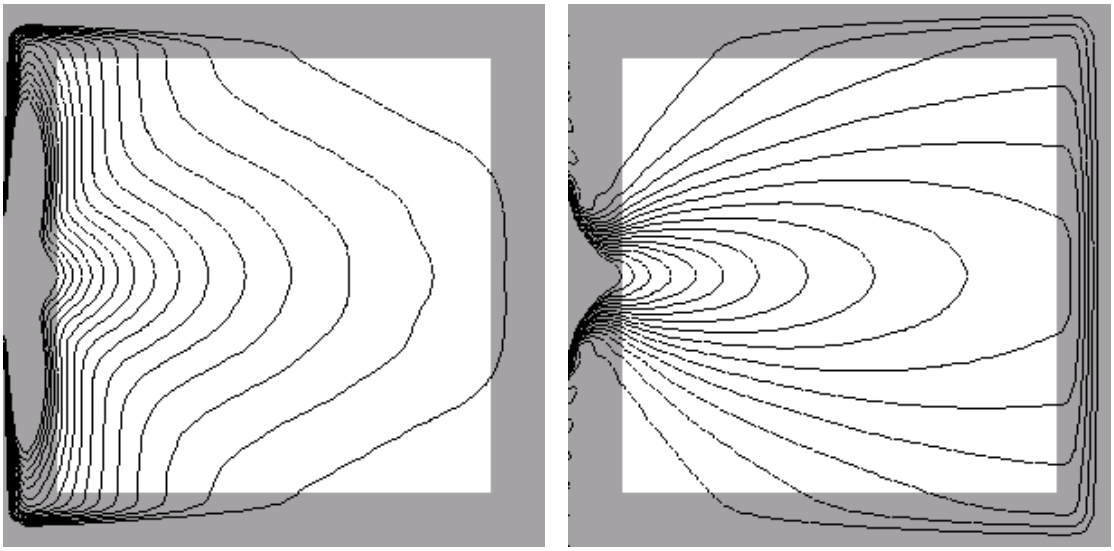


Figure 4.7: Logarithmic contour plots of a Gaussian beam with a waist of 0.5 mm. Left: uncorrected input field; right: input field with corrected modal amplitudes. The grey area is the PML region.

In the right-hand picture with the corrected boundary condition, the light diverges as the Gaussian beam should; the left-hand picture shows a field pattern that is no longer Gaussian at all. This shows that it is necessary to use the correct amplification factors of each mode.

In order to investigate the reflectivity at the PML, first the one-dimensional system is tested. A window containing 128 grid cells is used; the wavelength is 1.0 μm and the index of refraction 1.0 everywhere. Six different systems are simulated. The stepsize Δx is either 0.05 μm or 0.025 μm; the PML's are 4, 8 or 16 cells thick. For each of these systems, the PML has a quadratic envelope and its strength is varied such that the attenuation of the amplitude in a single traversal of the PML (attenuation factor) lies between 2 and 250. The field on the left-hand boundary is set to a value equal to the attenuation factor, so that the analytical amplitude of the right-traveling wave in the area between the PML's is 1.0. Reflections from the right-hand boundary will create a standing wave pattern in the central area; from the modulation depth of this pattern, one can directly deduce the ratio of the intensity of the reflected wave and the incoming wave. For each of these systems, the power reflected from the right-hand boundary is plotted in Figure 4.8.

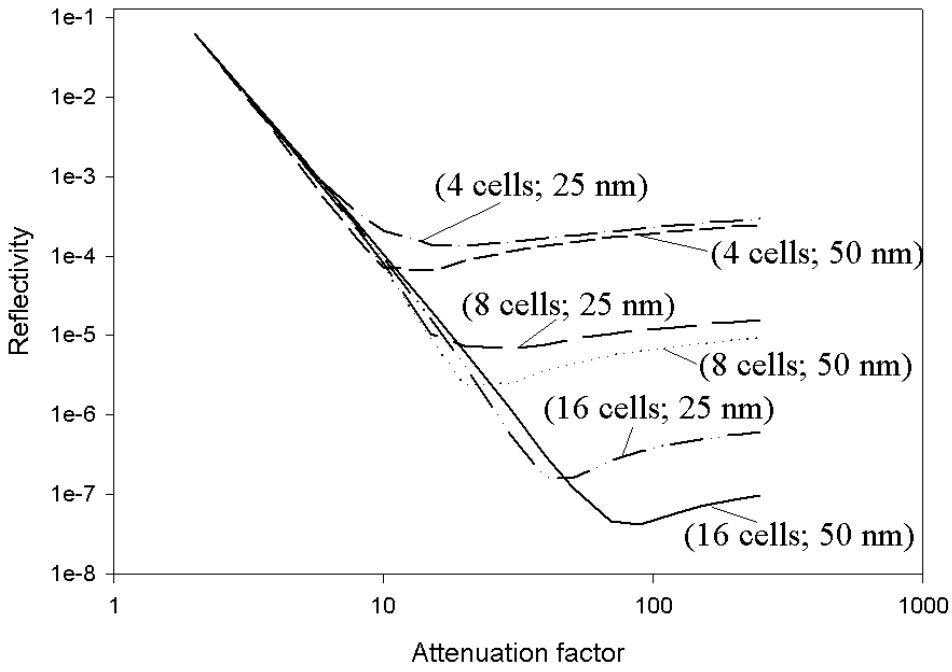


Figure 4.8: Reflectivity from a 1-dimensional PML versus the attenuation factor of the PML. For each line, the number of cells in the PML and the stepsize is shown.

For the analytical PML, the curves should all lie on a straight line, since the analytical power reflectivity is (attenuation factor)^4. However, in the numerical scheme, the quickly-changing amplitude of the light in a strong PML cannot be accurately taken into account, and neither can sudden changes in the strength of the PML. This causes the reflectivity to reach a minimum, after which it starts rising again.

It turns out that both the width of the PML with respect to the wavelength and the number of grid cells in the PML influence the behaviour of the PML. For the same number of grid cells in the PML, a wider PML has a better optimum, while for the same width, a higher number of points (i.e. a smaller stepsize) is generally better.

For the two-dimensional system, a similar test is performed. The fundamental mode is launched into a waveguide that is 0.4 μm wide, with an index of 1.5, surrounded by air. The strength of the PML is chosen such that for the analytical mode, the attenuation constant is as designed. The rest of the parameters for the six different systems are the same as those in the one-dimensional case; the stepsize Dx (perpendicular to the waveguide) is always 0.05 μm, the stepsize Dz (parallel to the waveguide) is either 0.05 μm or 0.025 μm. The input power and the reflectivity are calculated from modal overlap calculations. Figure 4.9 shows the reflectivity vs. the attenuation factor for this system.

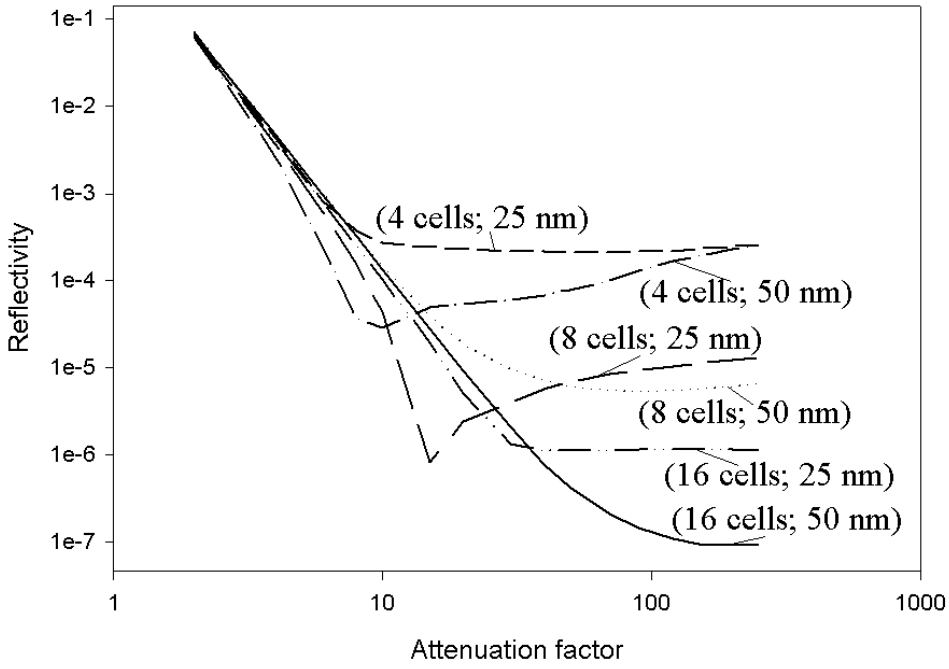


Figure 4.9: Reflectivity from a PML for a mode in the 2-dimensional system. For each line, the number of cells in the PML and the stepsize is shown.

Again, the reflectivity has a minimum for a certain attenuation constant. For these calculations, the reflectivity becomes quite a bit lower than the analytical line. This is likely due to the fact that modal overlaps are taken to calculate the reflectivity; some numerical reflection may occur into other modes than the fundamental one (e.g. into radiation modes), and is therefore not measured correctly.

The power in the guided mode is designed to be 1. However, the mode first has to traverse the numerical PML. Near the outer boundary, the strength of the PML is highest due to the quadratic envelope of the PML; also, the gradient of the strength of the PML is highest. This means that the amplitude drops very rapidly in that area, which will cause problems with the numerics; when the field or the strength of the PML change rapidly over a few grid cells, the operators in the Helmholtz equation are not discretized accurately anymore. Figure 4.10 shows the power in the fundamental mode of the waveguide in the central area for the same systems as those in Figure 4.9.

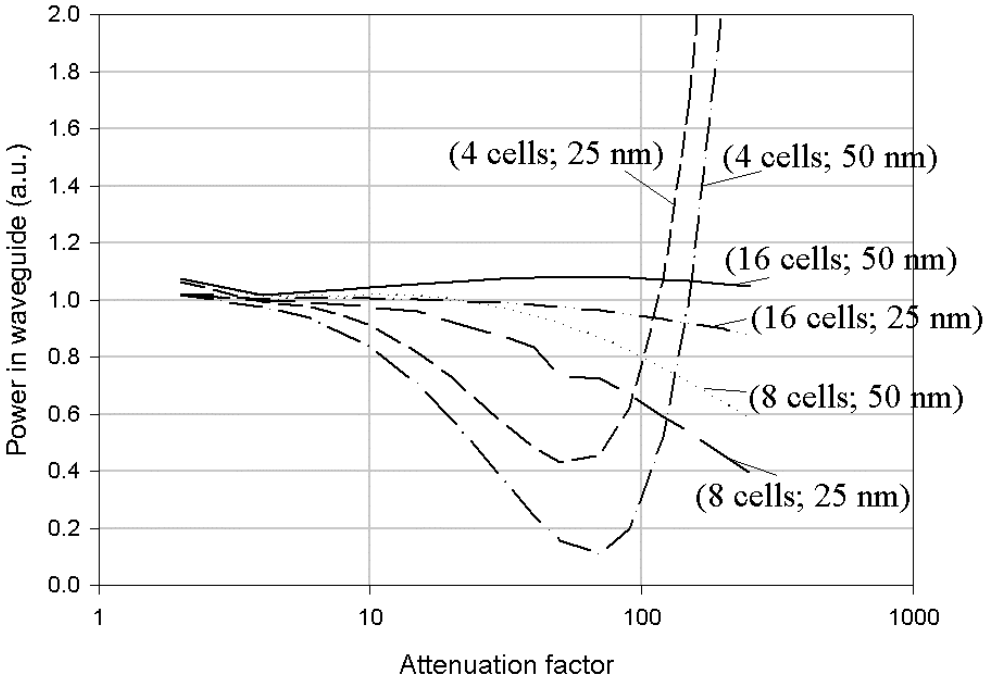


Figure 4.10: Right-traveling Power in the fundamental mode of the waveguide for various systems.

It is clear from these simulations that the error that is introduced by a thin PML is much more pronounced in the input power than in the reflectivity. This is due to the fact that the PML strength, and therefore its numerical inaccuracy, is highest near the edge. For the reflectivity, this does not matter much, since the amplitude near the edge for a reflecting field is not very high anyway; it has already been attenuated on its approach, so extra numerical scattering or reflection in this position has some influence, but not much. The incoming field, however, is high at the edge, so errors that are introduced into the field here have a large influence on the field inside the window. It turns out that in almost all cases, we see a considerable error in the intensity in the waveguide; even for the 16-cell PML, it becomes as large as 10%. A practical solution to overcome this problem is to do two calculations: One without the device under study (i.e. only the incoming waveguide) to determine the incoming modal intensity, and one with the rest of the device included to calculate the reaction of this modal field to the rest of the index structure.

Interfaces positioned somewhere in the grid cell (i.e. not coinciding with a grid boundary) are taken into account in the discretization, as discussed in Secion 4.1.3. In order to investigate this, a similar test is done as in the previous chapter: The same straight waveguide as in the previous example is simulated. Its fundamental mode is launched and its effective index in the window is determined through overlap calculations. Since Dz is kept constant, the effective index will not converge, in the limit of $Dx \rightarrow 0$, to the analytical one (which is 1.3234), but to another value, which

turns out to be 1.3208. In the graphs, the difference between the simulated value and this convergence value is plotted. The same calculations are done with and without the corrections to the interfaces. Δz, parallel to the waveguide, is kept constant at 0.05 μm; Δx is varied. Figures 4.11 (uncorrected) and 4.12 (corrected) shows the results of these calculations.

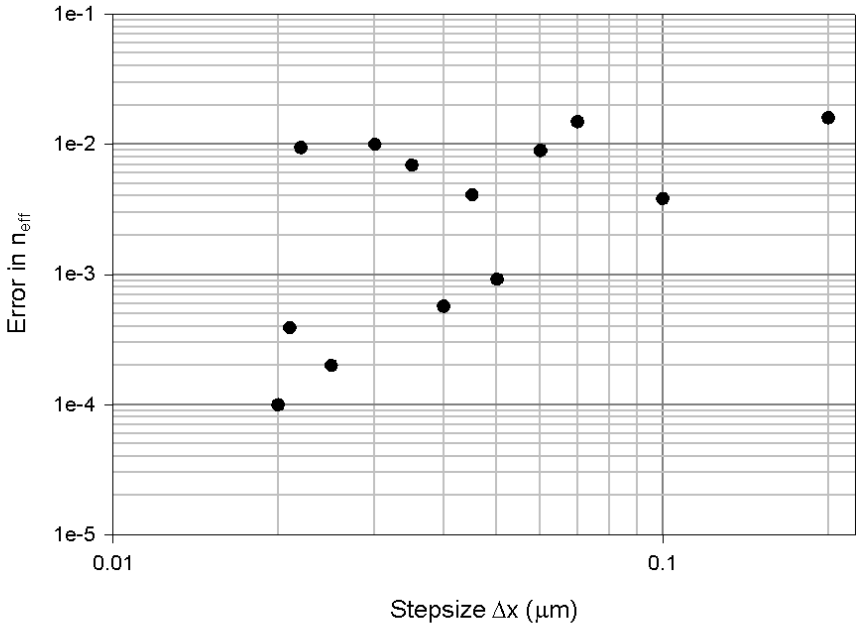


Figure 4.11: Difference between the simulated effective index and the converged effective index for the system in which interfaces running through grid cells are not taken into account accurately, i.e. all cells have a constant index of refraction.

In the uncorrected case, the points for which the interfaces lie exactly on grid points give a curve that converges as $O(\Delta x^2)$. However, all other points lie above this line, and the error seems to be of order 1, though due to the small amount of points, no definite concluations can be drawn about this; $O(\Delta x)$ is what we would expect. In the corrected case, the points all lie near or below the same line of convergence. The reason that the points with the interfaces exactly on grid cells do not give exactly the same result as in the uncorrected case is because the cells that contain the interface are still broken up into triangles, which means that the interpolations on the cell are different from the uncorrected case. It is clear that the corrections increase the accuracy of the method tremendously; the global error is $O(\Delta x^2)$.

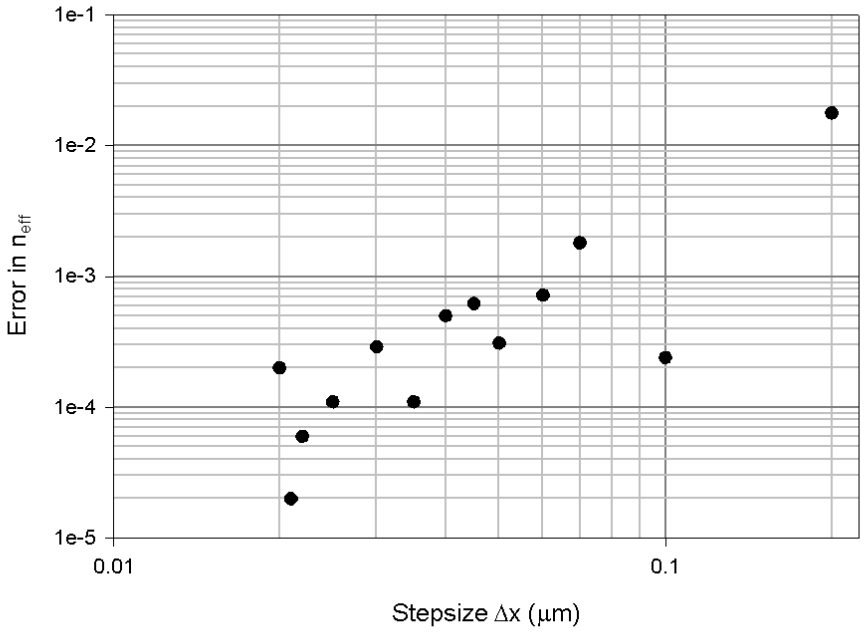


Figure 4.12: Difference between the simulated effective index and the converged effective index for the system in which interfaces running through grid cells are taken into account accurately.

4.2: FDTD with PML's and modal pulse generation

The Finite Difference Time Domain (FDTD) method is a well-known method for the simulation of electromagnetics [Taflove 1995, Yee 1966]. In this method, PML's [Bérenger 1994] are often used as nonreflecting boundary conditions. For incoming field generation, mostly methods like the so-called total field / scattered field method are used [Taflove 1995]. But it is also possible to do the same as in the previous section: Start the field on the outer boundary and let it propagate through the PML before it enters the 'real' window. This section will briefly describe the two-dimensional fdtd method with PML's, and the startfield generation. Since the FDTD method simulates the full time-dependent Maxwell equations, the startfield may contain more than one frequency. If a pulse is launched, this contains a range of frequencies that are all simulated simultaneously. Of course, if one wants to simulate the response of a structure to an incoming mode in a waveguide, each of the frequencies contained in the pulse should have its own modal profile. The generation of this modal pulse will be described after the general algorithm is considered.

4.2.1: FDTD with PML's

As described in Chapter 2, the time-dependent Maxwell equations must be split in order to have a good nonreflecting PML. The equations for TE polarization are:

$$\begin{aligned}
\frac{\partial}{\partial t}\varepsilon E_{yx} &= -\frac{\partial}{\partial x}H_z - \varepsilon\sigma_x E_{yx} \\
\frac{\partial}{\partial t}\varepsilon E_{yz} &= \frac{\partial}{\partial z}H_x - \varepsilon\sigma_z E_{yz} \\
\frac{\partial}{\partial t}\mu H_x &= \frac{\partial}{\partial z}\left(E_{yx}+E_{yz}\right) - \mu\sigma_z H_x \\
\frac{\partial}{\partial t}\mu H_z &= -\frac{\partial}{\partial x}\left(E_{yx}+E_{yz}\right) - \mu\sigma_x H_z
\end{aligned} \tag{4.16}$$

and for TM:

$$\begin{aligned}
\frac{\partial}{\partial t}\mu H_{yx} &= \frac{\partial}{\partial x}E_z - \mu\sigma_x H_{yx} \\
\frac{\partial}{\partial t}\mu H_{yz} &= -\frac{\partial}{\partial z}E_x - \mu\sigma_z H_{yz} \\
\frac{\partial}{\partial t}\varepsilon E_x &= -\frac{\partial}{\partial z}\left(H_{yx}+H_{yz}\right) - \varepsilon\sigma_z E_x \\
\frac{\partial}{\partial t}\varepsilon E_z &= \frac{\partial}{\partial x}\left(H_{yx}+H_{yz}\right) - \varepsilon\sigma_x E_z
\end{aligned} \tag{4.17}$$

As seen often in literature [Taflove 1995, and references therein], we use Yee's Mesh [Yee 1966] for the grid. This is a staggered grid in both space and time. The *E* and *H*-

fields are located at half a timestep away from each other; the spatial distribution is shown below in Figure 4.13.

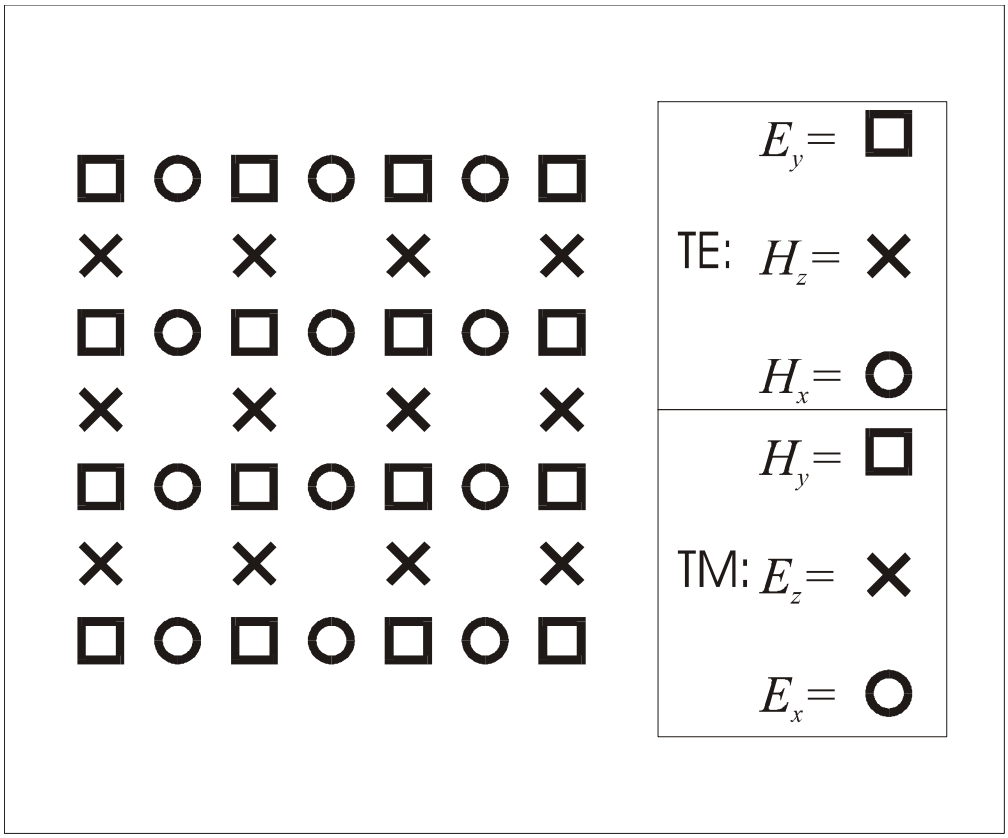


Figure 4.13: Spatial distribution of Yee's mesh for the Finite Difference Time Domain method

The index of refraction is assumed to be constant in each cell; this means that the index is approximated be a staircase. It may be possible to make similar corrections as those described in Section 4.1.3 for the Helmholtz equation, but this would have the disadvantage that the discretization is no longer similar on all points, and therefore, much more memory would be used to store the data. Due to time constraints, this option was not studied further.

The derivatives in equations (4.16) and (4.17) are replaced by simple central differences. For the terms with the conductivities, the average of the field at the old and the new timestep is taken. For the timestep, a simple relation must be satisfied in order to enforce stability of the scheme:

$$\Delta t < \frac{1}{c\sqrt{\frac{1}{\Delta x^2} + \frac{1}{\Delta z^2}}} \tag{4.18}$$

This means that when one increases the accuracy by decreasing the spatial step sizes, the total calculation time scales as $N^{3/2}$, with N the total number of grid points.

Some care must be taken with the conductivities. Due to the spatial staggering of the fields, the *E*- and *H*-fields are located at different positions within the PML. The

strength of the PML is chosen to increase as a quadratic function, as shown in Figure 4.14.

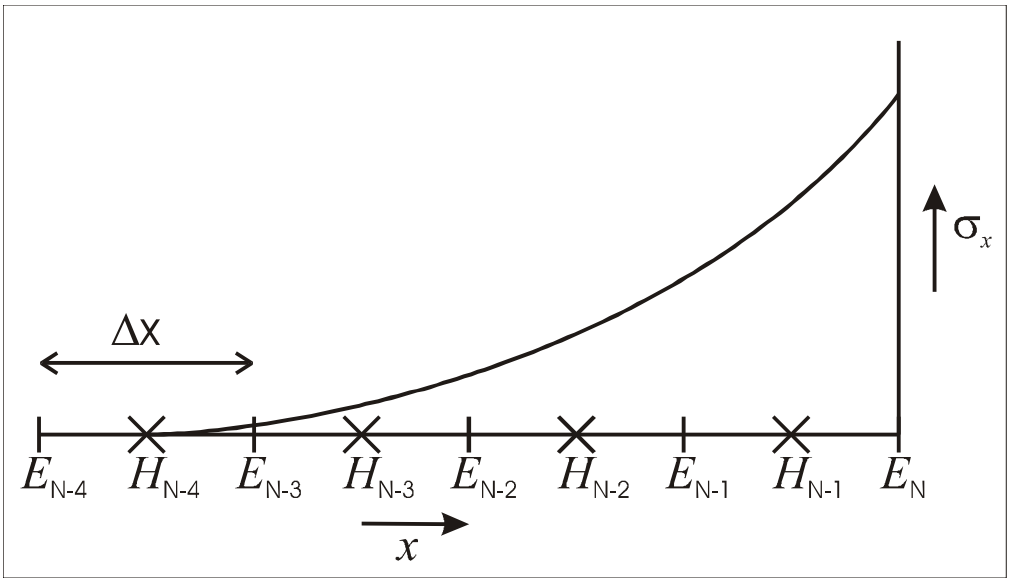


Figure 4.14: Example of a quadratic distribution of s_x at the right-hand side of the window; in this case, s_x is zero left of x=(N-3.5)Δx, the location of H_{N-4}. The positions of the E_y and Hx fields is shown for TE polarization. E_m is the E_y-field located at x=mΔx; H_m is the H_x-field located at x=(m+½)Δx.

The actual s_x that is used in the algorithm is a staircase approximation, with staircases with the length of the grid cell, of this quadratic function. The attenuation of the PML is determined by the integral of s_x over x, so the integral of the staircase approximation should be the same as the integral of the real, quadratic, shape. Since the *E*- and *H*-fields are staggered, the staircase approximation of s_x should be different, yielding two different formulas, $s_{x,E,m}$ and $s_{x,H,m}$, for this approximation:

$$s_{x,E,m} = \frac{1}{\Delta x}\int_{(m-\frac{1}{2})\Delta x}^{(m+\frac{1}{2})\Delta x} s_x dx$$
$$s_{x,H,m} = \frac{1}{\Delta x}\int_{m\Delta x}^{(m+1)\Delta x} s_x dx \qquad (4.19)$$

$s_{x,E,m}$ is the value that is used in the discretized analogue of the first equation of (4.16); $s_{x,H,m}$ should be used for the fourth equation. For the other two equations, similar staircase approximations are made for s_z.

When launching the field on the outer boundary of the window, it is only necessary to set the y-component of the field (i.e. E_y for TE and H_y for TM), since all the other components follow from this one. It does not matter whether one sets E_{yx} or E_{yz}, since in the calculation of the other quantities, only the sum of the two is used.

As shown in Chapter 2, the analytical attenuation of a PML does not depend on the frequency if the effective index is the same. But, just like in the previous section, a mode with a different effective index will also have a different attenuation. This

means that, again, the input field has to be adjusted; each mode needs to have its own amplification constant.

4.2.2 Modal pulse generation

When simulating light in integrated optical structures, one usually does not simulate the complete device; the input waveguides have often traveled a considerable distance from the edge of the chip to the area of interest. In that case, almost all the light will be in a mode of the waveguide. Therefore, in the simulations, the response of a device to a modal incoming field should be calculated.

As mentioned in the introduction to this section, the FDTD method can simulate multiple frequencies at once. For example, when a pulse is launched with a certain duration, this translates into a spectral width in the frequency domain. So, one could calculate a modal profile of a waveguide for a specified central wavelength, and launch this mode with a Gaussian envelope in time. However, each of the frequencies in the pulse will have its own modal profile in the waveguide. So the profile that is launched will be the correct shape for the central frequency, but for all other frequencies, the shape will not be equal to the modal field at that frequency, so some energy will go into other guided and radiation modes. This energy disturbs the calculation at that frequency, since one is only interested in the response to the modal field. Also, the amplitude of the modal field will not be correct, since each of the frequencies will also have a different effective index.

The way to solve this is to first calculate the modal profile for the frequencies involved. This profile is multiplied by the amplitude that that frequency should have and the amplification factor demanded by the PML. In this way, one obtains the spectrum of the fields at each lateral position. This spectrum can then be inverse Fourier transformed to obtain the time dependence of the pulse at this position. An example will be shown below.

Suppose, we want to launch a modal pulse form the left-hand side of the window, so at z=0. First, we determine the normalized modal fields and effective indices at all relevant frequencies (for example, a range of frequencies around the central frequency), yielding $e(x,\omega)$ and $n_{eff}(\omega)$. Then, the desired amplitude of the field at each of these frequencies is determined. For example, if one intends to launch a pulse with central frequency ω_0 and a Gaussian envelope with a width of w seconds, the time dependence of the uncorrected pulse is:

$$p(t) = e^{i\omega_0 t} e^{-\frac{2t^2}{w^2}} \tag{4.20}$$

whose Fourier transform is:

$$f(\omega) = F\left(p(t)\right) = w\sqrt{\frac{p}{2}} e^{-\frac{w^2(\omega-\omega_0)^2}{8}} \tag{4.21}$$

If the PML is designed to have an attenuation factor $A(\omega_0)$ for the amplitude of the mode of the central frequency, other modes will be attenuated by a factor:

$$A(\omega) = \left(A(\omega_0)\right)^{\frac{n_{eff}(\omega)}{n_{eff}(\omega_0)}} \tag{4.22}$$

Combining these equations, the pulse shape in ω-x-space on the outer boundary must be:

$$y(x,\omega) = A(\omega) f(\omega) e(x,\omega) \tag{4.23}$$

so the time dependence of the field on the outer boundary must be the inverse Fourier transform of this equation.

In order to numerically check the validity of this algorithm, a pulse in a waveguide is simulated. The pulse has a width w of 4 fs; the vacuum wavelength of the light is 0.9 μm. The waveguide is 0.5 μm wide; the core index is 1.6, the substrate is 1.45 and the cladding is 1.0. The polarization is TE. The stepsizes that are used are Δx=0.025 μm, Δz=0.05 μm and Δt=0.05fs. The PML is 0.8 μm wide on all sides, and attenuates approximately by a factor 25. Figure 4.15 shows two pictures of the pulse, both after 40 fs.

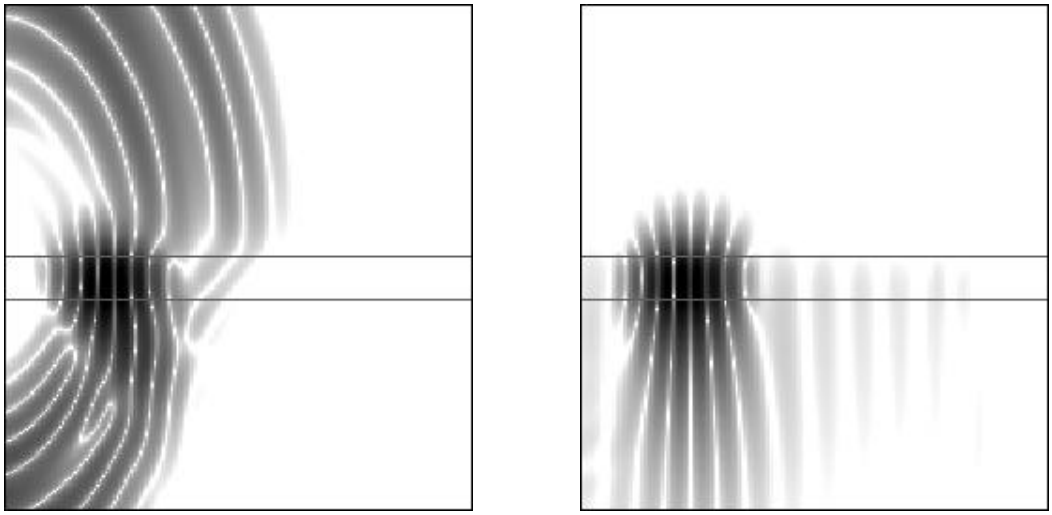


Figure 4.15: Modal pulse with constant lateral field profile (left) and with adjusted profiles at each frequency (right). The pictures show the logarithm of the absolute value of the field; the field at the lowest grey scale is about 0.2% of the field at the black areas.

It is clear that the pulse that takes different modal profiles at different frequencies into account gives a much better result. There are still some errors (e.g. the fields that are in front of the pulse), but their intensity is very low. This modal pulse generation has the advantage that one can use smaller windows in the simulations; no need to wait for the radiation to leave the system.

4.3 Conclusions

Following the concepts of the previous chapters, an omnidirectional solver for the TE and TM Helmholtz equation is developed. It is based upon the variational formalism, in which the solution of the equation is found by finding extremal points of a functional. The problem is reformulated: Instead of discretizing the wave equation, the functional is approximated by an interpolation using the fields on grid points, and the index distribution. The index distribution within a grid cell can be taken into account. Because of this, it is possible to do the calculations accurately with a rectangular grid that is independent of the structure, so one does not need complicated grid generation.

The calculation window is surrounded by Perfectly Matched Layers in order to have nearly reflectionless boundaries. Since the Helmholtz equation is a boundary value problem, the only way to give information about the incoming fields is by setting the field on the boundary. However, those fields must first traverse the PML before they enter the interior of the window, which means that they are attenuated by an amount that is related to the component of the wave vector perpendicular to the boundary. If a field is composed of multiple modes, i.e. if it is the sum of several orthogonal fields which have their own propagation constants, each of these modes will have its own wave number perpendicular to the boundary, which means that each mode is attenuated by a different amount. This must be compensated for by adjusting the amplitude of each mode accordingly.

Care must be taken with the choice for the strength and width of the PML; if it is too strong or too thin, there will be both numerical reflections and very inaccurate amplitudes of the input fields. By experimenting, an optimal value for the strength of the PML may be found for a minimal width. However, the amplitude of incoming fields will always have a significant error, so in general, two runs will be necessary for accurate simulations: One with only the incoming waveguide and no scattering structure, from which the amplitude of the incoming light as simulated can be deduced. This reference calculation may have a window size that is significantly reduced to reduce calculation time. This data can later be used in the analysis of the results of the second run, in which the full structure is present.

In the Finite Difference Time Domain method, the same boundary conditions are used, so the same care must be taken with the incoming field generation and the strength and width of the PML. In addition, the calculations are performed with full time dependence, so it is possible to simulate multiple frequencies at once. In order to calculate the response of a structure to modal input fields, the mode for each calculated frequency has to be different. A way to generate so-called modal pulsed is developed and works well.

For a truly good analysis of the data generated by FDTD simulations with modal pulse generation, the analysis should also be done on different modal profiles corresponding to the different frequencies, since if one is interested in modal input fields at many frequencies, one will also want to analyze the modal output fields at these different frequencies. A disadvantage is that the amount of memory needed is large; on a cross-section of a waveguide, the full time evolution of the field must be stored and Fourier transformed to get the spectrum at each point, from which the modal overlap with each frequency's mode may be calculated. This algorithm has not been implemented and tested yet.

The Helmholtz solver has the advantage of being able to be more accurate with a coarser grid, since interfaces through grid cells are taken into account accurately. However, on finer grids, the calculation time that the Helmholtz solver needs are much higher than those of the FDTD method, since the Helmholtz solver scales worse than N^2, while the FDTD scales as $N^{3/2}$ or better. Due to the additional advantage of the FDTD method that it can calculate for multiple frequencies at once, it is a more suitable method for most applications.

Chapter 5: Applications

In this chapter, the omnidirectional simulation methods that were described in the previous chapter are applied to two different types of structures. In both these structures, the light travels in all directions in the plane, so the omnidirectional simulation tools are necessary for these structures. The two structures that will be discussed are cylindrical microcavities and two-dimensional photonic crystal structures; for examples of these, see Figure 5.1.

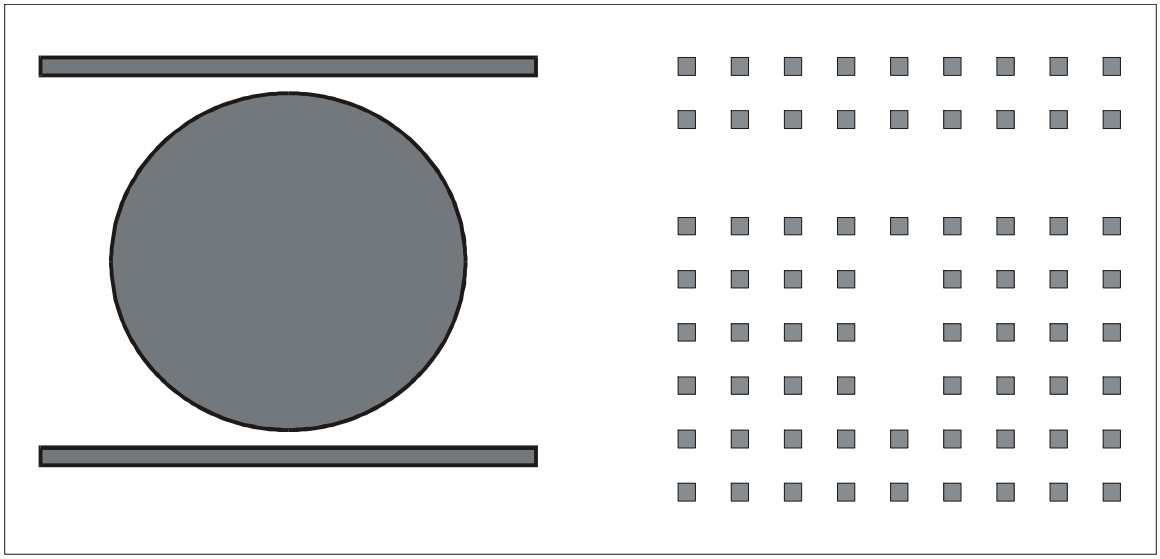


Figure 5.1: Examples of a cylindrical microcavity (left) and a filter in a photonic bandgap structure (right). The dimensions are on the scale of several wavelengths.

The types of structures that are discussed here are both promising structures for further miniaturization in integrated optics. Two main challenges in integrated optics are: to make the individual elements that perform the various functions on the light (e.g. power splitters, wavelength filters, add/drop filters, isolators) as small as possible in order to fit many functions on a small and cheap chip; and to make the interconnects between the elements as short and versatile (e.g. capable of changing direction quickly and crossing other interconnects without crosstalk) as possible. The cylindrical microcavity (or ring structure) can be used for e.g. add-drop filtering [Little 1997, Chin 1998, Hagness 1997] and passband flattening [Jinguji 1996, Oda 1988], and photonic crystal structures [Mekis 1996, Mekis 1998, Sigalas 1993] are promising candidates for interconnects and for e.g. filters.

In this chapter, both methods (Helmholtz solver and FDTD) are applied to gain insight into the performance of these structures; comparisons between the two methods are given. However, for some problems one method or the other is more applicable (i.e. more numerically efficient), in which case only one is used. This will be made clear in the text.

The first section deals with the cylindrical microcavity; the second one with various structures in photonic crystals.

5.1 Cylindrical microcavity

A cylindrical microcavity positioned between two waveguides may couple, depending on l, light from one of the waveguides to the other. Light from the incoming waveguide couples to the so-called whispering gallery modes of the microcavity [Little 1997, Hagness 1997, Balistreri 1999, Blom 1997, Chin 1998]. These whispering gallery modes travel along the edge of the cavity. The angular dependence of a whispering gallery mode is exp(-igf), in which g is the angular propagation constant. This g is complex for real w; g can also be written as $\gamma=k_0 n_{eff} R$. R is the radius of the disc. This n_{eff} corresponds to the standard effective index on the edge. For those frequencies for which the optical pathlength of one round-trip of a whispering gallery mode is an integral number times the wavelength (i.e. $\hat{A}(g)=m$, where $\hat{A}(g)$ denotes the real part of g), the cavity is said to be in resonance; light that is coupled into the cavity is in phase with the light already in the cavity. For these frequencies, nearly all the light will couple into the 'drop' waveguide, while not much remains in the 'through' waveguide. Note that the presence of the waveguides disrupts the whispering gallery mode locally, so the actual resonances will be at slightly different positions due to phase shifts at the waveguides. The efficiency of this process depends, among others, on the loss of energy during one round-trip through the cavity. In part, this loss is described by the imaginary part of the angular propagation constant g. However, during each roundtrip the mode also loses energy due to the coupling to the drop- and through-waveguides.

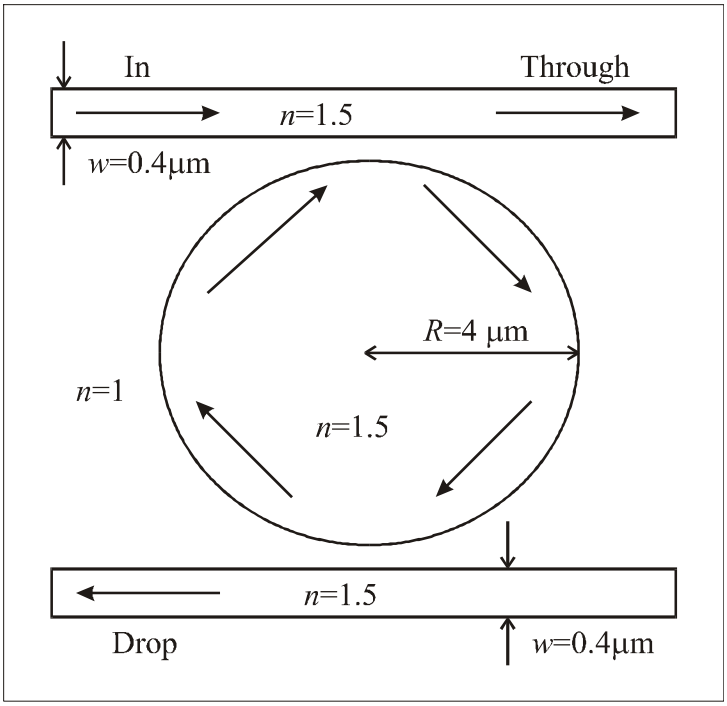


Figure 5.2: The cylindrical microcavity under study. Light comes in from the In port, and it leaves the system at the Through port, Drop port or as radiation.

From the spectrum of the fields at the drop- and through-ports of the system, two important quantities can be obtained for the whispering gallery modes: Their Free Spectral Range (FSR), which is the distance in w- or l-domain between resonances, and the Finesse, which is the ratio of the spectral Full Width at Half Maximum

(FWHM) of the peak in the power at the drop port to the FSR. These quantities are important because they define the channel spacing and the channel width of such devices in telecommunication applications.

The cylindrical microcavity under study is the one shown in Figure 5.2. The waveguides and the disc all have the same index of refraction of 1.5, while the background index is 1.0. The waveguides are both 0.4 μm wide and the radius *R* of the disc is 4 μm. The distance between the waveguide and the disc is 0.2 μm.

Using a method developed by D.J.W. Klunder, the angular propagation constant of the two whispering gallery modes that are supported by this structure are calculated. Only those modes with the real part of n_{eff} higher than the index of the background are considered to be whispering gallery modes here. The method is based on a transformation described in [Lui et al, 1998], with the addition of PML's located far away from the cavity (outside the so-called Radiation Caustic, which is the position at which the radial dependence of the field becomes mainly radiating). The whispering gallery modes are found by an eigenvalue search after a finite difference discretization of the fields and the operators. The stepsizes involved are chosen small enough for the method to be converged. The effective indices for the modes at λ=1.0 μm are:

$$\begin{aligned} n_{eff}^{0} &= 1.310 - i \cdot 1.1327 \cdot 10^{-5} \\ n_{eff}^{1} &= 1.134 - i \cdot 0.001888 \end{aligned} \tag{5.1}$$

For the calculation of the theoretical FSR of these modes, the real part of the effective index of these modes is also calculated at a slightly higher wavelength, λ=1.032 μm, and turn out to be 1.307 resp. 1.128. From these, the local FSR can be approximated by the following formula:

$$FSR = \frac{l_1 - l_2}{2p\,R\left(\frac{n_{eff,l_2}}{l_2} - \frac{n_{eff,l_1}}{l_1}\right)} \tag{5.2}$$

and is for the two modes:

$$\begin{aligned} FSR^{0} &= 29.1nm \\ FSR^{1} &= 30.8nm \end{aligned} \tag{5.3}$$

From the effective indices, the so-called Unloaded Finesse $F_{(0)}$, the finesse of the whispering gallery mode without the waveguides, can be calculated [Rafizadeh 1998] to be:

$$\begin{aligned} F_{(0)}^{0} &= 1756 \\ F_{(0)}^{1} &= 10.5 \end{aligned} \tag{5.4}$$

The loaded finesse (of the cavity with the waveguides) cannot be calculated easily. It will be attempted to extract this quantity, and the FSR, from the FDTD and Helmholtz simulations of the loaded system.

First, results of the finite difference time domain method will be given; then, the Helmholtz solver is applied to the problem, after which comparisons are made.

5.1.2 Results of FDTD applied to microcavity

A modal pulse, generated as described in the previous chapter, with a pulse width of 10 fs and a central wavelength of 1 μm is launched into the upper waveguide from the left. The polarization used is TE. The stepsize in both spatial directions is 0.05 μm; the temporal stepsize is 0.1 fs. The PML's are 0.8 μm wide and their attenuation factor for the amplitude of a mode at the central frequency is 50. The simulation is done twice; once without the disc to get a reference signal for the normalization of the results, and once with the disc. The simulations are run for 262,144 timesteps, i.e. 26.2 ps. In the window of 260x260 points that was used, this takes 1 hour and 50 minutes of calculation time on an 550 MHz Athlon personal computer, which means that this PC processes approx. 10^{10} points per hour.

Since the analysis of modes for pulses is not implemented yet, the analysis of the results is done by analyzing single points in the center of the waveguides. One point is at the through port, just left of the right-hand PML; the other point is at the drop port, just right of the left-hand PML. The time evolution of the field is stored, and then Fourier transformed. Of this Fourier transform the absolute value squared is taken to obtain the intensity at each frequency. This intensity is then normalized by dividing it by the input intensity at the same frequency, as determined from the field at the through port of the system without a disc.

In Figure 5.3 this normalized intensity at the through port is shown as a function of wavelength; Figure 5.4 shows the normalized intensity at the drop port.

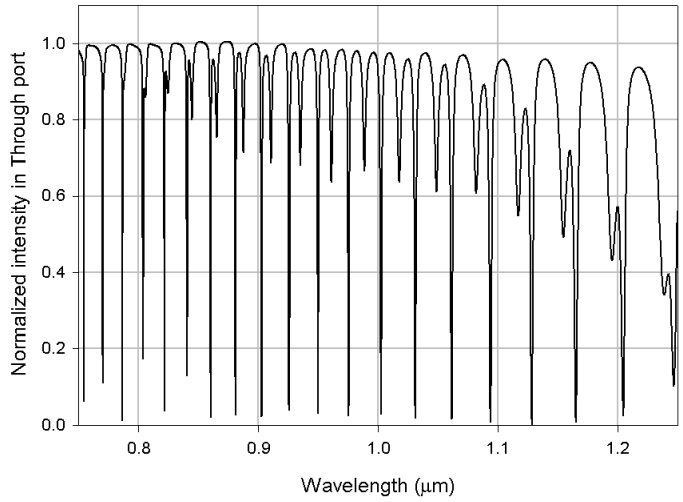


Figure 5.3: Normalized intensity at Through port of cylindrical microresonator as a function of wavelength

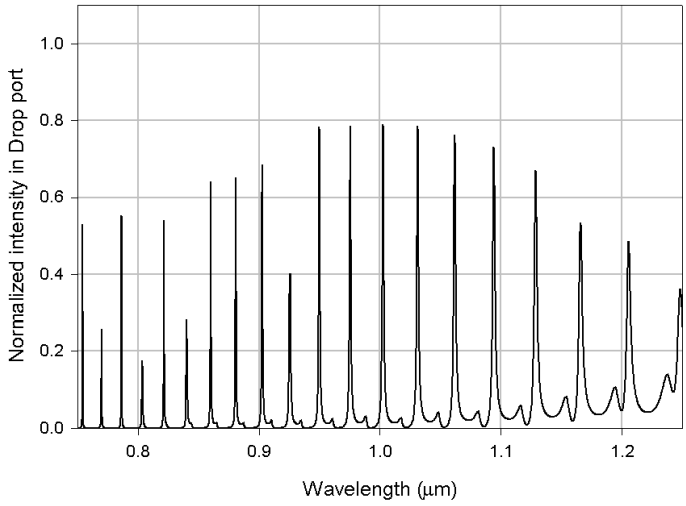


Figure 5.4: Normalized intensity at Drop port of cylindrical microresonator as a function of wavelength

The microcavity supports two modes in the studied wavelength region. These modes are both visible in the two graphs as peaks (in drop port) or dips (in through port). They have a different angular propagation constant, and therefore a different FSR, which can also be seen from the graphs; the period of the first-order mode is larger than the one of the fundamental mode. It can be inferred from the graphs that the first-order mode has large losses; the depth of the dips in the graph of the Through port is much larger than the height of the corresponding peak in the graph of the Drop port.

The rest of the energy is radiated away. For the fundamental mode, this effect is much less pronounced.

From the graph, the FSR of both modes for wavelengths around 1 μm can be determined to be:

$$FSR^0 = 28.7nm$$
$$FSR^1 = 31.0nm \tag{5.5}$$

These values are not exactly the same as the semi-analytical values (5.3), but they don't differ much from those. There are two possible reasons for this: Firstly, there is a certain amount of numerical dispersion, which causes the numerical wavelength to be somewhat different from the desired wavelength; and secondly, the presence of the waveguides disturbs the modes, and therefore also changes the real FSR.

From Figure 5.4, the finesse of the fundamental mode can be determined quite accurately at 21.7. Due to the low intensity at the Drop port of the first order mode, its finesse cannot directly be determined, but it is clear that it is quite low, the order of magnitude is 5. These loaded finesses are much lower than the unloaded finesses of the modes. This is due to the strong coupling of the waveguides to the microcavity; as mentioned before, this coupling introduces extra loss terms, which decrease the finesse. Figure 5.5 shows that the coupling from the waveguide to the cavity (and thus, the coupling from the cavity to the waveguide) is indeed large; it shows the absolute value of the field of simulated pulse as it has just passed the cavity in the upper waveguide for the system described above; the central wavelength of the pulse is 1.0192 μm. The field in the cavity is comparable to that in the waveguide; the maximum field strength in the cavity is approx. 40% of that in the waveguide.

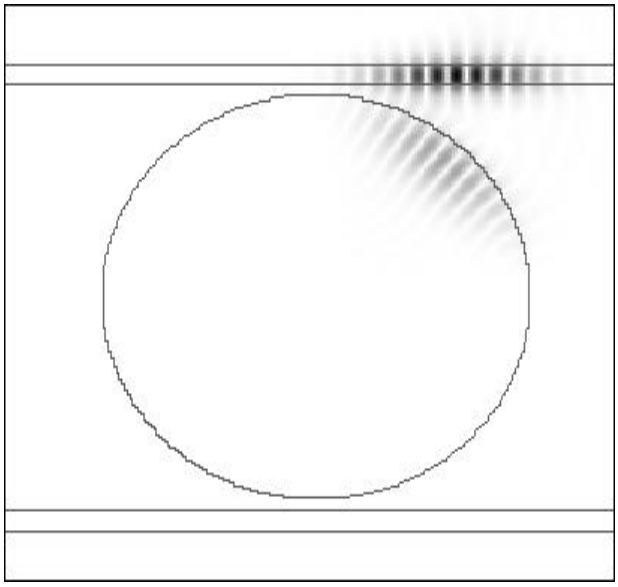


Figure 5.5: Absolute value of field after the incoming pulse passes the cavity

5.1.2 Results of Helmholtz solver applied to microcavity

The same structure described above is also simulated using the Helmholtz solver described in the previous chapter. Due to the fact that the Helmholtz solver only simulates for one wavelength at a time, and its relatively long calculation time, the resolution in wavelength space that is obtained is not as high as that for the FDTD simulations; each simulation takes between 1 and 2 hours to converge, so obtaining the same resolution would take prohibitively long. Only the wavelength range from l=1.0 μm to 1.060 μm is investigated with a resolution of 2 nm, except for one additional point at l=1.0192 μm. Figure 5.6 shows contour plots of the amplitude of the field at two different wavelengths: λ=1.0192 μm (near resonance of fundamental mode) and λ=1.010 nm (out of resonance).

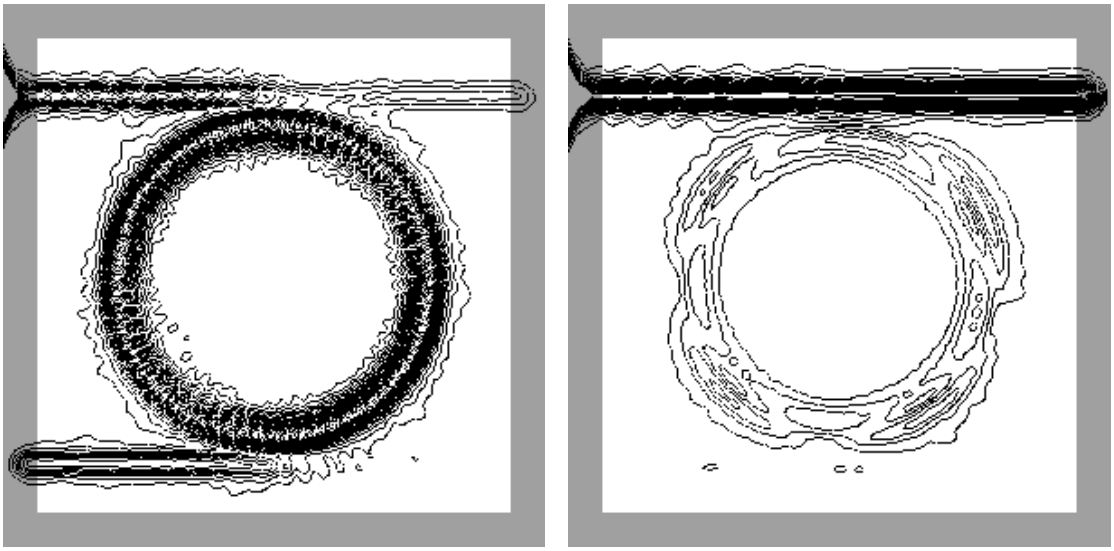


Figure 5.6: Contour plot of field at l=1.0192 μm (left) and l=1.010 μm (right). In the left-hand plot, the system is in resonance; in the right-hand plot, it is out of resonance.

Figures 5.7 and 5.8 show the normalized intensity at the through- respectively drop port; for comparison, the FDTD results for the same wavelength range is plotted as a dashed curve in the same graph.

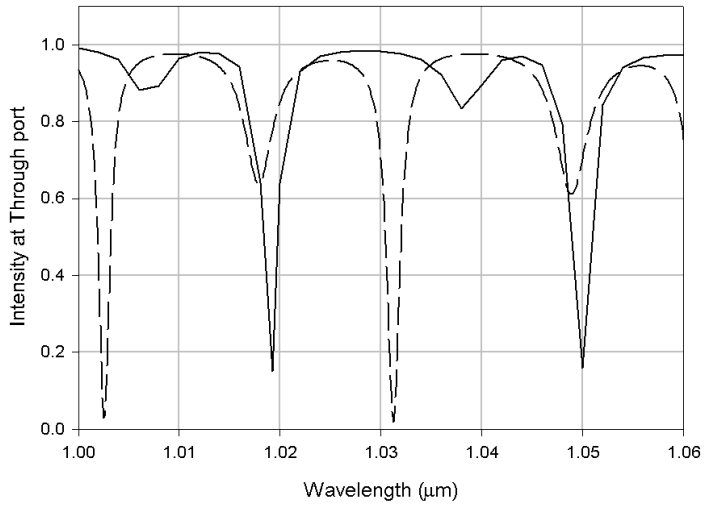


Figure 5.7: Normalized intensity at Through port for both FDTD (dashed line) and Helmholtz (solid line) simulations

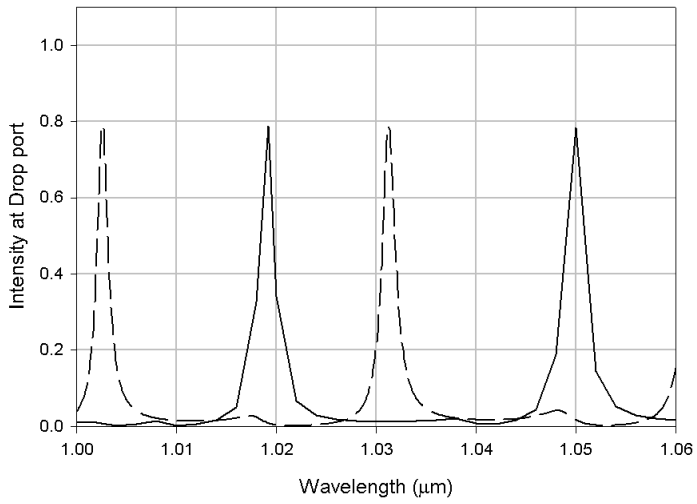


Figure 5.8: Normalized intensity at Drop port for both FDTD (dashed line) and Helmholtz (solid line) simulations

The results obtained by the Helmholtz solver look similar to the ones obtained using the FDTD method, but the peaks are shifted considerably. Also, the FSR is slightly different. Note that, due to the low sampling resolution, the exact peaks of the resonances cannot be determined accurately. This is most pronounced in the dips corresponding to the first order whispering gallery mode (located at approx. 1.007 and 1.039 in the Helmholtz simulations). The extra sampling point at λ=1.0192 was inserted to sample the peak at that position more accurately.

There are two reasons the peaks are shifted with respect to the FDTD simulations. Firstly, the discretization is different, and therefore the numerical dispersion as well. Secondly, the index distribution is taken into account more accurately in the Helmholtz solver, due to the fact that interfaces through grid cells are taken into account. The approximate FSR's of the two modes for the Helmholtz simulations are approx. 30.8 nm respectively 32 nm. According to the same theory of the unloaded microcavity as described in the previous section, they should be 30.2 nm and 31.2 nm.

5.1.3 Conclusions

Both the Helmholtz solver and the FDTD method are shown to be appropriate tools to simulate small microcavity structures. For the same grid size, the Helmholtz solver is more accurate due to its better treatment of slanted interfaces. However, the calculation time for a single wavelength for the Helmholtz solver is, for this particular structure, about equal to the calculation of a complete wavelength scan of the FDTD method; the calculation of the spectra in Figures 5.7 and 5.8 take approx. 3.5 hours for FDTD, and 50 hours for Helmholtz; and still, the sampling resolution of the FDTD is about 15 times higher.

Both methods consistently calculate a free spectral range that is higher than the predicted, unloaded FSR. As mentioned in Section 5.1.1, there are two reasons for this: Firstly, the numerical dispersion causes the numerical wavelength to be unequal to the real wavelength; and secondly, the presence of the waveguides disturb the cavity, also changing its free spectral range.

With both methods, the FSR can be calculated quite accurately, while the exact position of the peaks is not very well-determined. In practice, this FSR is much more important than the exact position of the peaks. The fabricated device will always differ from the designed one by some tolerance, which will also shift the peaks. The FSR, however, is influenced much less by these fabrication errors, and will thus correspond quite well with the calculated one. It is also the more important one, since it determines what wavelength channels can be filtered. The exact position of the resonances can easily be tuned, while the FSR cannot be tuned by a large amount, and must thus be known in advance.

For this particular problem, the FDTD is a more suitable method, since one is interested in the behaviour of the light in a range of wavelengths. It could be argued that it is best to use the FDTD to find the position of resonances and then calculate the field at those resonances using the Helmholtz solver, since it is more accurate at the same grid size. However, the two methods yield results that differ too much; the resonance peak calculated with one method is not located at the same position as the one calculated with the other method. So, it is best to use FDTD for both the spectral response and the fields at single frequencies, if those frequencies are chosen from the spectral response.

5.2 Photonic crystal structures

This section is an adaptation of reference [Stoffer et al 2000].

Photonic bandgap crystals are a promising class of structures for tight photonic integration. They consist of a regular lattice of dielectric materials with a high refractive index compared to the rest of the space. When the index contrast is high enough, a range of wavelengths exists for which propagation is forbidden in all directions, the so-called photonic band gap [Joannopoulos et al. 1995]. Calculations show that it should be possible to make very tight bends in waveguides in these crystals, with a radius in the order of the wavelength, with (nearly) 100% transmission [Mekis et al. 1996, Baba et al. 1999]. Other devices like ultracompact wavelength filters are also possible [Fan et al. 1998, Sigalas et al. 1993, Mekis et al. 1998, Scherer et al. 1998].

This section will focus on a particular 2-dimensional photonic bandgap crystal. It consists of a square lattice of square rods of silicon in air. A simple method to determine the bandgap will be presented in Section 5.2.1.

When one row of rods is removed, this line defect can act as a waveguide [Sakoda et al. 1997, Benisty 1996, Centeno and Felbacq 1999], which we will call a crystal waveguide. For the considered lattice, this turns out to be a monomodal waveguide for frequencies within the lowest bandgap. However, this mode does not behave in the same way as dielectric waveguide modes. Consider such a wave that is guided in the z-direction along a dielectric waveguide. Such a wave has a field distribution in the x-y plane (the mode profile) which is independent of the z-coordinate, and a single effective index. On the other hand, waves guided along crystal waveguides show mode profiles and phase velocities corresponding to the different Floquet modes. For such a mode an effective index corresponding to the first Brillouin zone can be determined numerically. From this wavelength dependent effective index the group velocity can be evaluated. Results of calculations on these quantities, and also on the effective modal width, will be presented in Section 5.2.2.

In practical applications, the coupling efficiency between a photonic crystal waveguide and a conventional dielectric waveguide (e.g. an optical fiber) is an important issue. In order to gain some insight into this subject, we investigated the coupling between a dielectric slab and a crystal waveguide. The results are discussed in Section 5.2.3.

In Section 5.2.4, an ultra-compact filter that is obtained by coupling a cavity to a waveguide will be described. When a row of rods is partially removed, so that a waveguide is formed that terminates in the crystal, a cavity is formed that acts as a resonator. If such a resonator is attached as a side-branch to a crystal waveguide, similar to a tuned stub in microwave engineering, a structure results that acts as a wavelength filter. [Danglot et al. 1998] have shown this effect for metallic photonic crystals in the microwave regime, but to our knowledge it has not been studied in dielectric photonic crystals. Increasing the quality factor Q of the cavity by not removing the rod directly adjacent to the crystal waveguide will result in a filter

having a smaller bandwidth. Such high-Q cavities show a large field enhancement at resonance, which may be exploited for non-linear optics applications.

In photonic crystal waveguides, the PML boundaries must be made thicker than usually. These thick PML's are used throughout the chapter; in Section 5.2.6, the reason behind this is explained.

5.2.1 Photonic crystal structure and bandgap calculation

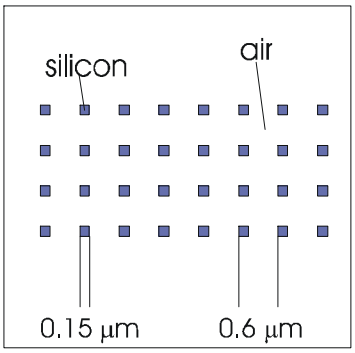


Figure 5.9: Two-dimensional photonic crystal structure composed of square silicon rods in air

We considered a crystal as shown in Figure 5.9, consisting of a square lattice of silicon rods in air, having a lattice constant of 600 nm. The rods have a 150x150 nm square cross-section and a refractive index of 3.4 (assumed to be constant in the wavelength range of interest, around 1550 nm). For a very similar structure [Joannopoulos et al. 1995], a bandgap is known to exist for wavelengths around λ=1.5 μm for TE polarization (i.e. E-field parallel to the rods, perpendicular to the calculation window). A similar structure has been realized in silicon by [Zijlstra et al, 1999].

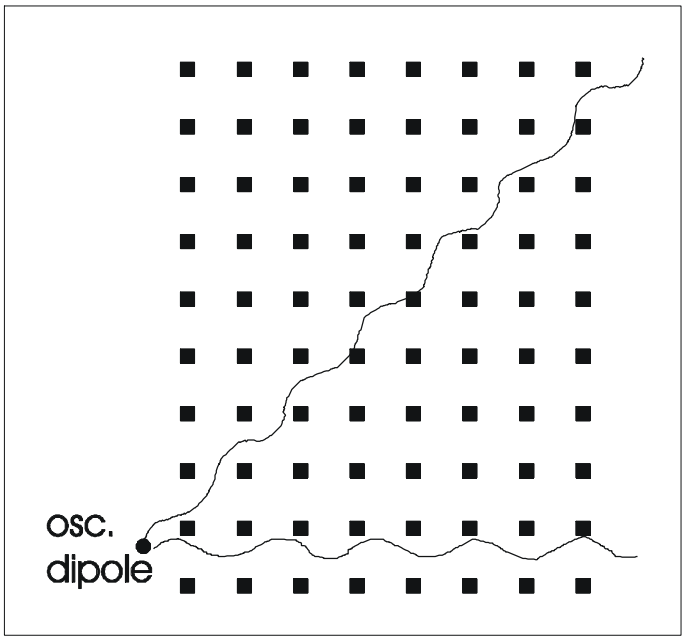


Figure 5.10: Configuration for determining the bandgap: A dipole oscillator is positioned at the left-hand side of the crystal, while the transmission is determined along a vertical line at the right-hand side

A straightforward method for determining the bandgap is to calculate the transmission of a broad-band electromagnetic signal through the crystal for all directions in the first Brillouin zone, which may be reduced for symmetry reasons [Joannopoulos et al 1995]. For that purpose, the 2-dimensional finite difference time domain (FDTD) scheme described in Chapter 4 is used. The spatial stepsize used is 25 nm. As shown in Figure 5.10, a finite crystal of 8 rods width is defined. To the left of this, a dipole is oscillating having a central frequency that corresponds to a wavelength of 1.5μm; the pulse is chosen short enough (50 fs) to contain all wavelengths in the bandgap. This dipole will radiate in all directions. The Fourier transform of the time dependence of the transmitted field at the right-hand side of the crystal for a range of propagation angles between 0 and 45 degrees is taken. This covers the reduced Brillouin zone of the square lattice, and yields the transmitted amplitude for a sufficiently wide wavelength range around the bandgap.

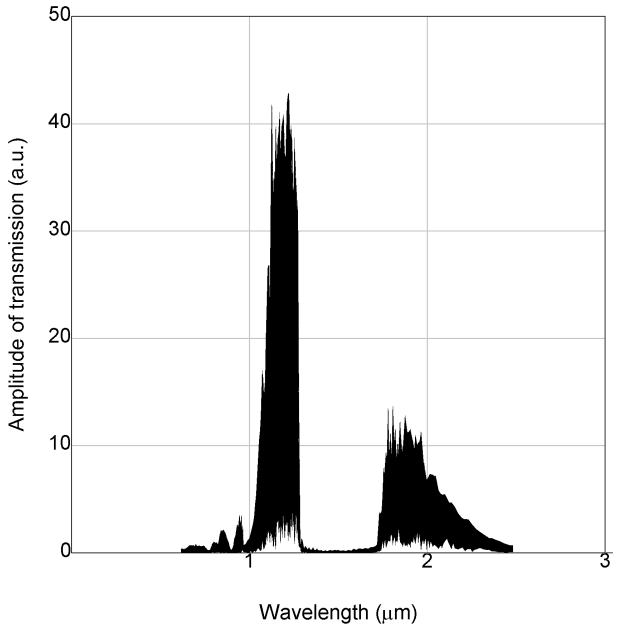


Figure 5.11: Superposition of the transmission amplitudes over all relevant angles. The low area between ~1.3 and 1.7 μm is the bandgap for TE polarization.

Figure 5.11 shows a superposition of the graphs, for 100 uniformly distributed propagation angles between 0 and 45 degrees, of the transmitted amplitude versus the wavelength. In the range from approximately 1.3 μm to 1.7 μm, the transmission is nearly zero for all directions, revealing the photonic bandgap.

While there are other methods [Joannopoulos et al. 1995], which use a single unit cell for the band-structure calculation, the presented method is fast and gives a good approximation of the total bandgap of the crystal, nearly equal to the bandgap for the similar structure studied in [Joannopoulos et al. 1995].

5.2.2 Crystal waveguides

Removing a row of rods creates a channel through which light can propagate, a so-called crystal waveguide. Such a waveguide can also be seen as two semi-infinite crystals brought into proximity. We have found from numerical calculations that this particular waveguide supports one symmetrical mode for each wavelength inside the bandgap. In the calculations, the mode is excited through a thick Perfectly Matched Layer. Later on in this chapter, in Section 5.2.5, it will be made clear that a thick PML is necessary; we use a 3.2 μm wide PML. Since only the waveguide mode can propagate in the structure, it does not matter much exactly what input field is generated; after some distance, only the crystal waveguide mode will be present, so we simply launch a Gaussian-shaped beam.
In integrated optics, waveguide modes are usually characterized by their effective index, which is directly related to the phase velocity along the waveguide. In order to determine the effective index of crystal waveguide modes, we applied the two

different calculation methods: Finite Difference Time Domain (FDTD) and the omnidirectional Helmholtz solver. Light is propagated through a waveguide with a length of about 25 lattice periods. In the Helmholtz solver, the effective index was calculated directly from the phase velocity. In the FDTD method, the waveguide was terminated after a number of rods, creating a mirror, since beyond that point, only the uniform crystal remains, in which no light can propagate for this wavelength. From the interference pattern in the waveguide before the mirror, the effective wavelength can be determined and from that the effective index. Figure 5.12 shows the result for both methods.

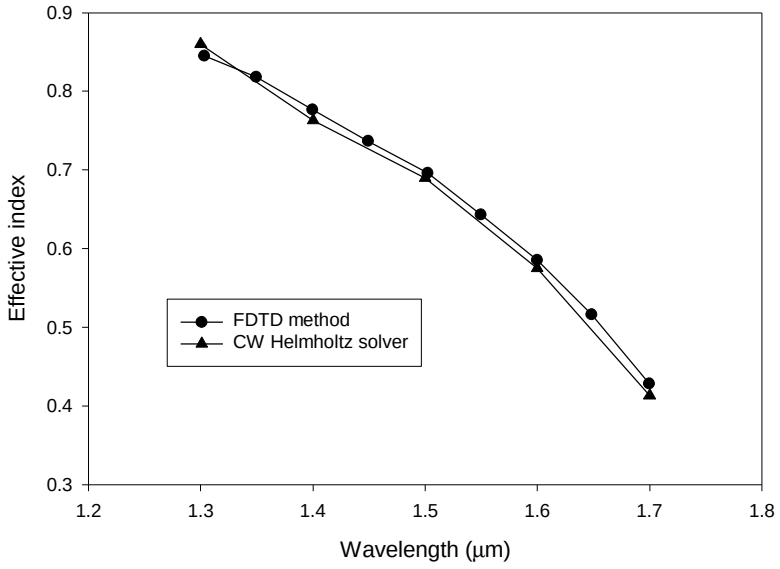


Figure 5.12: Effective index of crystal waveguide, calculated by the two different methods, as a function of wavelength.

The effective index is lower than one, which means that the phase velocity is larger than the speed of light. In dielectric waveguides, the effective index is also lower than the index of the core layer. Since the core layer in a crystal waveguide is air, with a refractive index of one, this means that the effective index must be lower than one. This is similar to the case of modes guided between parallel conducting plates.

In order to determine the group velocity, a third-order polynomial is fitted to the effective index values obtained from the FDTD calculations, and the following formula is used:

$$v_{group} = \frac{\partial w}{\partial k_x} = \frac{c}{n_{eff} - l\,\dfrac{\partial n_{eff}}{\partial l}} \tag{5.6}$$

This gives the group velocity graph shown in Figure 5.13. The group velocity is smaller than the speed of light in vacuum and decreases with increasing wavelength.

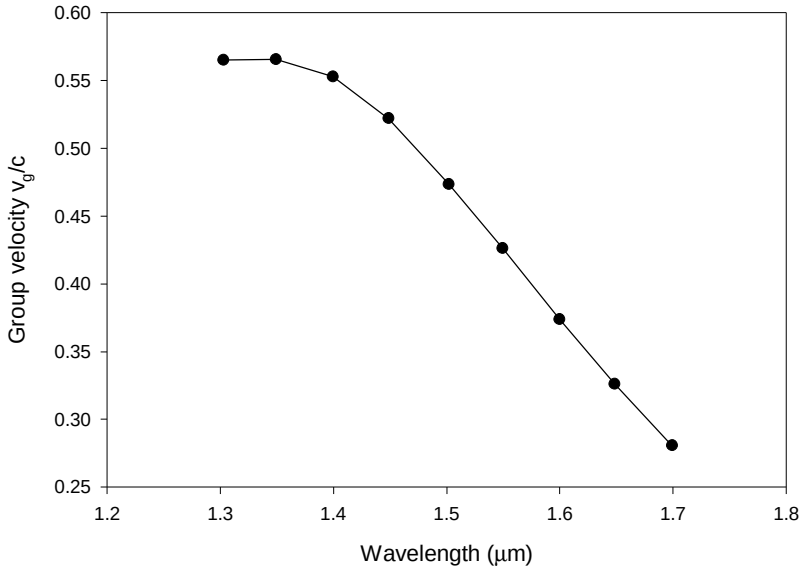


Figure 5.13: Group velocity of crystal waveguide mode, as a function of wavelength.

Another important practical property of a waveguide mode, be it a crystal or a dielectric waveguide, is its effective width. For dielectric waveguides, this is calculated by taking into account the Goos-Hänchen shift on the interfaces [Kogelnik 1979]; however, for crystal waveguides this shift cannot be readily defined. First of all, the width of the waveguide is not well-defined since the structure is not invariant in the propagation direction. One may choose the width across which the index is constant; in this case, that width is 1.05 μm. Another choice is the lattice constant, since a single row of rods was removed; then, the width is 0.6 μm. An effective width, however, can be defined for the lowest-order mode as:

$$w = \frac{l}{2\sqrt{n_{core}^2 - n_{eff}^2}} \tag{5.7}$$

corresponding to the width of a parallel-plate waveguide with perfectly conducting walls that supports a lowest order TE mode with the same effective index. This effective width is an important factor when coupling light from free space or from another waveguide, since the overlap between the incoming field and the mode (and thus the coupling efficiency) is high when the effective width is equal. The thus obtained effective width as a function of the wavelength is given in Figure 5.14.

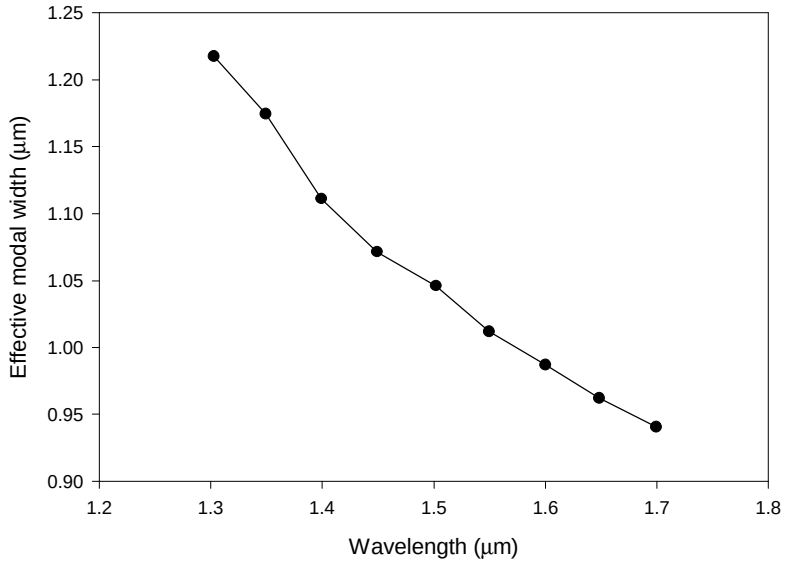


Figure 5.14: Effective modal width of crystal waveguide mode, as a function of wavelength.

For the crystal waveguide mode, the calculated effective width lies between 0.95 and 1.22 μm. When defining the physical width of the crystal waveguide, a choice of 0.6 μm would be the natural one, since one row of rods is removed, and the lattice constant is 0.6 μm.
Crystal waveguide modes share many properties with those of standard dielectric waveguides. The effective index is smaller than the index of the core material, the group velocity is smaller than the speed of light. If the width of the crystal waveguide is defined as the lattice constant, it is smaller than the effective modal width, as is the case in dielectric waveguides. As mentioned in the introduction, a big difference between dielectric and crystal waveguides is the fact that the transverse shape of the field is not constant along the propagation direction in the case of the crystal waveguide; it is periodic, with the same period as the period of the lattice.

5.2.3 Coupling from a dielectric waveguide to a crystal waveguide

In this section, the coupling of light from a normal dielectric waveguide to a crystal waveguide will be investigated. The structure is given in Figure 5.15.

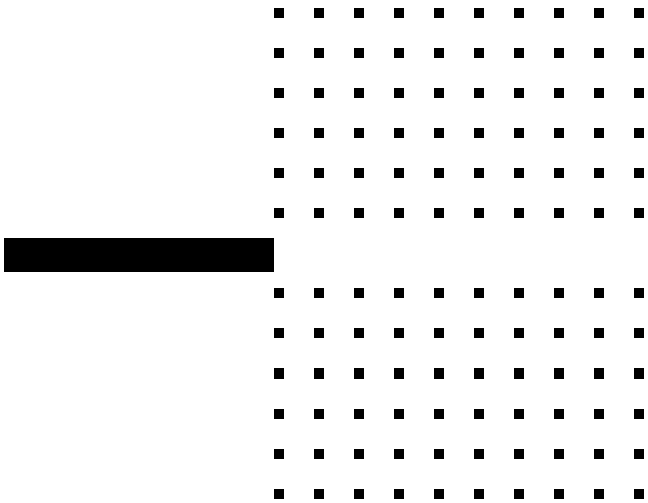
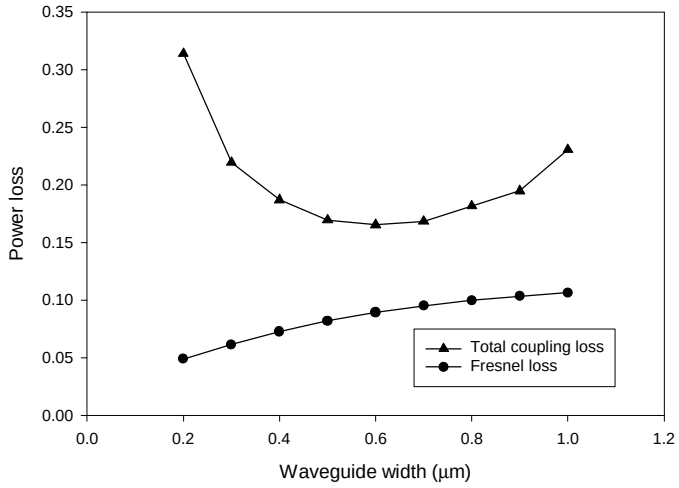


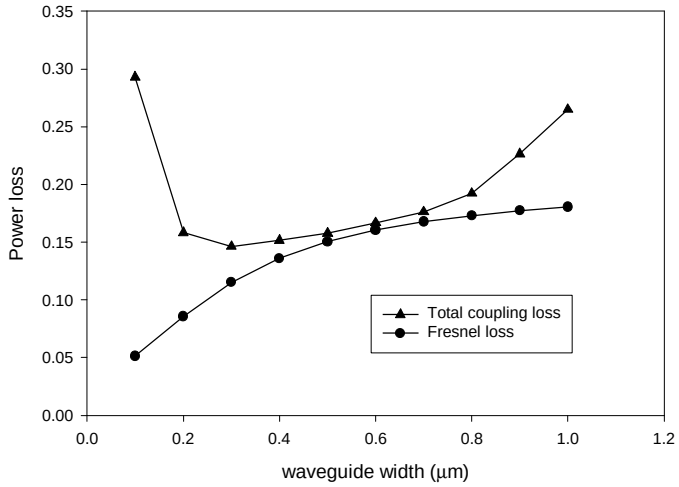
Figure 5.15: Structure used for coupling from dielectric slab wagevuide to crystal waveguide. The index of the slab (1.45 or 1.8, as used in Figures 5.16 and 5.17) is different from that of the rods (3.4).

We will only consider the central wavelength of the bandgap, λ=1.5 μm. For this wavelength, the effective width of the crystal waveguide as calculated in the previous section is 1.05 μm. We chose to investigate two different materials for the input waveguide: one with a refractive index of 1.45, and one with a refractive index of 1.8. The Fresnel losses, caused by the difference in effective index between the two waveguides, are lowest for the low-index input waveguide; however, the effective modal width of the higher-index waveguide can be smaller. The reflection was calculated using the Helmholtz solver, since only a single frequency is considered.

Figures 5.16a and 5.16b show the total coupling loss and the theoretical Fresnel losses of both structures. In Figure 5.17, the effective modal width of the incoming waveguides is given. In approximation, the difference between the Fresnel losses and the total coupling loss is the extra loss caused by the mismatch of the modal field. It is difficult to draw conclusions from these results. For the waveguide with the high refractive index, the extra loss is smallest around d=0.6 μm. The effective modal width of the dielectric guide is around 1 μm, which is close to the effective modal width of the crystal waveguide. This seems to tell us that the effective modal width as calculated in the previous section is a useful tool when analyzing the coupling from other waveguides or free space into crystal waveguides. However, the other waveguide, with the refractive index of 1.45, does not comply with this theory; the extra coupling loss is lowest at a width of around 0.6 μm, where the effective modal width is not closest to the effective width of the crystal waveguide mode.



a



b

Figure 5.16: Total coupling loss and theoretical Fresnel loss caused by coupling from a dielectric waveguide with index 1.45 (a) and 1.8 (b) to a crystal waveguide, at a wavelength of 1.5 μm, as a function of the dielectric waveguide width.

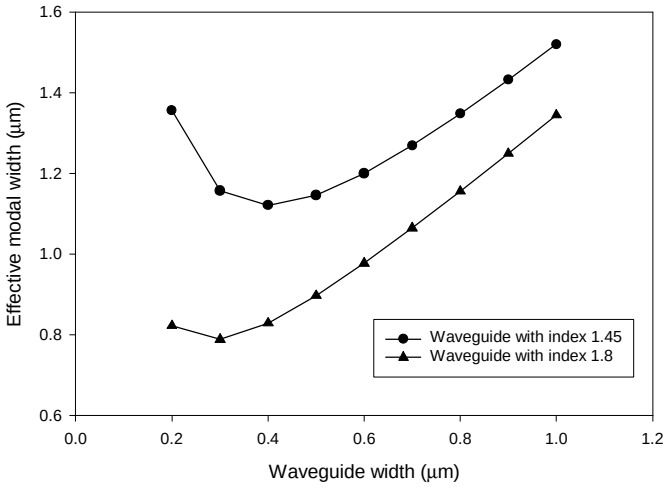


Figure 5.17: Effective modal width for dielectric waveguides at 1.5 µm wavelength, as a function of the physical waveguide width.

As anticipated, the simple modal width picture is not sufficiently accurate to predict the coupling of dielectric waveguides to crystal waveguides. This may be because the crystal mode is not the same at the entrance of the photonic crystal as it is deeper into the waveguide. Also, the mode consists of a set of Floquet modes, all of which should be considered together.

It would be very interesting to examine how this coupling loss may be further minimized. The following are possibilities for improvements:

- Changing the refractive index of the dielectric waveguide to a value between 1.45 and 1.8
- Moving the end of the dielectric waveguide into the crystal guide
- Changing the size or the index of the rods near the entrance of the crystal waveguide

5.2.4 An ultracompact wavelength filter

In literature, one can find several examples of wavelength filters in photonic bandgap materials, based upon resonant effects in cavities. Most of these filters transmit a small range of wavelengths while reflecting the larger part of the spectrum [Sigalas et al. 1993]. Also, an add-drop filter has been shown in which the index or the size of single rods must be tuned to achieve proper operation [Fan et al. 1998]. We will show a filter based upon waveguide side-branches, which transmits most wavelengths, while reflecting a number of small wavelength ranges. This type of filter is also known as a notch filter.

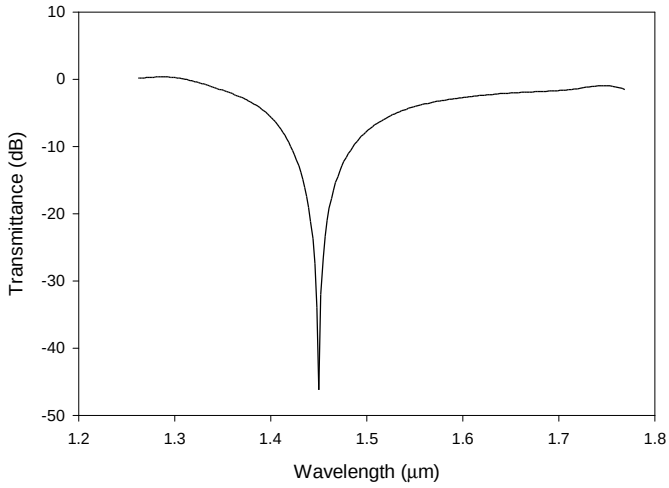


Figure 5.18: Wavelength dependence of transmission through a crystal waveguide with a defect (missing rod) located directly adjacent to it.

The first observation of a filter function arises when a single rod directly adjacent to the crystal waveguide is removed. This will start acting as a cavity; for a certain wavelength, all light will be reflected into the input channel. This filter characteristic is shown in Figure 5.18. Increasing the side-branch length to 3 rods, as shown in Figure 5.19, gives a narrower filter characteristic, as shown in Figure 5.20. This is to be expected, since the free spectral range decreases while the finesse of the cavity should remain more or less constant, since the finesse mainly depends on the loss per roundtrip in the cavity, which is constant.

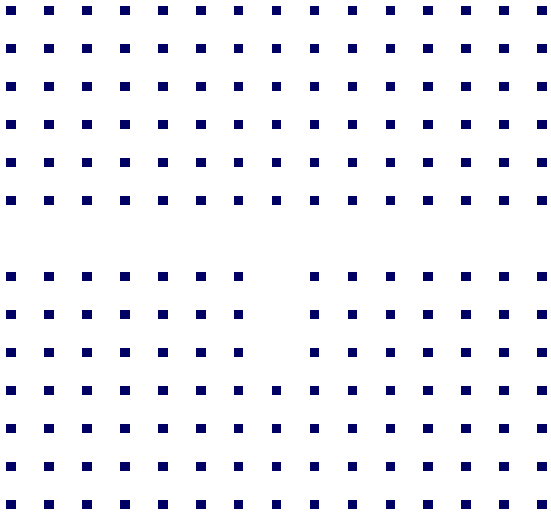


Figure 5.19: Crystal waveguide with q 3-rods-deep side-branch

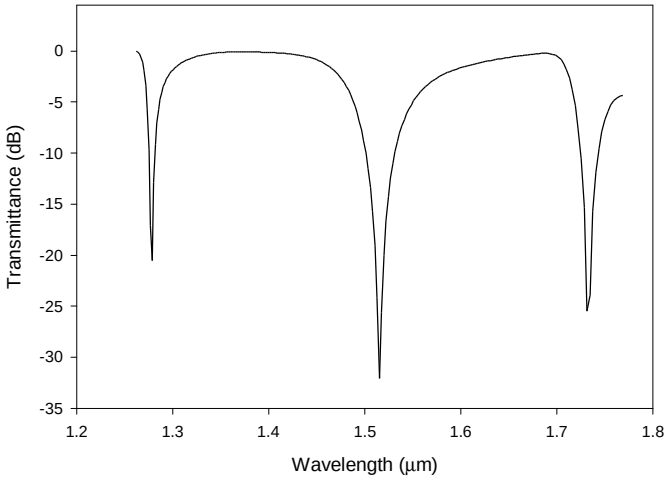


Figure 5.20: Filter characteristic of the structure shown in Figure 5.19; transmission as a function of wavelength.

The filter function can be explained by means of scattering matrix theory [Haus 1982]. The spot where the branch connects to the main waveguide is considered as a node with three input and three output ports, as drawn in Figure 13. Light coming in from an input port will be distributed across the three output ports according to the following matrix:

$$\begin{pmatrix} A_1^{out} \\ A_2^{out} \\ A_3^{out} \end{pmatrix} = S \cdot \begin{pmatrix} A_1^{in} \\ A_2^{in} \\ A_3^{in} \end{pmatrix} \qquad S = \begin{bmatrix} r_1 & t_1 & s_1 \\ t_1 & r_1 & s_1 \\ s_1 & s_1 & r_3 \end{bmatrix} \tag{5.8}$$

with

$$S = S^T = S^{*-1} \tag{5.9}$$

in which the vectors contain the amplitudes of the incoming and outgoing modes. Due to the symmetry of the system, some of the elements of the matrix are the same.

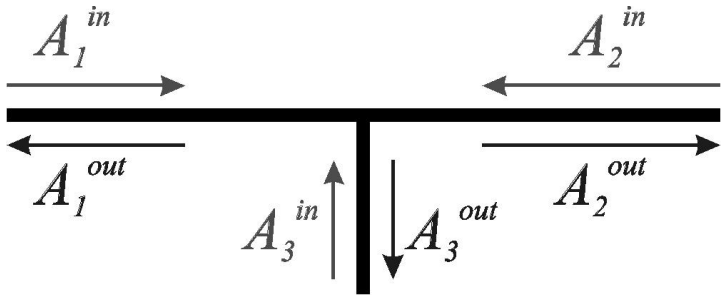


Figure 5.21: 3-port representation of the crystal waveguide with a side-branch

When a perfect mirror with a wavelength-dependent phase-shift is introduced at the output of port 3, while only having input into port 1, the input of port 3 can be calculated as follows:

$$A_3^{in} = e^{ij\,(l)} A_3^{out} \tag{5.10}$$

$$A_3^{in} = e^{ij\,(l)} A_1^{in} s_1 \left(1 + r_3 e^{ij\,(l)} + r_3^2 e^{2ij\,(l)} + \ldots\right) = \frac{e^{ij\,(l)} A_1^{in} s_1}{1 - r_3 e^{ij\,(l)}} \tag{5.11}$$

so, using (5.8) and (5.11), the output amplitude of port 2 is:

$$A_2^{out} = t_1 A_1^{in} + s_1 A_3^{in} = A_1^{in} \left(\frac{t_1 - t_1 r_3 e^{ij\,(l)} + s_1^2 e^{ij\,(l)}}{1 - r_3 e^{ij\,(l)}} \right) \tag{5.12}$$

Using the fact that $S^* = S^{-1}$ [Haus 1982], it may be proved that

$$-t_1 r_3 + s_1^2 = t_1^* \equiv t_1 e^{iq\,(l)} \tag{5.13}$$

Hence

$$A_2^{out} = A_1^{in} t_1 \frac{1 + e^{ij\,(l) + q\,(l)}}{1 - r_3 e^{ij\,(l)}} \tag{5.14}$$

So, for certain values of λ, the numerator of this formula becomes zero, yielding zero transmission and full reflection.

This configuration is similar to a transmission line with a parallel short-circuited stub.

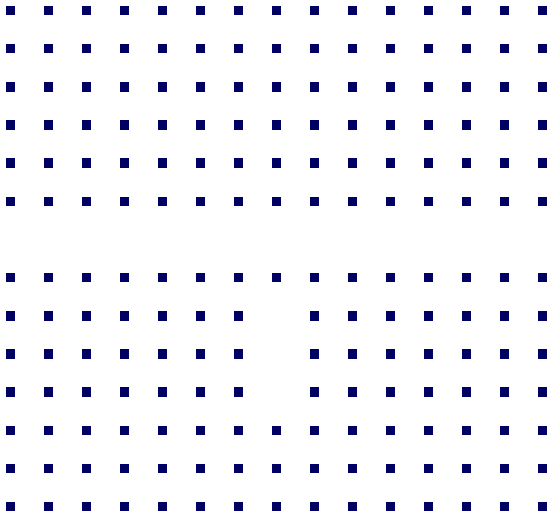


Figure 5.22: Crystal waveguide with a side-branch with higher reflection coefficient at the junction, creating a high-Q cavity

The device described above has a filter characteristic with dips in the transmission (and peaks in the reflection) that may be too wide for some applications. The bandwidth of the filter is directly related to the finesse of the resonating cavity, which is dependent on the reflection coefficients on either side of the cavity. At the far side of the cavity, the reflection is 100%; at the junction, it is only approximately 33%. By moving the cavity down one rod and introducing an extra rod as shown in Figure 5.22, the reflection coefficient at the junction increases dramatically, while still retaining the symmetry of the system. This leads to the response curve in Figure 5.23. Using the same arguments as before, the transmission can be shown to be zero for certain wavelengths.

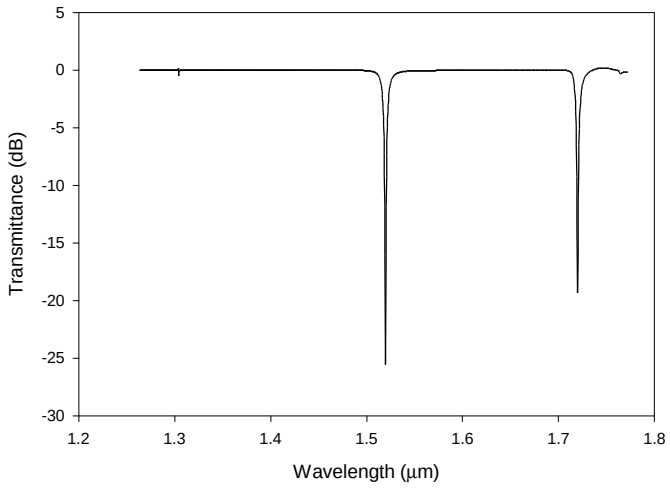


Figure 5.23: Filter characteristic of the structure shown in Figure 5.22; transmission as a function of wavelength.

A high-Q resonator like this builds up a high intensity in the cavity. This high intensity could be used for nonlinear applications. In this cavity, the intensity is about 20 times larger than the input intensity.

5.2.5 PML's for photonic crystal waveguides

As was seen in the previous chapters, thin Perfectly Matched Layers (PML's) perform well as nonreflecting boundary conditions for e.g. waveguides or free space. A PML with a width of 0.8 μm can give a reflectivity of less than 10^{-5}. However, for crystal waveguides, thin PML's don't perform well anymore. This is because the crystal waveguide mode is very dependent on the structure of the crystal in its vicinity; in the previous section, is was seen that one missing rod can give complete reflection of the crystal waveguide mode. If the PML is very thin, and therefore very strong if the absorption in the traversal of the PML is kept constant, there is no interaction of the light with rods after the onset of the PML; it is like the photonic crystal structure is terminated at the PML.

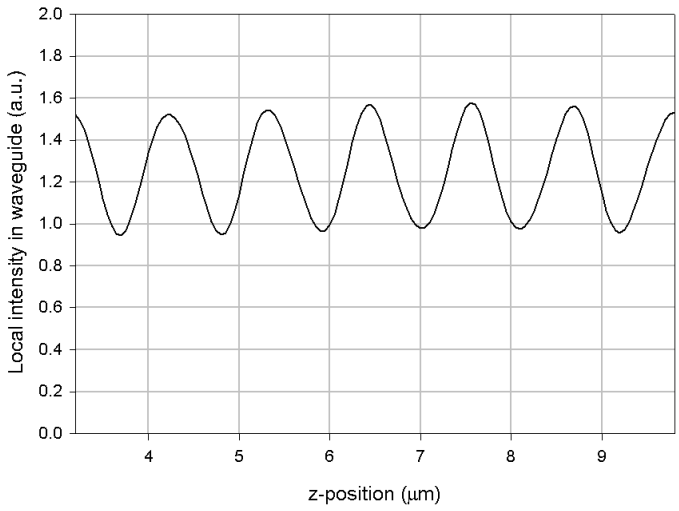


Figure 5.24: Local intensity in the crystal waveguide; a 0.8 μm-wide PML is used. The used calculation window is 13 μm long, including the PML regions.

Figure 5.24 shows the local intensity in the center of the crystal waveguide for a PML with a width of 0.8 μm. There is a large modulation of the intensity that does not have the periodicity of the crystal, corresponding to a power reflection at the boundary of ~1.4%. The modulation itself also seems to be modulated with a larger period; this is likely due to the interplay between the modulation due to reflection and the modulation that exist due to the fact that the Bloch crystal mode is composed of multiple Floquet modes, with the same period as the crystal.

Figure 5.25 shows the intensity in a crystal waveguide that is terminated at z=8 μm. Here, the power reflection is 4.8%. This shows that the effects of a thin PML are similar to those of the termination of the crystal. In order to drive down the reflection, the PML must be made much thicker.

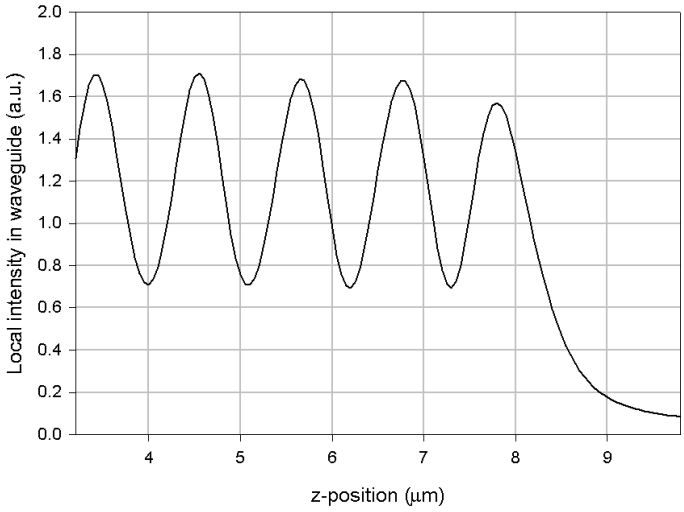


Figure 5.25: Local intensity in the crystal waveguide. After z=8 µm, the crystal is terminated and the light propagates in free space.

Figure 5.26 shows the intensity with a PML with a width of 3.2 µm. The modulation of the intensity mainly has the same periodicity as the crystal (0.6 µm), which is caused by the Floquet modes of the system. There is still some modulation on top of this with a lower period, which is caused by a counter-propagating mode; the reflectivity caused by this thick PML, however, is lower than 10^{-4}.

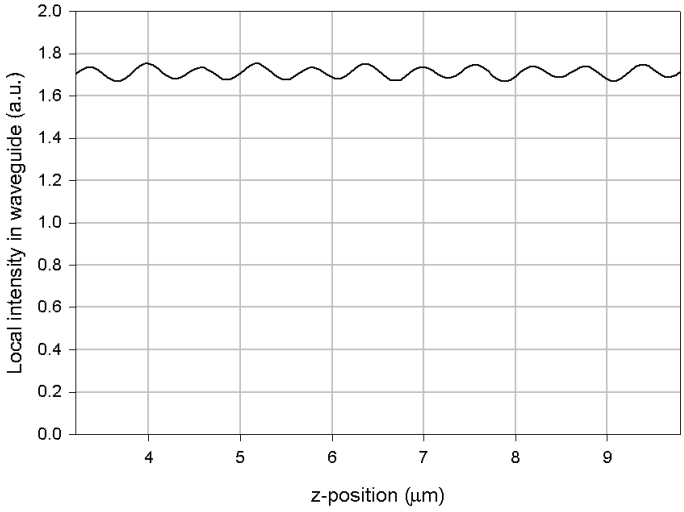


Figure 5.26: Local intensity in the crystal waveguide. A 3.2 µm-wide PML is used.

5.2.6 Conclusions

This section presented the results of some calculations on waveguides in a 2-dimensional photonic bandgap crystal. For such a configuration to be practical, the light will have to be confined in the third dimension, which might be hard to achieve. However, we feel that the concepts presented here can be used in the description, calculation and design of general photonic crystal devices that have confinement in the third dimension.

The used photonic crystal is a square lattice of silicon rods in air. A simple method was used to determine the bandgap. Removal of a row of rods creates a monomodal crystal waveguide for all wavelengths in the gap. The effective index of these modes is lower than one, but, as it should, the group velocity is lower than the speed of light in vacuum. An effective modal width is introduced, which is found to decrease with increasing wavelength. This effective modal width may play an important role in the losses on coupling dielectric waveguides to crystal waveguides. It is shown that, for the considered photonic structure, such coupling losses can be as low as approximately 15%. The Fresnel losses, caused by the effective index difference between the dielectric waveguide mode and the crystal waveguide mode, are an important factor; in the coupling from a waveguide with a higher refractive index, these losses are the dominant part of the total loss.

In normal waveguide-to-waveguide coupling, the effective modal width is a powerful design parameter, which can be used to minimize the coupling loss approximately. The results indicate that this picture holds only approximately for coupling to a crystal waveguide, and that the physics is too complicated for the coupling to be described purely in terms of effective modal widths and Fresnel losses. The coupling should be described using all crystal modes, including all of their relevant Floquet components.

When a single rod is removed next to a waveguide, this will create a filter characteristic, which means that for some wavelengths no light will be transmitted. As a consequence, waveguides in photonic crystals are quite sensitive to technological imperfections of this kind. One can use this phenomenon to design an ultracompact filter. The wavelength characteristic of the filter is determined by the length of the side-branch (which affects the free spectral range) and the modal reflection coefficients of the junction point. We have shown a filter that blocks a wavelength range of approximately 4 nm. Cavities of this kind may be used for intensity enhancement.

Perfectly matched layers do not work as well for crystal waveguides as they do for normal, dielectric waveguides, due to the sensitivity of the crystal waveguide mode to perturbations of the crystal. One needs thick PML's (~5 lattice periods) in order to decrease the reflectivity to an acceptable level. Another way to solve this problem might be by not continuing the periodic structure inside the PML, but substituting the crystal by a structure of which the index is constant in the direction perpendicular to the boundary, and suitably chosen parallel to the boundary; e.g. the average index of the lattice on each x-position [Mekis et al 1996].

Chapter 6: Conclusions and Summary

This thesis presents several improvements on simulation methods in integrated optics, as well as some new methods. Both uni- and omnidirectional tools are presented; for the unidirectional methods, the emphasis is on higher-order accuracy; for the omnidirectional methods, the boundary conditions are extremely important, and care must be taken with the incoming field definition. The omnidirectional methods are applied to two different types of structures, a cylindrical microcavity and photonic bandgap crystals.

Two different methods are developed to obtain fourth-order accurate unidirectional beam propagation methods. Interfaces between materials at arbitrary positions within grid cells are taken into account, so there are no restrictions the positioning of the discretization grid. The first method applies Taylor decompositions of the fields around interfaces to obtain corrections on the discretization of the SVEA Helmholtz equation. It yields very simple and numerically efficient formulas for TE polarization, in which the only correction terms are on the diagonal of the matrices. In the second method, a finite difference scheme is used in the propagation direction and a variational method in the transverse direction. This variational approach causes the implementation for TE and TM to be very similar. Both second-order accurate and fourth-order accurate methods are shown. The variational approach leads to a norm differing from the standard dot product. The error made in the effective index converges as $O(\Delta x^2)$ or $O(\Delta x^4)$ for the 3- and 5-point operator, respectively. The increased accuracy of the fourth-order scheme decreases the amount of grid points that is needed, so the total computation time and memory usage can be much smaller at a given accuracy. At the boundary, Perfectly Matched Layers (PML's) are shown to be more robust nonreflecting boundaries than Sommerfeld-type boundary conditions.

In the presence of PML's the analytical field solutions correspond to light that is attenuated in these layers without any backreflection. Numerically, some care must be taken to attain this situation. For the two- or three-dimensional PML, the time-dependent equations must be split in order to have directional attenuation and to have zero reflection at the interface for radiative and guided modal fields. In the time-harmonic case (Helmholtz equation), the PML corresponds to the addition of extra coefficients into the equation. For an incoming modal field, the lateral field distribution does not change in the PML; however, the amplitude of the modal field decreases as the mode traverses the PML. A very important property of modal propagation in the PML is the fact that the attenuation of a modal field depends on its propagation constant β.

An omnidirectional solver for the 2D TE and TM Helmholtz equation is developed. It is based upon the variational formalism, in which the solution of the equation is found by finding extremal points of a functional. The problem is reformulated: Instead of discretizing the wave equation, the functional is approximated by an interpolation using the fields on grid points, and the index distribution. The index distribution within a grid cell can be taken into account. Because of this, it is possible to do the calculations accurately with a rectangular grid that is independent of the structure, so one does not need complicated grid generation. The calculation window is surrounded by Perfectly Matched Layers in order to have nearly reflectionless boundaries.

Incoming fields are necessarily defined on the outer boundary, outside the PML. Those fields must first traverse the PML before they enter the interior of the window, which means that they are attenuated by an amount that is related to the component of the wave vector perpendicular to the boundary. If a field is composed of multiple modes, each of these modes will have its own wave number perpendicular to the boundary. Therefore, the amplitude of each mode must be adjusted independently. Care must be taken with the strength and shape of the PML in order to obtain little numerical reflection and accurate amplitudes of the input fields. The amplitude of incoming fields will always have a significant error, so in general, two runs will be necessary for accurate simulations: One with only the incoming waveguide and no scattering structure from which the amplitude of the incoming light as simulated can be deduced.

In the Finite Difference Time Domain method, the same boundary conditions are used, so the same care must be taken with the incoming field generation and the strength and width of the PML. In addition, the calculations are performed with full time dependence, so it is possible to simulate multiple frequencies at once. In order to calculate the response of a structure to modal input fields, the modal field for each calculated frequency has to be different. A way to generate so-called modal pulsed is developed and works well.

The Helmholtz solver has the advantage of being able to be more accurate with a coarser grid, since interfaces through grid cells are taken into account accurately. However, on finer grids, the calculation time that the Helmholtz solver needs are much higher than those of the FDTD method for the same problem, since the Helmholtz solver scales worse than N^2, while the FDTD scales as $N^{3/2}$ or better. Due to the additional advantage of the FDTD method that it can calculate for multiple frequencies at once, it is a more suitable method for most applications.

Both the Helmholtz solver and the FDTD method are shown to be appropriate tools to simulate small microcavity structures. For the same grid size, the Helmholtz solver is more accurate due to its better treatment of slanted interfaces. However, the calculation time for a wavelength scan is orders of magnitude higher for the Helmholtz solver than for the FDTD method. For the cylindrical microcavity, the free spectral range (FSR) and the finesse are calculated and shown to comply well with another theory. Due to numerical dispersion, the exact position of the resonances is a slowly converging function of the stepsizes. However, in practice, the FSR and the finesse are the truly important parameters, they do converge rapidly.

For calculations on photonic bandgap crystals, the same two methods are applied. The following results are obtained:

- The bandgap is calculated.
- The waveguide dispersion of a crystal waveguide is shown.
- Some estimations of the efficiency that may be expected in the coupling of light from a normal dielectric waveguide to a crystal waveguide are made. As expected, comparisons of the results with a simple picture involving only the modal width of the dielectric waveguide and the effective modal width of the crystal waveguide show clear differences.
- A cavity near a crystal waveguide creates a notch filter characteristic.
- Nonreflecting boundaries in crystal waveguides require a thicker PML, since there is significant reflection from thin PML's in these structures.

References

Baba, T., Fukaya, N., Yonekura, Y., *Observation of light propagation in photonic crystal optical waveguides with bends*, Electr. Letters **35** (8), pp. 654-655, 1999

Balistreri, M. L. M., Klunder, D. J. W., Blom, F. C., Driessen, A., Hoekstra, H. J. W. M., Korterik, J. P., Kuipers, L., Hulst, N. F. van, *Visualizing the whispering gallery modes in a cylindrical optical microcavity*, Optics Letters **24** (24), pp. 1829-1831, 1999

Benisty, H., *Modal analysis of optical guides with two-dimensional photonic band-gap boundaries*, J. Appl. Phys. **79** (10), pp. 7483-7492, 1996

Bérenger, J.P., *A perfectly matched layer for the absorption of electromagnetic waves*, Journ. of Comp. Phys. **114**, pp. 185-200, 1994

Blom, F. C., Dijk, D. R. van, Hoekstra, H. J. W. M., Driessen, A., Popma, Th. J. A., *Experimental study of integrated-optics microcavity resonators: Toward an all-optical switching device*, Appl. Phys. Lett. **71** (6), pp. 747-749, 1997

Boisvert, F., *A Fourth-Order Accurate Fourier Method for the Helmholtz Equation in Three Dimensions*, ACM Trans. on Math. Softw. **13** (3), pp. 221-234, 1987

Centeno, E., Felbacq, D., *Guiding waves with photonic crystals*, Optics Communications **160**, pp 57-60, 1999

Chin, M. K., Ho, S. T., *Design and Modeling of Wavequide-Coupled Single-Mode Microring Resonators*, Journ. of Lightw. Techn. **16** (8), pp. 1433-1446, 1998

Danglot, J., Carbonell, J., Fernandez, M., Vanbésien, O., Lippens, D.,*Modal analysis of guiding structures patterned in a metallic photinic crystal*, Appl. Phys. Lett. **73** (19), pp 2712-2714, 1998

Fan, S., Villeneuve, P. R., Joannopoulos, J. D., Haus, H. A., *Channel drop tunneling through localized states*, Phys. Rev. Letters **80** (5), pp 960-963, 1998

Fan, S., Appelbaum, I., Joannopoulos, J. D., *Near-field scanning optical microscopy as simultaneous probe of fields and band structure of phonotic crystals: A computational study*, Appl. Phys. Letters **75** (22), pp. 3461-3463, 1999

Hadley, G.R., *Transparent boundary condition for beam propagation*, Optics Letters **16** (9), pp. 624-626, 1991

Hagness, S. C., Rafizadeh, D., Ho, S. T., Taflove, A., *FDTD Microcavity Simulations: Design and Experimental Realization of Waveguide-Coupled Single-Mode Ring and*

Whispering-Gallery-Mode Disk Resonators, Journ. Of Lightw. Techn. **15** (11), pp. 2154-2165, 1997

Haus, H. A., *Waves and Fields in Optoelectronics*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1982

Heideman, R.G., Lambeck, P.V., *Remote opto-chemical sensing with extreme sensitivity: design, fabrication and performance of a pigtailed integrated optical phase modulated Mach-Zehnder interferometer system*, Sensors and Actuators B **61**, pp. 100-127, 1999

Hoekstra, H.J.W.M., Krijnen, G.J.M., Lambeck, P.V., J. Lightwave Technol. **10**, pp. 1352, 1992[a]

Hoekstra, H.J.W.M., Krijnen, G.J.M., Lambeck, P.V., *On the accuracy of the finite difference method for application in beam propagation techniques*, Opt. Comm. **94**, pp.506-508, 1992[b]

Hoekstra, H.J.W.M., Krijnen, G.J.M., Lambeck, P.V., Opt. Comm. **97**, pp. 301, 1993

Hoekstra, H.J.W.M., *Theory and Numerical Strategies of BPMs: On Beam Propagation Methods in Integrated Optics*, Opt. And Quant. Electr. **29**, pp. 157-171, 1997

Jinguji, K., *Synthesis of Coherent Two-Port Optical Delay-Line Circuit with Ring Waveguides*, Journ. of Lightw. Techn. **14** (8), pp. 1882-1898, 1996

Joannopoulos, J. D., Meade, R. D., Winn, J. N., *Photonic crystals: Molding the flow of light*, Princeton University Press, Princeton, New Jersey, ISBN 0-691-03744-2, 1995

Koch, T, *Laser sources for amplified and WDM lightwave systems*, in "Optical Fiber Telecommunications IIIB", I. Kaminov and T. Koch, Eds. San Diego: Academic Press, pp. 115-162, 1997

Kogelnik, H., *Theory of dielectric waveguides*, in *Integrated optics*, ed. T. TAMIR, Springer-Verlag New York Heidelberg Berlin, ISBN 0-387-09673-6, 1997

Little, B. E., Chu, S. T., Haus, H. A., Foresi, J., Laine, J.-P., *Microring Resonator Channel Dropping Filters*, Journ. Of Lightw. Techn. **15** (6), pp. 998-1005, 1997

Lui, W. W., Xu, C.-L., Hirono, T, Yokoyama, K., Huang, W.-P., *Full-Vectorial Wave Propagation in Semiconductor Optical Bending Waveguides and Equivalent Straight Waveguide Approximations*, J. of Lightw. Techn. **16**(5), pp. 910-914, 1998

Mekis, A., Chen, J., Fan, S., Villeneuve, P. R., Joannopoulos, J. D., *High transmission through sharp bends in photonic crystal waveguides*, Phys. Rev. Letters **77** (18), pp. 3787-3790, 1996

Mekis, A., Fan, S., Joannopoulos, J. D., *Bound states in photonic crystals and waveguide bends*, Phys. Rev. B **58** (8), pp. 4809-4817, 1998

Moore, T.G., Blaschak, J.G., Taflove, A., Kriegsman, G.A., *Theory and Application of Radiation Boundary Operators*, IEEE Trans. on Antenn. and Prop. **36** (12), pp. 1797-1811, 1988

Oda, K., Takato, N., Toba, H., Nosu, K., *A Wide-Band Guided-Wave Periodic Multi/Demultiplexer with a Ring Resonator for Optical FDM Transmission Systems*, Journ. of Lightw. Techn **6** (6), pp. 1016-1022, 1988

Onaka, H., Miyata, H., Ishikawa, Otsuka, K., Ooi, H., Kai, Y., Kinoshite, S., Seino, M., Nishimoto, H., Chikama, T., *1.1 Tb/s WDM transmission over a 150 km 1.3 micron zero-dispersion single-mode fiber*, PD-19, Vol. 2 of 1996 Technical Digest Series (Optical Society of America, Washington, DC), Optical Fiber Communication Conference, 1996

Pregla, R., *Method of Lines Based Beam Propagation Method*, in *Methods for Guided-Wave Optoelectronic Devices Part II: Waves and Interactions, Progress in Electromagnetics Research* (Piers 11), EMW Publishing, Editor: W.P. Huang, Cambridge Massachusetts, USA, 1995

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., *Numerical Recipes in Pascal: The Art of Scientific Computing*, Cambridge University Press, 1989

Rogge, U., Pregla, R., *Method of Lines for the Analysis of Dielectric Waveguides*, Journ. of Lightw. Techn. **11** (12), 1993

Rafizadeh, D., Zhang, J. P., Tiberio, R. C., Ho, S. T., *Propagation Loss Measurements in Semiconductor Microcavity Ring and Disk Resonators*, Journ. Of Lightw. Techn. **16** (7), pp. 1308-1313, 1998

Sakoda, K., Ueta, T., Ohtaka, K., *Numerical analysis of eigenmodes localized at line defects in photonic lattices*, Phys. Rev. B **56** (23), pp. 14905-14908, 1997

Scherer, A., Painter, O., D'Urso, B., Lee, R., Yariv, A., *InGaAsP photonic band gap crystal membrane microresonators*, J. Vac. Sci. Technol. B **16** (6), pp. 3906-3910, 1998

Schmidt, F., *Computation of discrete transparent boundary conditions for the 2D Helmholtz equation*, Opt. And Quant. Electr. **30**, pp.427-441, 1998

Sigalas, M., Soukoulis, C. M., Economou, E. N., Chan, C. T., Ho, K. M., *Photonic band gaps and defects in two dimensions: Studies of the transmission coefficient*, Phys. Rev. B **48** (19), pp. 14121-14126, 1993

Song, G.H., *Transparent boundary conditions for beam-propagation alalysis from the Green's function method*, J. Opt. Soc. Am. A **10** (5), pp. 896-904, 1993

Stoffer, R., Hoekstra, H.J.W.M., Ridder, R.M. de, Groesen, E. van, Beckum, F.P.H. van, *Numerical studies of 2D photonic crystals: Waveguides, coupling between wavegudies and filters*, Opt. and Quant. Electr. **32**, pp. 947-961, 2000

Taflove, A., *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, Artech House inc, Norwood, MA, ISBN 0-89006-792-9, 1995

Tamir, T., *Integrated Optics*, Springer Verlag, Berlin, 1979

Thylen, L., Opt. And Quantum Electr. **15**, pp. 433, 1983

Vasallo, C., Collino, F., *Highly Efficient Absorbing Boundary Conditions for the Beam Propagation Method*, Journ. of Lightw. Techn. **14** (6), pp. 1570-1577, 1996

Vasallo, C., Keur, J. M. van der, *Comparison of a few transparent boundary conditions for finite-difference optical mode-solvers*, Journ. of Lightw. Techn. **15**, pp.397-402, 1997

Yamauchi, J., Shibayama, J., Sekiguchi, M., Nakano, H., *Finite-Difference Beam Propagation Method Based on the Generalized Douglas Scheme for a Nonuniform Grid*, IEEE Phot. Techn. Lett. **9** (1), pp. 67-69, 1997

Yee, K.S., *Numerical solution of boundary value problems involving Maxwell's equations in isotropic media*, IEEE Trans. Antennas Prop. AP-14, pp 302, 1966

Yevick, D., Hermansson, B., Electronic Letters **25**, pp. 461, 1984

Yevick, D., Opt. And Quant. Electr. **26**, pp 185, 1994

Yevick, D., Yu, J., Schmidt, F., *Analytic Studies of Absorbing and Impedance-matched Boundary Layers*, IEEE Phot. Techn. Lett. **9** (1), pp. 73-75, 1997

Zijlstra, T., Drift, E.W.J.M. van der, Dood, M.J.A. de, Snoeks, E., and Polman, A., *Fabrication of two-dimensional photonic crystal waveguides at 1.5 um in silicon by deep anisotropic etching*, J. Vac. Sci. Technol. B17, pp. 2734, 1999

Samenvatting

Deze samenvatting is een bewerking van hoofdstuk 6 (‘Conclusions and Summary’).

In dit proefschrift worden verscheidene verbeteringen aan simulatiemethodes voor de geïntegreerde optica gepresenteerd, alsmede enkele nieuwe methodes. De gepresenteerde methodes zijn ofwel uni- ofwel omnidirectioneel. Voor de unidirectionele methodes ligt de nadruk op hogere-orde nauwkeurigheid; voor de omnidirectionele methodes zijn de randvoorwaarden extreem belangrijk, en moet gezorgd worden voor een goede definitie van de inkomende velden. De omnidirectionele methodes zijn toegepast op twee verschillende types structuren, namelijk een cilindrische microresonator en zogenaamde Photonic Bandgap Crystals.

Twee verschillende methodes zijn ontwikkeld om vierde-orde nauwkeurige Beam Propagation Methods (bundelpropagatiemethodes) te maken. Grensvlakken tussen materialen op een willekeurige positie in een rekencel kunnen worden meegenomen in de berekeningen, zodat er geen restricties zijn aan de positionering van het rekenrooster. De eerste methode past een Taylor-decompositie toe op de velden rondom de grensvlakken om correcties te verkrijgen op de discretisatie van de SVEA (langzaam-variërende) Helmholtz vergelijking. Dit levert erg simpele en numeriek efficiënte formules op voor TE polarisatie, waarin alleen de termen op de diagonaal van de matrices hoeven te worden gecorrigeerd. In de tweede methode wordt een eindige-differentie-schema gebruikt in de propagatierichting, en een variationele methode in de transversale richting. In deze variationele aanpak is de implementatie van TE en TM polarisatie vrijwel gelijk. Zowel tweede-orde als vierde-orde nauwkeurige schema’s zijn ontwikkeld. Door de variationele aanpak moet de norm die word gebruikt bij bv. berekeningen van energiestromen en overlappen worden veranderd; het standaard inproduct is niet nauwkeurig meer. De fout die gemaakt wordt in de effectieve index convergeert als $O(\mathbb{D}x^2)$ en $O(\mathbb{D}x^4)$ voor respectievelijk een drie- of vijfpuntsoperator. Door de verhoogde nauwkeurigheid van het vierde-orde schema is de hoeveelheid roosterpunten die nodig is voor een bepaalde nauwkeurigheid lager dan in het tweede-orde schema, zodat de totale rekentijd en geheugenconsumptie veel lager kunnen zijn bij een gegeven nauwkeurigheid. Aan de randen van het rekenvenster blijken Perfectly Matched Layers (PML’s, Perfect Passende Lagen) meer robuuste niet-reflecterende randen te geven dan Sommerfeld-achtige randvoorwaarden.

In de PML’s corresponderen de analytische oplossingen met licht dat in deze lagen geabsorbeerd wordt zonder reflecties te veroorzaken. Ook numeriek is het mogelijk om de reflectie laag te houden, maar enige zorg moet besteed worden aan zaken als de dikte en de sterkte van de PML. Voor de twee- en driedimensionale PML moeten de tijdsafhankelijke vergelijkingen gesplitst worden om richtingsafhankelijke demping te krijgen, en hierdoor nul reflectie aan het grensvlak voor stralende en geleide modi. In het tijdsharmonische geval (Helmholtz-vergelijking) correspondeert de PML met de toevoeging aan de vergelijkingen van extra complexe coëfficiënten. Voor een inkomend

modaal veld verandert de laterale vorm niet tijdens de propagatie door de PML; alleen zijn amplitude vermindert. Een belangrijke eigenschap van modale propagatie door een PML is het feit dat de demping afhangt van de propagatieconstante b van de mode.

Een omnidirectionele solver voor de 2-dimensionale TE en TM Helmholtz-vergelijking is ontwikkeld. De solver is gebaseerd op het variationele formalisme, waarin de oplossing van een differentiaalvergelijking wordt gevonden door te zoeken naar extrema van een functionaal. Het probleem wordt geherformuleerd: in plaats van de golfvergelijking te discretiseren, wordt de functionaal benaderd met behulp van een interpolatie van het veld tussen de roosterpunten, en de brekingsindexverdeling. De indexverdeling binnen een roostercel kan worden meegenomen. Hierdoor is het mogelijk om de berekeningen nauwkeurig te doen met een rechthoekig, structuuronafhankelijk rooster, zodat het niet nodig is complexe roosters te genereren. Het rekenvenster wordt omringd door PML's om (vrijwel) reflectievrije wanden te maken. Inkomende velden kunnen alleen op de buitenste rand, buiten de PML, worden gedefinieerd. Deze velden moeten door de PML gaan voor ze het binnenste deel van het venster bereiken, wat betekent dat ze gedempt worden met een factor die afhankelijk is van de component van de golfvector loodrecht op de rand. Als het inkomende veld uit meerdere modi bestaat, zal ieder van deze modi een eigen golfvector hebben, en dus zal de amplitude van iedere mode verschillend aangepast moeten worden om de juiste amplitude binnen het venster te verkrijgen. Enige zorg moet worden besteed aan de sterkte en de vorm van de PM om weinig numerieke reflectie te hebben en om nauwkeurige amplitudes van de inkomende velden te garanderen. Echter, er zal altijd een significante fout ontstaan in de amplitude van de inkomende velden, zodat er in het algemeen twee berekeningen dienen te worden gedaan: Eén met alleen de ingaande golfgeleider en geen verstrooiende structuren in het venster, waarmee de amplitude van de numerieke inkomende velden kan worden bepaald, en één met de verstrooiende structuren. Deze laatste berekeningen kunnen dan genormaliseerd worden met behulp van de eerste.

In de Finite Difference Time Domain (FDTD) methode worden dezelfde randvoorwaarden gebruikt, dus ook hier zijn de sterkte en breedte van de PML belangrijk, en dient de startveldgeneratie op de juiste manier te gebeuren. Aangezien de berekeningen met volledige tijdsafhankelijkheid gebeuren is het mogelijk om meerdere frequenties tegelijk te simuleren. Om de respons van een structuur voor modale inkomende velden te berekenen, moet dit modale veld voor elk van de frequenties anders zijn. Een manier om zogenaamde modale pulsen te generen is ontwikkeld en werkt goed.

Een voordeel van de Helmholtz-solver is dat hij nauwkeuriger kan zijn met een grover rooster, aangezien de grensvlakken door roostercellen nauwkeurig kunnen worden meegenomen. Op fijnere roosters wordt de rekentijd die de Helmholtz-solver nodig heeft echter veel groter dan die van de FDTD voor hetzelfde probleem, aangezien de Helmholtz-solver slechter schaalt dan N^2, terwijl de FDTD schaalt als $N^{3/2}$ of beter. Vanwege het bijkomende voordeel van de FDTD dat hij meerdere frequenties tegelijk kan doorrekenen, is dit een meer bruikbare methode voor de meeste toepassingen.

Zowel de Helmholtz-solver als de FDTD zijn bruikbare algoritmes om kleine cilindrische microresonatoren te simuleren. Bij dezelfde roostergrootte is de Helmholtz-solver nauwkeuriger door zijn betere benadering van schuine grensvlakken. De rekentijd voor een golflengtescan is echter ordes van grootte hoger dan bij de FDTD methode. Van een cilindrische microresonator zijn de Free Spectral Range (FSR) en de finesse uitgerekend, en die blijken goed overeen te komen met een andere theorie. Door de numerieke dispersie is de exacte positie van de resonanties een langzaam convergerende functie van de stapgroottes. In de praktijk zijn de FSR en de finesse echter de meest belangrijke parameters, en zij convergeren snel.

Op fotonische bandgap structuren zijn dezelfde twee methodes toegepast. De volgende resultaten zijn behaald:

- De bandgap is berekend.
- De golfgeleiderdispersie van een kristalgolfgeleider is bepaald.
- Een schatting van de efficiëntie die verwacht kan worden van de koppeling van licht van een normale diëlectrische golfgeleider naar een kristalgolfgeleider is gemaakt. Zoals verwacht zijn er duidelijke verschillen te zien in de vergelijking van deze resultaten met een simpel model waarin alleen de breedtes van de modale veldprofielen worden vergeleken.
- Een trilholte in de nabijheid van een kristalgolfgeleider creëert golflengteafhankelijkheid van de transmissie; voor bepaalde golflengtes wordt alle energie gereflecteerd.
- Om niet-reflecterende wanden te maken voor een kristalgolfgeleider moet de PML dikker gemaakt worden, aangezien een dunne PML significante reflecties veroorzaakt.

Dankwoord

Zo, na deze vier jaren OIO-schap ligt er dan eindelijk een boekje. Natuurlijk heb ik dit niet alleen voor mekaar gekregen; er zijn altijd veel mensen bereid geweest wetenschappelijke en morele steun te geven. Hierbij wil ik allen van harte bedanken; ik hoop iedereen te noemen die het verdient, maar mocht je er niet bij staan, dan komt dat gewoon door mijn verstrooide geest.

Als eerste en belangrijkste is daar natuurlijk mijn promotor, Brenny van Groesen. Ondanks de soms Babylonische spraakverwarringen tussen de natuurkundigen en de wiskundigen denk ik toch dat de samenwerking bijzonder nuttig en vooral aangenaam is verlopen. Ik heb de maandagochtendbesprekingen ondanks het vroege uur toch altijd leuk en nuttig gevonden. En natuurlijk bedankt voor het altijd weer vlak voor een deadline nakijken van een te publiceren paper. Als tweede wil ik Hugo Hoekstra bedanken voor de dagelijkse begeleiding, waarbij ik veel vrij werd gelaten, maar op de nodige momenten toch net even in de goeie richting werd geduwd; en natuurlijk voor de vele conferenties, waar het met een vodka of een biertje erbij altijd wel gezellig was (ehm, en natuurlijk wetenschappelijk interessant). Verder zijn alle leden van de maandagochtendbijeenkomsten bijzonder behulpzaam geweest. De belangrijkste is natuurlijk Frits van Beckum, maar ook de anderen, Peter, Theo, Manfred, Agus, Israel en Stefan worden hartelijk bedankt. In het eerste jaar van mijn werk heb ik bijzonder veel met Peter 'Old Spice' Bollerman samengewerkt, die me een hoop enthousiasme voor de wiskunde heeft bijgebracht, en met wie het ook goed in de kroeg zitten was.

Ik heb bij twee verschillende vakgroepen gezeten in deze tijd; zowel bij wiskunde als bij natuurkunde. Altijd leuk om de cultuurverschillen tussen deze twee richtingen, die toch niet eens zo ver uit mekaar liggen, te mogen observeren. In beide groepen zitten bijzonder leuke mensen, die ik nu dus zal proberen bij naam te noemen. Bij wiskunde was natuurlijk mijn kamergenoot het merendeel van de tijd Wilbert IJzerman. Dan hebben we nog de spelletjesfanaat Timco; we moeten nog eens wat spelen! En natuurlijk bedank ik ook alle anderen, Edi, Barbera, Stefan, Frits, Manfred, Judith, Jaap-Harm, en Helena, voor de leuke tijd. Bij natuurkunde heb ik in een boel kamers gezeten, en dus ook een hoop kamergenoten gehad. Chris 'Roelie / Savant' Roeloffzen, Libor Kotacka, Harsoyono, Szabolcs, Rene 'Opperzombie' Heideman, Sami Musa Abbas Mohammed (om het maar eens helemaal uit te schrijven), Freddy, Henk 'Rambling Mammot' Bulthuis, en Dion 'Bionic Klungel / Klungel de Bas / Quak Atheist' Klunder. En verder zijn er natuurlijk nog vele anderen binnen LDG die een groot of klein bedankje verdienen: Anton Hollink en Henk van Wolveren voor het altijd prima voor mekaar hebben van het computersysteem; de altijd geinteresserde (en met interessante problemen komende) stafleden René de Ridder, Paul Lambeck en Alfred Driessen; de meestersnookeraars Ton, Rubberen Robbie, Toon en Harm; de gezelligheid van Joris 'Rope' van Lith, Geert 'Onvrijwillige Borrelcommissaris' Altena, Marcel 'Chateau 1967' Hoekman, Gertje, en verder alle studenten die hier in de loop der jaren zijn

voorbijgekomen; dat zijn er te veel om te noemen. Een bijzondere vermelding moet ik wel geven aan Gerard Klaasse, ‘mijn’ student, die het heeft aangedurfd om onder mijn leiding een vrij praktisch georiënteerde opdracht aan te pakken, en daar nog bijzonder leuke resultaten in heeft geboekt ook.

Begin 2000 heb ik drie maanden aan de Australian National University in Canberra gewerkt. Ondanks dat ze dit niet kunnen lezen, wil ik de volgende mensen bedanken voor een onvergetelijke tijd: Yuri, Martin, Matthias, Sam, Carsten, Andreas, Chiara, Ruth, Sjnezana, Elena, en Keith.

Uiteraard is hier ook een bedankje op zijn plaats voor de geldschieters, STW/NWO.

En natuurlijk, als laatste maar niet als minste, een groot Dankjewel voor mijn familie en buiten-universitaire vrienden. Mijn ouders en mijn zus Fiona voor de niet-aflatende morele (en in sommige gevallen noodzakelijke financiële) steun. Opa en oma en opa: Ik hoop dat jullie na dit gebeuren iets gemakkelijker kunnen uitleggen wat die rare kleinzoon van jullie nou toch eigenlijk doet. Remco, Pety, Harry, Egon en Anne-Marie: ik hoop dat we mekaar na deze voor mij drukke periode weer eens wat vaker zullen zien!

Dat moet het wel zo’n beetje zijn. De groeten,

Remco.

Bibliography

R. Stoffer & H.J.W.M. Hoekstra, *Efficient interface conditions based on a 5-point finite difference operator*, Optical and quantum electronics, 30 (nr: 5-6), ISSN 0306-8919, (1998), pp. 375-383

H.P. Nolting, M. Gravert, H.J.W.M. Hoekstra, R. Stoffer, G.J.M. Krijnen, C. Sibilia, M. de Minicis, R. Pregla, & O. Conradi, *Beam-propagation method benchmark test: symmetrical coupler and gain-loss waveguide*, Proceedings of the Integrated Photonics Research Conference (IPR 98), 1998, pp. 3

R. Stoffer, *Computational methods in integrated optics*, Proceedings of the IEEE/LEOS 1998 Benelux Chapter, Gent / Belgium, 26/11/98, ISBN 90-76546-01-0, pp. 225-228

R.M. de Ridder, A. Driessen, H.J.W.M. Hoekstra, D.J.W. Klunder, R. Stoffer, M.L.M. Balistreri, & N.F. van Hulst, *Modelling, fabrication and characterisation of microcavities at MESA Lightwave Devices Group*, Proceedings of the COST 268 Workshop on Wavelength Scale Photonic Components for Telecommunications, Stockholm / Sweden, 15/3/99, 2 pp.

R. Stoffer, H.J.W.M. Hoekstra, E. van Groesen, F.P.H. van Beckum, *Pandirectional simulation scheme with Perfectly Matched Layers allowing accurate definition of incoming field*, Proceedings of the 9th European Conference on Integrated Optics and Technical Exhibition, Torino / Italy, 13/4/99, pp. 47-50

R. Stoffer, C. Stolk, S.W. Rienstra, J.K.M. Jansen, *Windtunnel model position and orientation*, Proc. Of the 36th European Study Group with Industry, ISSN 0167-9708, Eindhoven, 1999, pp. 33-39

R. Stoffer, P.A.A.J. Bollerman, H.J.W.M. Hoekstra, E.W.C. van Groesen, F.P.H. van Beckum, *New true fourth-order accurate scalar beam propagation methods for both TE and TM polarizations*, Optical and quantum electronics, 31, ISSN 0306-8919, (1999), pp. 705-720

R. Stoffer, R.M. de Ridder, H.J.W.M. Hoekstra, E.W.C. van Groesen, F.P.H. van Beckum, *Calculations on 2-dimensional waveguides in photonic crystals*, Proceedings of LEOS '99 annual meeting, 1999, pp. 421-422

R. Stoffer, H.J.W.M. Hoekstra, R.M. de Ridder, E. van Groesen, F.P.H. van Beckum, *Numerical studies of 2D photonic crystals: Waveguides, coupling between waveguides and filters*, Optical and Quantum Electronics, 32, ISSN 0306-8919, (2000), pp. 947-961

R. Stoffer, Yu. S. Kivshar, *Numerical study of optical bistability in a nonlinear photonic crystal notch filter*, Poster presentation at Annual Symposium of the IEEE/LEOS Benelux chapter, Delft, November, 2000