

Dielectric Waveguides

Peter Hertel

**Physics Department
Osnabrück University
Germany**

Lectures delivered at
TEDA Applied Physics School
Nankai University, Tianjin, PRC

Dielectric waveguides are the key components of modern integrated optics. A region of increased permittivity hinders light from spreading in space.

In section 1 we explain why waves cannot be kept confined unless the medium is inhomogeneous. We then summarize Maxwell's equations for the electromagnetic field and specialize to modes of definite angular frequency.

Planar waveguides are the subject of section 2. Just as light propagating in free space has two states of polarization, there are also two kinds of modes, transversal electric or transversal magnetic. We calculate the modes of a graded index waveguide by the finite difference method and discuss a semi-analytic approach for slab waveguides.

We next turn, in section 3, to the more realistic case of strip waveguides which are the integrated optics counterpart to wires and bonds in electronics. Again, modes are either quasi transversal electric or transversal magnetic. A realistic rib waveguide is investigated by the finite difference method. A few alternative methods for calculating the guided modes of strip waveguides are mentioned as well.

Section 4 is about wave propagation. We derive the Fresnel equation and implement a finite difference propagation method. A necessarily finite computational window must be equipped with transparent boundary conditions in order to prevent spurious reflections. We comment on the method of lines and on the operator splitting beam propagation method.

Finally, in section 5, some effects of optical anisotropy are discussed, in particular non-reciprocal propagation which is required for an optical isolator.

September 2009

Contents

1	Basics	3
1.1	Plane Waves and wave packets	3
1.2	Maxwell's equations	4
1.3	Monochromatic waves and modes	5
2	Planar Waveguides	8
2.1	TE modes	8
2.2	TM modes	9
2.3	Graded index waveguides	10
2.4	Slab waveguides	12
3	Strip Waveguides	15
3.1	Quasi TE and TM modes	15
3.2	Finite difference method	16
3.3	Various other methods	18
3.3.1	Galerkin methods	19
3.3.2	Method of lines	20
3.3.3	Collocation methods	21
4	Propagation	22
4.1	Fresnel equation	22
4.2	Finite differences	23
4.3	Transparent boundary conditions	24
4.4	Propagation in a slab waveguide	27
4.5	Other propagation methods	29
4.5.1	Method of lines	29
4.5.2	Operator splitting	30
5	Optical Anisotropy	33
5.1	Permittivity tensor	33
5.2	Anisotropic waveguides	35
5.3	Non-reciprocal effects	37
5.3.1	The Faraday effect	37
5.3.2	Waveguide isolators	38
A	Program Listings	40

1 Basics

We discuss waves and wave packets and show why they have to spread out if the transmitting medium is homogeneous. Only if the permittivity varies with location, waves or wave packets may be confined. We recapitulate Maxwell's equations for the electromagnetic field and discuss harmonic in time solutions.

1.1 Plane Waves and wave packets

A plane wave is described by

$$f(t, \mathbf{x}) \propto e^{-i\omega t} e^{i\mathbf{k} \cdot \mathbf{x}}. \quad (1.1)$$

The wave (1.1) is called plane because, at a particular time t , the surfaces of constant phase are planes being orthogonal to the wave vector \mathbf{k} . f stands for the field strength which might be the air pressure deviation δp for sound waves, a component u_i of the displacement vector for elastic waves, any component of the electromagnetic field (\mathbf{E} , \mathbf{B}), or the amplitude of a particle wave.

In all these cases there are linear field equations which allow to derive a relation between the angular frequency ω and the wave vector, $\omega = \omega(\mathbf{k})$. It will be explained soon why we speak of a dispersion relation.

For sound in air, $\omega = v|\mathbf{k}|$ is a rather good approximation up to frequencies of tens of kilohertz. The same applies for longitudinally or transversally polarized elastic waves. Only for rather high frequency (ultrasound) deviations from a linear relation between frequency and wave number¹ become noticeable.

Light in the infrared, visible, and ultraviolet region, if propagating in matter, exhibits marked deviations from a linear relationship between ω and k . This is to be expected since the photon energy, one or a few electron volts, matches typical energies of atomic physics.

Free particles of mass m are characterized² by

$$\omega = \frac{\hbar}{2m} k^2 \quad (1.2)$$

which is very non-linear.

Now, plane waves are an idealization. They are there everywhere and always. In fact, waves are excited and more or less localized. Realistic waves consist of packets as described by

$$f(t, \mathbf{x}) = \int \frac{d^3k}{(2\pi)^3} \phi(\mathbf{k}) e^{-i\omega t} e^{i\mathbf{k} \cdot \mathbf{x}}. \quad (1.3)$$

Note that

$$\int d^3x |f(t, \mathbf{x})|^2 = \int \frac{d^3k}{(2\pi)^3} |\phi(\mathbf{k})|^2 \quad (1.4)$$

¹magnitude $k = |\mathbf{k}|$ of the wave vector
² \hbar is Planck's constant

does not depend on time. We interpret it as the wave's energy and normalize to unity.

The wave packet is located at

$$\langle \mathbf{X} \rangle_t = \int d^3x \mathbf{x} |f(t, \mathbf{x})|^2 \quad (1.5)$$

which is equal to

$$\langle \mathbf{X} \rangle_t = \int \frac{d^3k}{(2\pi)^3} \phi^*(\mathbf{k}) e^{i\omega t} i \nabla_{\mathbf{k}} \phi(\mathbf{k}) e^{-i\omega t}, \quad (1.6)$$

or

$$\langle \mathbf{X} \rangle_t = \langle \mathbf{X} \rangle_0 + t \int \frac{d^3k}{(2\pi)^3} |\phi(\mathbf{k})|^2 \nabla_{\mathbf{k}} \omega. \quad (1.7)$$

Hence, the wave packet moves with constant velocity

$$\langle \langle \nabla_{\mathbf{k}} \omega \rangle \rangle = \int \frac{d^3k}{(2\pi)^3} |\phi(\mathbf{k})|^2 \nabla_{\mathbf{k}} \omega. \quad (1.8)$$

In the same way we may define $\langle \mathbf{X}^2 \rangle_t$. This expectation value turns out to be quadratic in t , the leading term being

$$\langle \mathbf{X}^2 \rangle_t = t^2 \langle \langle (\nabla_{\mathbf{k}} \omega)^2 \rangle \rangle + \dots \quad (1.9)$$

Thus, for large times, the root-mean-square extension of the wave increases with time like

$$\delta X(t) = \sqrt{\langle \mathbf{X}^2 \rangle_t - \langle \mathbf{X} \rangle_t^2} = t \sqrt{\langle \langle (\nabla_{\mathbf{k}} \omega)^2 \rangle \rangle - \langle \langle \nabla_{\mathbf{k}} \omega \rangle \rangle^2} + \dots \quad (1.10)$$

Wave packets spread out more and more. There are two reasons. One is that the packet is made up of waves travelling in different directions. The other reason is that even for parallel wave vectors angular frequency and wave number are not proportional. The spreading of wave packets is unavoidable—in a homogeneous medium.

1.2 Maxwell's equations

Maxwell's equations in matter read

$$\mathbf{div} \mathbf{D} = \rho, \quad \mathbf{div} \mathbf{B} = 0 \quad (1.11)$$

and

$$\mathbf{curl} \mathbf{H} = \mathbf{j} + \dot{\mathbf{D}}, \quad \mathbf{curl} \mathbf{E} = -\dot{\mathbf{B}}. \quad (1.12)$$

ρ is the charge density, \mathbf{j} the current density. \mathbf{D} , \mathbf{B} , \mathbf{H} and \mathbf{E} stand for the dielectric displacement, induction, magnetic and electric field strength, respectively. They are fields, i. e. depend on time t and location \mathbf{x} . The divergence of a vector field \mathbf{F} is $\mathbf{div} \mathbf{F} = \nabla \cdot \mathbf{F}$, the curl is defined by $\mathbf{curl} \mathbf{F} = \nabla \times \mathbf{F}$, and $\dot{\mathbf{F}} = \partial_t \mathbf{F}$ denotes the partial derivative with respect to time. $\nabla = (\partial_x, \partial_y, \partial_z)$ is the nabla operator.

The conservation of charge is expressed by the continuity equation

$$\dot{\rho} + \mathbf{div} \mathbf{j} = 0, \quad (1.13)$$

a consequence of Maxwell's equations.

In a linear medium, which is characterized by $\mathbf{D} \propto \mathbf{E}$ and $\mathbf{B} \propto \mathbf{H}$, there is an energy density

$$\eta = \frac{\mathbf{E} \cdot \mathbf{D} + \mathbf{H} \cdot \mathbf{B}}{2} \quad (1.14)$$

and an energy current density

$$\mathbf{S} = \mathbf{E} \times \mathbf{H} \quad (1.15)$$

to be associated with the electromagnetic field. The following balance equation holds true:

$$\dot{\eta} + \mathbf{div} \mathbf{S} = -\mathbf{j} \cdot \mathbf{E}. \quad (1.16)$$

\mathbf{S} is Poynting's vector, and $-\mathbf{j} \cdot \mathbf{E}$ is Joule's heat. Equation (1.15) is known as Poynting's theorem. Like the continuity equation (1.13), it is a consequence of Maxwell's equations.

At interfaces between two homogeneous media the following field components are continuous:

- the normal component B_{\perp} of the induction
- the normal component D_{\perp} of the dielectric displacement, if there is no surface charge
- both tangential components \mathbf{E}_{\parallel} of the electric field strength, and
- both tangential components \mathbf{H}_{\parallel} of the magnetic field strength, if there is no surface current.

1.3 Monochromatic waves and modes

Assume $f = f(t, \mathbf{x})$ any of the electromagnetic field components. We may decompose it into its frequency components:

$$f(t, \mathbf{x}) = \int \frac{d\omega}{2\pi} e^{-i\omega t} \hat{f}(\omega, \mathbf{x}). \quad (1.17)$$

Now, $f = f(t, \mathbf{x})$ is a real field, and this implies

$$\hat{f}^*(-\omega, \mathbf{x}) = \hat{f}(\omega, \mathbf{x}) . \quad (1.18)$$

Therefore, we may write

$$f(t, \mathbf{x}) = \int_0^\infty \frac{d\omega}{2\pi} e^{-i\omega t} \hat{f}(\omega, \mathbf{x}) + \text{cc} , \quad (1.19)$$

where cc denotes the complex conjugate of the term to the left. Hence, only positive frequencies matter. We will always bear in mind that the complex conjugate has to be added, but usually omit cc.

In the following we pick out one component with a well-defined positive angular frequency ω . In $\hat{f}(\omega, \mathbf{x})$ we drop the $\hat{}$ indicator (for Fourier transform), and do not mention ω in the list of arguments.

We investigate a dielectric medium without charges and currents. This situation is characterized by

$$\rho = 0 , \quad \mathbf{j} = 0 , \quad \mathbf{D} = \epsilon\epsilon_0\mathbf{E} \quad \text{and} \quad \mathbf{B} = \mu_0\mathbf{H} . \quad (1.20)$$

Maxwell's equations for the electric and magnetic field strengths now become

$$\mathbf{div} \epsilon\mathbf{E} = 0 , \quad \mathbf{div} \mathbf{H} = 0 \quad (1.21)$$

and

$$\mathbf{curl} \mathbf{H} = -i\omega\epsilon_0\epsilon\mathbf{E} , \quad \mathbf{curl} \mathbf{E} = i\omega\mu_0\mathbf{H} . \quad (1.22)$$

Note that the permittivity $\epsilon = \epsilon(\mathbf{x})$ may depend on location, but not on time. The two first order equations (1.22) are inserted into each other such that a second order equation results:

$$\mathbf{curl} \mathbf{curl} \mathbf{E} = k_0^2\epsilon \mathbf{E} , \quad (1.23)$$

where $k_0 = \omega/c$ and $c = 1/\sqrt{\epsilon_0\mu_0}$.

An alternative version is

$$\mathbf{curl} \epsilon^{-1} \mathbf{curl} \mathbf{H} = k_0^2\mathbf{H} . \quad (1.24)$$

Let us define the following scalar product for vector fields:

$$(\mathbf{b}, \mathbf{a}) = \int d^3x \mathbf{b}^*(\mathbf{x}) \cdot \mathbf{a}(\mathbf{x}) . \quad (1.25)$$

It is a simple exercise to show that

$$(\mathbf{b}, \mathbf{curl} \mathbf{a}) = (\mathbf{curl} \mathbf{b}, \mathbf{a}) \quad (1.26)$$

holds true for square integrable³ and differentiable vector fields \mathbf{a} and \mathbf{b} . The **curl** operator is Hermitian. The (**curl curl**) operator in (1.23), the square of a Hermitian operator, is therefore non-negative. This is in agreement with $k_0^2\epsilon$ being non-negative.

Any solution of (1.23) for $k_0^2 \neq 0$ obeys (1.11). This is evident for the displacement field if the divergence of (1.23) is worked out. However, because of $\mathbf{B} \propto \mathbf{E}$, the divergence of the induction field vanishes as well.

The curl is a differentiation operator to be applied to vector fields, and we think of $\epsilon = \epsilon(\mathbf{x})$ as a multiplication operator. (1.23) defines a generalized eigenvalue problem, the eigenvalue being k_0^2 . (1.24) is a normal eigenvalue problem.

Given a permittivity profile, the allowed light frequency values $\omega = ck_0$ may be calculated. (1.23) or (1.24) describe a resonator. If there is a three-dimensional region Ω of increased permittivity, light of certain frequencies may be stored in it. We will not discuss this here. Instead, we shall study structures where the permittivity varies along one or two dimensions only. $\epsilon = \epsilon(x)$ describes a planar waveguide, $\epsilon = \epsilon(x, y)$ a strip waveguide.

For equations which are linear in the fields we may omit the cc reminder. For quadratic expressions we have to be more careful. For example, the Poynting vector is given by

$$\mathbf{S} = (\mathbf{E} + \mathbf{E}^*) \times (\mathbf{H} + \mathbf{H}^*) \quad (1.27)$$

which are altogether four contributions. Two of them oscillate with angular frequency 2ω and -2ω , they should be dropped. What remains is the zero frequency, or time averaged contribution

$$\mathbf{S} = 2 \operatorname{Re} \mathbf{E} \times \mathbf{H}^* . \quad (1.28)$$

³ \mathbf{f} is square integrable if $(\mathbf{f}, \mathbf{f}) < \infty$

2 Planar Waveguides

A planar waveguide is characterized by a permittivity profile $\epsilon = \epsilon(x)$ which does not depend on y or z . The wave vector lies in the (y, z) plane, and we choose the z direction without loss of generality. All components of the electromagnetic field are shaped according to

$$F(t, x, y, z) = F(x) e^{-i\omega t} e^{i\beta z}. \quad (2.1)$$

Just as a free photon has two states of polarization, there are two differently polarized modes, TE and TM. It is the Electric or Magnetic field strength, respectively, which is Transversal, i. e. orthogonal to the waveguide normal as well as to the propagation direction.

We study two different kinds of planar waveguides. A graded index waveguide is characterized by a smoothly varying permittivity profile while a slab waveguide consists of one or more homogeneous films of different optical properties.

2.1 TE modes

The electromagnetic field of a TE mode is

$$\mathbf{E} = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \frac{1}{\omega\mu_0} \begin{pmatrix} -\beta E \\ 0 \\ -i E' \end{pmatrix}. \quad (2.2)$$

The electric field strength $E = E(x)$ has to obey⁴

$$\frac{1}{k_0^2} E'' + \epsilon E = \epsilon_{\text{eff}} E. \quad (2.3)$$

This is an eigenvalue problem, the eigenvalue being the effective permittivity $\epsilon_{\text{eff}} = (\beta/k_0)^2$. For a given light source frequency ω , the mode equation (2.3) allows to calculate the possible propagation constants β .

According to (1.28) and (2.2) the energy current density is

$$S = \frac{2\beta}{\omega\mu_0} |E(x)|^2. \quad (2.4)$$

By integrating over x we obtain the power flux per lateral unit length:

$$\frac{dP}{dy} = \frac{2\beta}{\omega\mu_0} \int dx |E(x)|^2. \quad (2.5)$$

It is therefore natural to define the following scalar product:

$$(g, f) = \int dx g^*(x) f(x). \quad (2.6)$$

⁴ $n_{\text{eff}} = \beta/k_0$ is called an effective index, its square $\epsilon_{\text{eff}} = n_{\text{eff}}^2 = \beta^2/k_0^2$ is an effective permittivity.

Note that the mode operator⁵

$$L_{\text{TE}} = \frac{1}{k_0^2} \frac{d^2}{dx^2} + \epsilon \quad (2.7)$$

is self-adjoint with respect to the scalar product (2.6). Hence, the ϵ_{eff} are real. Since the second derivative operator is negative⁶, the eigenvalues ϵ_{eff} are smaller than the largest permittivity. A guided mode is characterized by $(E, E) < \infty$, by a finite total power flux per unit lateral length. Therefore, ϵ_{eff} must be larger than the permittivities at infinity. Otherwise the solutions would be of sine type at infinity and could not be normalized.

All continuity requirements are fulfilled if $x \rightarrow E(x)$ and $x \rightarrow E'(x)$ are continuous.

2.2 TM modes

The electromagnetic field of a TM mode is

$$\mathbf{E} = \frac{1}{\omega \epsilon_0 \epsilon} \begin{pmatrix} \beta H \\ 0 \\ i H' \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} 0 \\ H \\ 0 \end{pmatrix}. \quad (2.8)$$

The magnetic field strength has to obey the following mode equation⁷

$$\frac{1}{k_0^2} \epsilon \frac{d}{dx} \epsilon^{-1} \frac{d}{dx} H + \epsilon H = \epsilon_{\text{eff}} H. \quad (2.9)$$

This is again an eigenvalue problem, the eigenvalue being ϵ_{eff} .

(2.8) implies the following expression for the power flux per lateral unit length:

$$P = \frac{\beta}{\omega \epsilon_0} \int dx \frac{1}{\epsilon(x)} |H(x)|^2. \quad (2.10)$$

It is therefore natural to define the scalar product

$$(g, f) = \int dx \frac{1}{\epsilon(x)} g^*(x) f(x). \quad (2.11)$$

It is not difficult to show that the TM mode operator

$$L_{\text{TM}} = \frac{1}{k_0^2} \epsilon \frac{d}{dx} \epsilon^{-1} \frac{d}{dx} + \epsilon \quad (2.12)$$

is self-adjoint with respect to the scalar product (2.11), hence its eigenvalues are real and the eigenvectors are orthogonal in the sense of (2.11).

⁵ L_{TE} is dimensionless which is required for numerical solutions.

⁶ A is negative if $(f, Af) \leq 0$ for all f , here: $(f, f'') = -(f', f') \leq 0$.

⁷This is just one of many forms

Again, as for TE modes, the differential operator part of (2.12) is negative, since $(f, f'') = -(f', f')$ holds true. Therefore, the allowed effective permittivities are smaller than the maximum permittivity and larger than the permittivity at infinity.

All continuity requirements are fulfilled if $x \rightarrow H(x)$ and $x \rightarrow H'(x)/\epsilon(x)$ are continuous.

2.3 Graded index waveguides

Think of a substrate like glass or lithium niobate. Its surface may be treated by various processes in order to modify the permittivity at the surface, such as in-diffusion or exchange of ions. Lithium niobate may be covered by a thin titanium layer which is then allowed to diffuse into the substrate at high temperatures. Another procedure is to apply benzoic acid which replaces a certain amount of lithium ions by protons. With glass, one can offer silver ions which are drawn into the substrate by an electric field.

In any case, a permittivity profile is produced which exceeds the substrate value by $\Delta\epsilon(x)$, where x is the depth below the surface at $x = 0$. The region $x < 0$ is the cover, usually vacuum or air, or a protective substance. Its permittivity is denoted by ϵ_c . The substrate permittivity ϵ_s is larger than ϵ_c . Since the concentration of in-diffused ions follows a Gaussian and since, for not too high concentrations, the permittivity change is proportional to the concentration of in-diffused ions, we assume

$$\epsilon(x) = \begin{cases} \epsilon_c & \text{for } x < 0 \\ \epsilon_s + \Delta\epsilon e^{-(x/W)^2} & \text{for } x > 0 \end{cases} . \quad (2.13)$$

w denotes the width of the permittivity increase, and $\Delta\epsilon$ the maximum permittivity enhancement. (2.13) is a rather good approximation for titanium in-diffused planar waveguides.

The standard procedure to solve such an eigenvalue problem is to approximate the infinite x axis \mathbb{R} by a finite number $x_j = x_{\min}, x_{\min} + h, \dots, x_{\max}$ of representative points. The field values $F_j = F(x_j)$ form a vector. A linear operator is represented by a square matrix. Here we describe the method of finite differences: infinitesimals dx are approximated by finite differences, h in our cases. The second derivative in particular is represented by

$$f''(x_j) = f_j'' = \frac{f_{j+1} - 2f_j + f_{j-1}}{h^2} , \quad (2.14)$$

which can be translated into a matrix to be applied to a vector f . This matrix has a diagonal $-2/h^2$ and adjacent diagonals $1/h^2$. A multiplication operator, such as $f \rightarrow \epsilon f$ is represented by a diagonal matrix with elements $\epsilon_j = \epsilon(x_j)$.

Here is a realization.

Our MATLAB program begins by defining constants. All lengths are measured in microns.

```
1 % this file is gi_wg.m
```

```

2  LAMBDA=0.6328;
3  k0=2*pi/LAMBDA;
4  EC=1.000;
5  ES=4.800;
6  ED=0.045;
7  W=4.00;

```

The wavelength is that of a cheap helium-neon laser, the cover is air, the substrate permittivity is that of lithium niobate, and the permittivity profile parameters $\Delta\epsilon$ (ED) and W are realistic.

We next represent the real axis by a finite set of representative values. The run from -1 to $4W$, in steps of h .

```

8  xmin=-1.0;
9  xmax=4*W;
10 h=0.1;
11 x=(xmin:h:xmax)';
12 dim=size(x,1);

```

The next line defines the permittivity profile:

```

13 prm=(x<0).*EC+(x>=0).*(ES+ED*exp(-(x/W).^2));

```

The following piece of code sets up the mode operator L :

```

14 next=ones(dim-1,1)/h^2/k0^2;
15 main=-2.0*ones(dim,1)/h^2/k0^2+prm;
16 L=diag(next,-1)+diag(main,0)+diag(next,1);

```

Its eigenvectors and eigenvalues are calculated by

```

17 [evec, eval]=eig(L);

```

Only eigenvectors with eigenvalues $(\beta/k_0)^2 = \epsilon_{\text{eff}} > \epsilon_s$ make sense. We isolate and plot them:

```

18 eff_eps=diag(eval);
19 guided=evec(:,eff_eps>ES);
20 plot(x,guided);

```

Figure 1 shows the output. There are three guided TE modes which are indexed by TE0, TE1, and TE2. The basic mode, the one with the largest propagation constant, has no node. The next one has one node, and so forth.

On my laptop computer the above program requires 0.3 s to build up the 171×171 matrix, to diagonalize it, and to produce the graphical representation. It is simply not worthwhile to improve it.

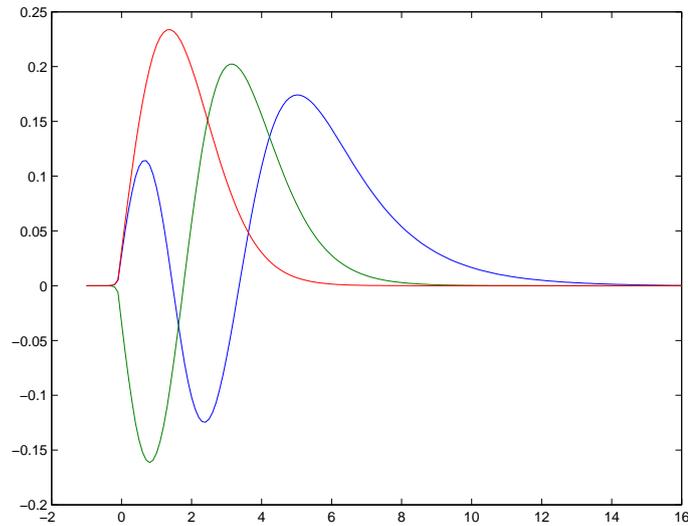


Figure 1: Electric field strength of guided modes vs. depth below surface (microns) of a graded index waveguides. See the MATLAB program for waveguide parameters.

2.4 Slab waveguides

A slab waveguide is made up of a substrate carrying one or more homogeneous films of enhanced permittivity. On top is a cover layer. Here we study a very simple device. There is just one film with permittivity $\epsilon_f > \epsilon_s$ of thickness w . The permittivity profile is

$$\epsilon(x) = \begin{cases} \epsilon_s & \text{for } x < 0 \\ \epsilon_f & \text{for } 0 < x < w \\ \epsilon_c & \text{for } w < x \end{cases} \quad (2.15)$$

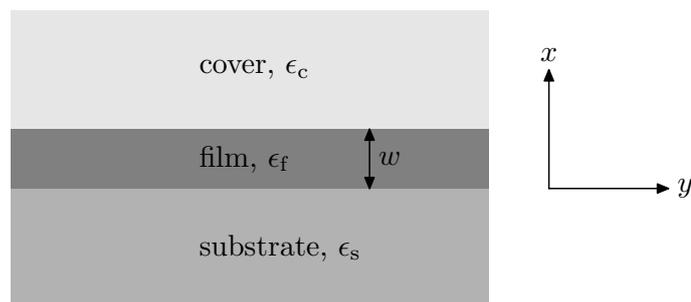


Figure 2: Layout of a slab waveguide with one film of enhanced permittivity.

Formally, the TE and the TM mode operators, (2.7) and (2.12) respectively, coincide for such a stepwise constant profile. However since the continuity re-

quirements differ, the propagation constants and the modal fields as well are different for TE and TM modes.

Let us define

$$\kappa_c = k_0 \sqrt{\epsilon_{\text{eff}} - \epsilon_c} \quad , \quad \kappa_s = k_0 \sqrt{\epsilon_{\text{eff}} - \epsilon_s} \quad \text{and} \quad k_f = k_0 \sqrt{\epsilon_f - \epsilon_{\text{eff}}} \quad . \quad (2.16)$$

These expressions are positive as long as we insist on $\epsilon_c, \epsilon_s < \epsilon_{\text{eff}} < \epsilon_f$.

In the substrate region, the mode equation has two fundamental solutions, namely $E \propto \pm \exp(\kappa_s x)$. We chose the positive sign in order to guarantee decay at $x \rightarrow -\infty$.

For the TE mode, the field and its derivative have to be continuous at interfaces between different materials. The solutions in the film region is $c \cos(k_f x) + s \sin(k_f x)$. We therefore have to determine the amplitudes c and s such that

$$1 = c \quad \text{and} \quad \kappa_s = s k_f \quad (2.17)$$

hold true at the interface ($x = 0$) between substrate and film, therefore

$$E \propto \cos k_f x + \frac{\kappa_s}{k_f} \sin k_f x \quad \text{for} \quad 0 < x < w \quad . \quad (2.18)$$

In the cover $x > w$ the field is a linear combination of two exponential functions, namely $E \propto a \exp(-\kappa_c x) + b \exp(-\kappa_c x)$. The continuity requirements for TE modes at $x = w$ are

$$\cos k_f w + \frac{\kappa_s}{k_f} \sin k_f w = a e^{-\kappa_c w} + b e^{\kappa_c w} \quad (2.19)$$

and

$$k_f (-\sin k_f w + \frac{\kappa_s}{k_f} \cos k_f w) = \kappa_c (-a e^{-\kappa_c w} + b e^{\kappa_c w}) \quad . \quad (2.20)$$

The condition for a guided mode reads $b = 0$. There must not be an exploding contribution. (2.19) and (2.20) are compatible only if

$$\cot k_f w = \frac{k_f^2 - \kappa_s \kappa_c}{k_f (\kappa_s + \kappa_c)} \quad (2.21)$$

holds true.

An analogous calculation for TM modes results in

$$\cot k_f w = \frac{\bar{k}_f^2 - \bar{\kappa}_s \bar{\kappa}_c}{\bar{k}_f (\bar{\kappa}_s + \bar{\kappa}_c)} \quad (2.22)$$

where $\bar{\kappa}_c = \kappa_c / \epsilon_c$, $\bar{\kappa}_s = \kappa_s / \epsilon_s$ and $\bar{k}_f = k_f / \epsilon_f$.

Note that the right hand side of (2.22) is smaller than its TE counterpart. Therefore, the propagation constants of TM modes lie below the corresponding TE values.

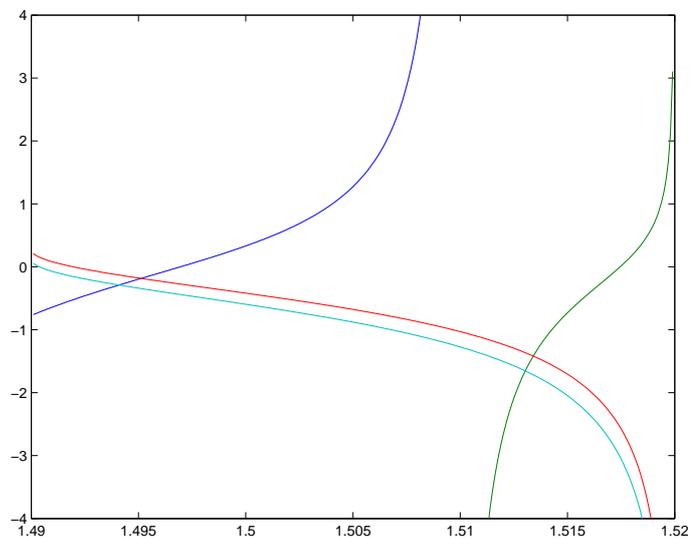


Figure 3: Graphical representation of (2.21) and (2.22). The cotangent as well as the right hand sides are plotted vs. effective index n_{eff} . The film (refractive index 1.52, thickness 1.8 microns) is deposited on a glass substrate (refractive index 1.49) and covered by air. The simulation is for a helium-neon laser. There are two guides TE and two guided TM modes.

Formulae (2.21) or (2.22) allow for an inverse procedure. Assume that at least two TE modes are guided. One can then, for a guessed film refractive index n_f , solve (2.21) for the film thickness w . For each mode i , a film thickness w_i results, but they will coincide only if the guessed film refractive index is correct. If there are more than two modes, the root mean square deviation of calculated film thicknesses must be minimized. Applying the same procedure to TM modes should result in the same film parameters.

3 Strip Waveguides

A linear, or strip waveguide is characterized by a permittivity profile depending on the cross section coordinates, $\epsilon = \epsilon(x, y)$. Strip waveguides confine light in a cross section, they are the counterparts of wires for electric currents. Unlike wires, strip waveguides must be rather straight. If the bending radius, as compared with the waveguide cross section, becomes too small, then power is radiated off or reflected. The latter phenomenon is well known for microwave guides.

Strip waveguides may be produced by depositing on a substrate a thin film of higher permittivity and removing it apart from a small rib. This is the most common technique. On lithium niobate a very thin film of titanium is deposited which, by common structuring procedures, is removed up to a small rib which is then in-diffused. Thereby, a smooth permittivity profile $\epsilon = \epsilon(x, y)$ is generated. We look for solutions of Maxwell's equations of the following form:

$$F(t, x, y, z) = F(x, y) e^{-i\omega t} e^{i\beta z} . \quad (3.1)$$

F may be any component of the electromagnetic field. The mode equation now is a system of coupled partial differential equations.

3.1 Quasi TE and TM modes

Recall the general mode equation

$$\mathbf{curl} \mathbf{curl} \mathbf{E} = k_0^2 \epsilon \mathbf{E} . \quad (3.2)$$

The curl operator is now

$$\mathbf{curl} = \begin{pmatrix} 0 & -i\beta & \partial_y \\ i\beta & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \end{pmatrix} . \quad (3.3)$$

Applying it twice results in

$$\begin{pmatrix} \beta^2 - \partial_y^2 & \partial_x \partial_y & i\beta \partial_x \\ \partial_x \partial_y & \beta^2 - \partial_x \partial_x & i\beta \partial_y \\ i\beta \partial_x & i\beta \partial_y & -\partial_x^2 - \partial_y \partial_y \end{pmatrix} \mathbf{E} = k_0^2 \epsilon \mathbf{E} . \quad (3.4)$$

Now, this is a complicated coupled system of partial differential equations. First, because it is redundant: there are three field components, but only two polarization states. Second, because the searched for propagation constant β appears linearly and quadratic, so (3.4) is not a proper eigenvalue problem.

Applying the divergence operator results in zero. We therefore may express the z -component of the electric field strength as

$$-i\beta E_z = \epsilon^{-1} \partial_x \epsilon E_x + \epsilon^{-1} \partial_y \epsilon E_y . \quad (3.5)$$

Inserting this into (3.4) gives

$$\begin{pmatrix} k_0^2\epsilon + \partial_x\epsilon^{-1}\partial_x\epsilon + \partial_y^2 & \partial_x\epsilon^{-1}\partial_y\epsilon - \partial_x\partial_y \\ \partial_y\epsilon^{-1}\partial_x\epsilon - \partial_y\partial_x & k_0^2\epsilon + \partial_x^2 + \partial_y\epsilon^{-1}\partial_y\epsilon \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} = \beta^2 \begin{pmatrix} E_x \\ E_y \end{pmatrix}. \quad (3.6)$$

This is now a proper eigenvalue problem for β^2 .

In most cases the waveguide is sufficiently broad such that $\epsilon\partial_y \approx \partial_y\epsilon$ is a good approximation. If we set $E_x \approx 0$ and solve

$$(\partial_x^2 + \partial_y^2 + k_0^2\epsilon)E_y = \beta^2 E_y \quad (3.7)$$

we have found an approximate solution. It is called a quasi TE mode since the normal component of the electrical field strength vanishes, at least approximately. (3.7) is the well known Helmholtz equation.

The counterpart is a quasi TM mode. It is derived from the alternative wave equation

$$\begin{pmatrix} k_0^2\epsilon + \partial_x^2 + \epsilon\partial_y\epsilon^{-1}\partial_y & \partial_x\partial_y - \epsilon\partial_y\epsilon^{-1}\partial_x \\ \partial_y\partial_x - \epsilon\partial_x\epsilon^{-1}\partial_y & k_0^2\epsilon + \epsilon\partial_x\epsilon^{-1}\partial_x + \partial_y^2 \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix} = \beta^2 \begin{pmatrix} H_x \\ H_y \end{pmatrix}. \quad (3.8)$$

(3.8) results from (1.27) by inserting the mode form (3.1) and eliminating H_z by making use of $\partial_x H_x + \partial_y H_y + i\beta H_z = 0$.

Assuming again $\epsilon\partial_y \approx \partial_y\epsilon$ and setting $H_x \approx 0$ results in

$$(\epsilon\partial_x\epsilon^{-1}\partial_x + \partial_y^2 + k_0^2\epsilon)H_y = \beta^2 H_y, \quad (3.9)$$

a generalized Helmholtz equation.

Compare these results with the mode equations for planar waveguides. Now the permittivity profile and the mode fields depend on the cross section coordinates x and y . In quasi TE and TM approximation the only change is the additional ∂_y^2 operator.

3.2 Finite difference method

Let us work out a simple example. We want to find out the first two guided TE modes of a rib waveguide. In order to keep things as simple as possible we refrain from discussing continuity requirements at interfaces between different media, we simply smooth out permittivities there.

We choose a computational window the boundaries of which imitate infinity. The field has to vanish there. The cross section is represented by points $(x_i, y_j) = (ih_x, jh_y)$ within the computational window.

The first step is to setup the permittivity profile. We describe a yttrium iron garnet as substrate and a modified garnet as rib. All lengths are in microns.

```
1  % this file is rib_wg.m
2  EC=1.00; % cover permittivity
3  ES=3.80; % substrate permittivity
```

```

4  ER=5.20; % rib permittivity
5  NX=50;
6  NY=80;
7  % lengths are in micrometers
8  x=linspace(0,2.5,NX);
9  y=linspace(0,4.0,NY);
10 xlo=1.5; xhi=2.0;
11 ylo=1.5; yhi=2.5;
12 [X,Y]=meshgrid(x,y);
13 RIB=(X>xlo)&(X<xhi)&(Y>ylo)&(Y<yhi);
14 SUB=(X<=xlo);
15 COV=(X>=0)&~RIB;
16 prm=ES*SUB+ER*RIB+EC*COV;

```

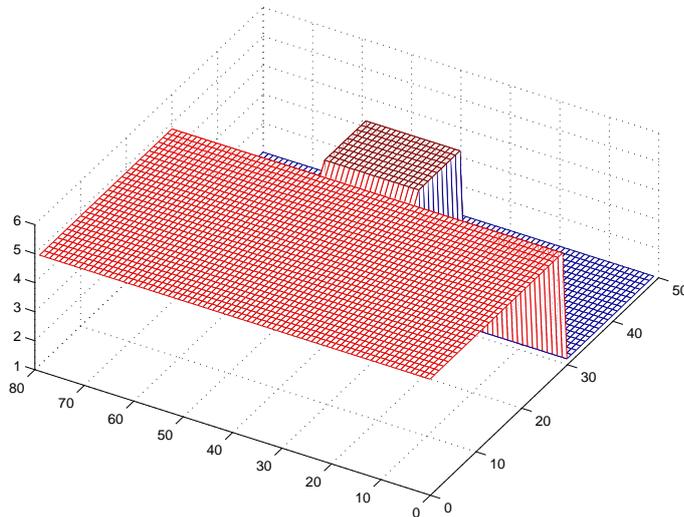


Figure 4: The computational window and the permittivity profile of our example.

The waveguide is operated with infrared light:

```

17 LAMBDA=1.500; % infrared light
18 k0=2*pi/LAMBDA;

```

A subprogram `helmholtz.m` establishes the sparse matrix H to be diagonalized which is achieved by `eigs`. The first argument is the sparse matrix, the second denotes the number of desired eigenvalues and eigenvectors, the third indicates that the largest algebraic eigenvalues are searched for. The diagonalized matrix (d , 2×2) and a matrix of column eigenvectors (u , $NX \times NY \times 2$) is returned.

```

19 HX=x(2)-x(1);
20 HY=y(2)-y(1);
21 H=helmholtz(HX,HY,k0*k0*prm);

```

```

22 [u,d]=eigs(H,2,'la');
23 mode1=reshape(u(:,1),NX,NY);
24 mode2=reshape(u(:,2),NX,NY);

```

Note that each mesh index pair (i, j) is one running index n . Therefore, the resulting eigenvectors `mode1` and `mode2` have to be reshaped to the computational window.

And here comes the code for establishing the Helmholtz operator $H = \Delta + d(x, y)$. The two dimensional Laplacian $\Delta = \partial_x^2 + \partial_y^2$ is represented by adding the expressions for the second derivatives, with possibly different spacing at the two dimensions. In general, indices $n, n - 1, n + 1, n - N_x$ and $n + N_x$ are linked by non-vanishing entries. This does not apply at the borders where the links must be set to zero.

```

1 % this file is helmholtz.m
2 function H=helmholtz(HX,HY,d)
3 [NX,NY]=size(d');
4 N=NX*NY;
5 ihx2=1/HX/HX; ihy2=1/HY/HY;
6 md=-2*(ihx2+ihy2)*ones(N,1)+reshape(d',NX*NY,1);
7 xd=ihx2*ones(N,1);
8 yd=ihy2*ones(N,1);
9 H=spdiags([yd,xd,md,xd,yd],[-NX,-1,0,1,NX],N,N);
10 for n=NX:NX:N-NX
11     H(n,n+1)=0;
12     H(n+1,n)=0;
13 end;

```

Note that the device is symmetric with respect to reflection at the x -axis. The lowest order mode is also symmetric while the first excited mode is antisymmetric.

In our case we have to diagonalize a 2950×2950 matrix which would require 72 MB of storage. However, most entries vanish, the matrix is sparsely populated. In our case there are only 14532 non-vanishing entries. It is therefore advisable to store the matrix as a list of index pairs and values of its non-vanishing entries.

Diagonalizing such a sparse matrix is virtually impossible since the matrix of eigenvectors is not sparse. Therefore, only a few eigenvectors closest to a value σ are calculated by an iterative algorithm. In our case, we have specified that the two largest eigenvalues are to be worked out.

Sparse matrix technology is essential for solving partial differential equations.

3.3 Various other methods

The method of finite differences is simple to program. However, it is almost always only the second best choice. One of the disadvantages is that the finite difference method is simple only if the mesh is equally spaced. Although the field changes most rapidly within the rib region, the computational window

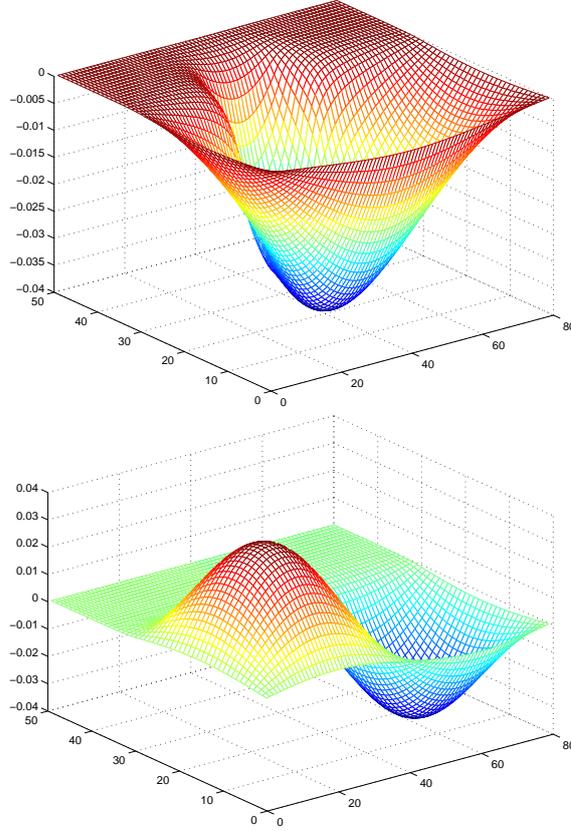


Figure 5: Ground and first excited TE mode of a typical rib waveguide.

should be large enough so that $E = 0$ at the boundary is a good approximation for $E(x, y) \rightarrow 0$ with $|x|, |y| \rightarrow \infty$ (note that the field vanishes exponentially towards infinity).

3.3.1 Galerkin methods

Consider a region Ω and functions on it.

$$(g, f) = \int_{\Omega} dx dy g^*(x, y) f(x, y) \quad (3.10)$$

is a scalar product which defines a Hilbert space \mathcal{H} . We look for solutions of the Helmholtz equation

$$(\partial_x^2 + \partial_y^2 + u)f = \Lambda f \quad (3.11)$$

with Dirichlet boundary conditions $f = 0$ on $\partial\Omega$. This problem is equivalent to demanding

$$-(\partial_x g, \partial_x f) - (\partial_y g, \partial_y f) + (g, u f) = \Lambda(g, f) \quad (3.12)$$

for all $g \in \mathcal{H}$. (3.13) is the weak form of the partial differential equation. Only first derivatives are involved.

We choose a set of square integrable expansion functions f_1, f_2, \dots with $f_j = 0$ on $\partial\Omega$. We select another set of square integrable test functions g_1, g_2, \dots . Now

$$H = \Lambda D \tag{3.13}$$

has to be solved where $H_{ji} = -(\partial_x g_j, \partial_x f_i) - (\partial_y g_j, \partial_y f_i) + (g_j, u f_i)$ and $D_{ji} = (g_j, f_i)$.

The Galerkin approximation method consists in choosing only a finite number N of expansion and test functions. Then (3.13) is an ordinary (generalized) eigenvalue problem. It becomes particularly simple if the set of expansion and test functions are the same and if they are normalized and mutually orthogonal such that $D_{ij} = \delta_{ij}$.

Trigonometric functions One may choose products $\sin(k_a x) \sin(k_b y)$ functions. They can be differentiated and integrated analytically, and by choosing proper k values, the boundary conditions on a rectangle may be met easily.

Finite element method This is today's method of choice. Ω is divided into triangles. Each interior points is the corner of two or more triangles. A tent functions is continuous, assigns the value 1 to the interior points, is linear in the adjacent triangles, and vanishes outside of them. There is one tent function for each interior point. Such tent functions are easy to construct and to differentiate. As opposed to trigonometric functions, they are localized. If $a = 1, 2, \dots$ enumerates the interior points (x_a, y_a) and if $t_a(x, y)$ are the corresponding tent functions, then the expansion

$$f(x, y) = \sum \phi_a t_a(x, y) \tag{3.14}$$

guarantees $f(x_a, y_a) = \phi_a$. Put otherwise, the field values at the interior points are the coefficients of a decomposition into tent functions.

In most cases, the set of tent functions serve as expansion as well as test functions.

There are software packages which provide for the triangulization of arbitrary regions, which set up the required matrices and which allow for refinements of the triangulization. The matrices involved are sparse. It is beyond the scope of this lecture series to introduce a finite element tool package.

3.3.2 Method of lines

The method of finite differences consists in covering the computational window by $N_x \times N_y$ representative points and approximating differential operators by finite differences. In many cases the computational windows may be covered by straight lines, N_y say. The field is no longer represented by $N_x \times N_y$ variables, but by N_y functions $f_i = f_i(x)$, for $i = 1, 2, \dots, N_y$. We thus have approximated the partial differential equation by a finite system of ordinary differential

equations which in many cases can be solved quasi-analytically. The method of line, if applicable, is usually the most precise method although more difficult to program than the finite difference method.

3.3.3 Collocation methods

An interesting approach is to expand the searched for solution into products of orthogonal functions which can be differentiated analytically. At suitably chosen points the partial differential equation is solved exactly which gives rise to linear equations for the expansion coefficients. In particular, Gauss-Hermite functions have been tried which are Gaussians multiplied by polynomials (harmonic oscillator eigenfunctions). The charm of such an expansion is that solutions automatically vanish rapidly at infinity, as they should.

However, there is a serious flaw. Orthogonal polynomials necessarily have coefficients of alternating sign. The small sum is the result of an almost total cancellation of large contributions, and by a rule of thumb, not more than 60 terms can be summed before running into serious rounding error problems.

4 Propagation

So far we have discussed guided modes. They have definite propagation constants and are either TE or TM polarized. In many cases, however, mode analysis is not sufficient, either because the structure under study is not z -homogeneous or because energy is radiated off. In this section we want to describe the propagation of a beam. For simplicity, the scalar equation of a quasi TE polarized wave is discussed.

4.1 Fresnel equation

We relax the requirement that the field is a plane wave with respect to propagation along the z axis. Instead we write

$$E(t, x, y, z) = e^{-i\omega t} e^{i n k_0 z} E(x, y; z). \quad (4.1)$$

$\beta = n k_0$ is the carrier spatial frequency, it should be chosen such that $E(x, y; z)$ depends as weakly as possible on the propagation coordinate z . Put otherwise: we discuss situations where light is almost a mode travelling in z direction.

We have to solve the following wave equations⁸:

$$e^{-i n k_0 z} \left(\partial_x^2 + \partial_y^2 + \frac{d^2}{dz^2} \right) e^{i n k_0 z} E(z) + k_0^2 \epsilon E(z) = 0. \quad (4.2)$$

Because the fields $E(z)$ are assumed to depend but weakly on z , we neglect its second derivative. The result is

$$-i E'(z) = P E(z) \quad \text{where} \quad P = \frac{\partial_x^2 + \partial_y^2 - k_0^2 \delta \epsilon}{2 n k_0}, \quad (4.3)$$

the well-known Fresnel equation, with $\delta \epsilon(x, y) = \epsilon(x, y) - n^2$. P is an operator acting on fields $x, y \rightarrow f(x, y)$.

(4.3) is not of second, but of first order with respect to propagation: by insisting on a weak z -dependency of $E(z)$ we have singled out the forward propagating beam.

The Fresnel equation describes the propagation of a beam well provided n is chosen properly, namely such that

$$\|E''(z)\| \ll k_0 \|E'(z)\| \quad (4.4)$$

holds true⁹.

Note that in a homogeneous medium $E(x, y; z) = A$ is a solution if $n^2 = \epsilon$. This describes a plane wave travelling in z direction.

In the following subsections we discuss various methods how to solve the Fresnel equation.

⁸ $x, y \rightarrow E(x, y; z)$ is regarded as a family of fields being parameterized by z

⁹The norm of a field is defined by $\|f\|^2 = \int dx dy |f(x, y)|^2$

4.2 Finite differences

Let us assume that $\epsilon = \epsilon(x, y)$ does not depend on the propagation coordinate z . In this case (4.3) is formally solved by

$$E(z) = e^{-i z P} E(0) \quad (4.5)$$

or, discussing a propagation step h ,

$$E(z+h) = e^{-i h P} E(z). \quad (4.6)$$

A very crude approximation is

$$E(z+h) = (I - i h P) E(z). \quad (4.7)$$

Likewise, one may write

$$e^{i h P} E(z+h) = E(z) \quad (4.8)$$

and approximate by

$$E(z+h) = (I + i h P)^{-1} E(z). \quad (4.9)$$

We assume that the operator P is also represented by a finite difference scheme on the cross section x, y . It turns out that propagating forward in time by (4.7) is unstable. Sending all propagation steps to zero so that more and more propagation steps are required does not converge. Propagating backward in time by (4.9) is more cumbersome because a system of linear equations has to be solved, however, the method is stable. Both are of first order in h .

A combination of the two methods is one order more accurate and stable as well. We set

$$E(z + \frac{h}{2}) = (I - \frac{i h P}{2}) E(z) = (I + \frac{i h P}{2}) E(z + h) \quad (4.10)$$

and obtain

$$E(z+h) = (I + \frac{i h P}{2})^{-1} (I - \frac{i h P}{2}) E(z), \quad (4.11)$$

the Crank-Nicholson scheme.

Let us try out the Crank-Nicholson scheme by propagating a Gaussian beam in empty space.

```

1  % this file is refl_bc.m
2  CW=10.0; % computational window
3  BW=1.0; % Gaussian beam width
4  LAMBDA=0.633; % helium neon laser

```

```

5  n=1.0; % refractive index of the medium
6  k0=2*pi/LAMBDA;
7  NX=128; % points on x axis
8  HX=CW/(NX-1); % x axis spacing
9  HZ=5*HX; % propagation step
10 x=linspace(-0.5*CW,0.5*CW,NX)';
11 E=exp(-(x/BW).^2); % initial field
12 u=0.5i*HZ/(2*n*k0);
13 main=-2*u*ones(NX,1)/HX^2;
14 next=u*ones(NX-1,1)/HX^2;
15 % step forward
16 FW=eye(NX)+diag(next,-1)+diag(main,0)+diag(next,1);
17 % step backward
18 BW=eye(NX)-diag(next,-1)-diag(main,0)-diag(next,1);
19 NZ=100; % number of propagation steps
20 hist=zeros(NX,NZ); % storage for history
21 for r=1:NZ
22     hist(:,r)=abs(E).^2;
23     E=BW\FW*E;
24 end;

```

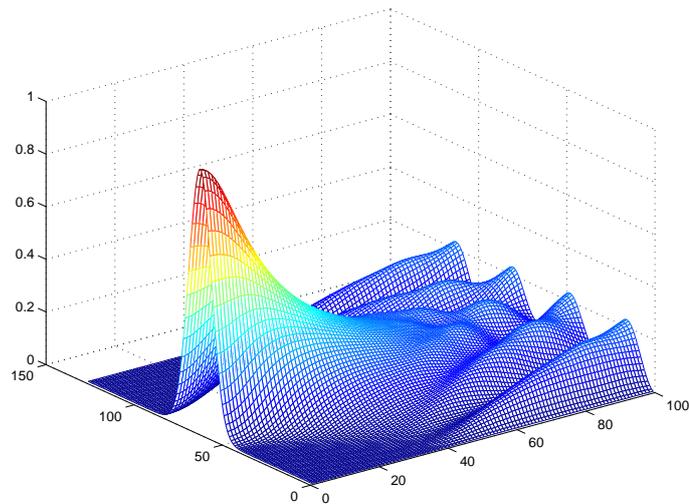


Figure 6: Propagation of a Gaussian beam by the Crank-Nicholson finite difference scheme. The field was silently assumed to vanish outside the computational window.

4.3 Transparent boundary conditions

Figure 6 shows the surprising result of propagating a Gaussian beam in free space. The boundary of the computational window obviously reflects the field, and we see a superposition of the propagated and multiply reflected field.

The reason for this odd behaviour is easily discovered. When constructing the Laplacian we did not care for the second derivative at the boundary. The first row of P is proportional to $-2, 1, 0, \dots$, the second row is $1, -2, 1, 0, \dots$, the third reads $0, 1, -2, 1, 0, \dots$, etc. The first row represents the second derivative at the boundary only if the field outside the computational window vanishes as if the computational window were surrounded by an ideally conducting material, which is not intended.

It was customary to implement absorbing boundary conditions. Outside the computational window an absorbing material was simulated. If absorption sets in too rapidly, it caused reflection. If it sets in too gently, the computational window became too large. Hence, all beam propagation calculations had to be hand-tuned.

Hadley¹⁰ in 1992 devised a clever algorithm how to avoid these nasty reflections and the computational window overhead.

The idea is rather simple. Determine from the field values close to the boundary the components of an outgoing and an incoming wave and suppress the latter.

Assume that $f_j^r = f(jh_x; rh_z)$ is already known, where $j = 1, 2, \dots, N$. From

$$e^{ikh_x} = \frac{f_N^r}{f_{N-1}^r} \quad (4.12)$$

we determine the wave number at the upper boundary. If its real part is positive—fine, it is an outgoing wave. If it is negative, then k is modified to $\bar{k} = \text{Im } k$. Now a factor $\gamma = \exp(i\bar{k}h_x)$ is calculated such that $f_{N+1}^r = \gamma f_N^r$ is used to work out the second derivative at the boundary. An analogous procedure is applied to the lower boundary.

The following program implements such transparent boundary conditions.

```

1  % this file is transp_bc.m
2  CW=10.0; % computational window
3  BW=1.0; % Gaussian beam width
4  LAMBDA=0.633; % helium neon laser
5  k0=2*pi/LAMBDA;
6  NX=128; % points on x axis
7  HX=CW/(NX-1); % x axis spacing
8  HZ=5*HX; % propagation step
9  x=linspace(-0.5*CW,0.5*CW,NX)';
10 E=exp(-(x/BW).^2); % initial field
11 u=0.5i*HZ/(2*k0);
12 main=-2*u*ones(NX,1)/HX^2;
13 next=u*ones(NX-1,1)/HX^2;
14 FW=eye(NX)+diag(next,-1)+diag(main,0)+diag(next,1); % forward
15 BW=eye(NX)-diag(next,-1)-diag(main,0)-diag(next,1); % backward
16 NZ=100; % number of propagation steps
17 hist=zeros(NX,NZ); % storage for history
18 for r=1:NZ

```

¹⁰G.R.Hadley, Transparent boundary condition for the beam propagation method, IEEE Journal of Quantum Electronics 28 (1992) 363-370

```

19 hist(:,r)=abs(E).^2;
20 E=one_step(HX,u,FW,BW,1e-4,E);
21 end;

```

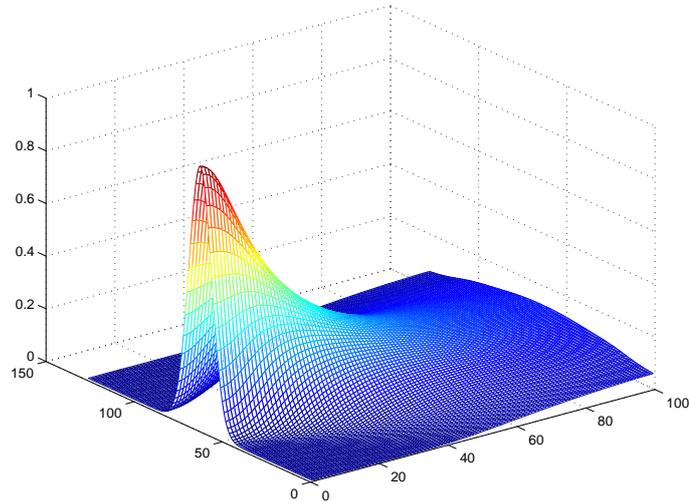


Figure 7: Propagation of a Gaussian beam by the Crank-Nicholson finite difference scheme. Transparent boundary conditions have been implemented.

The function `one_step` propagates the field by one step. It implements boundary conditions which guarantee transparency.

```

1 % this file is one_step.m
2 function new=one_step(HX,u,FW,BW,TINY,old)
3 NX=size(old,1);
4 FF=FW;
5 BB=BW;
6 if abs(old(1))>TINY
7 k=i/HX*log(old(2)/old(1));
8 if real(k)<0
9 k=i*imag(k);
10 end;
11 tbc=exp(i*k*HX)*u/HX^2;
12 FF(1,1)=FF(1,1)+tbc;
13 BB(1,1)=BB(1,1)-tbc;
14 end;
15 if abs(old(NX))>TINY
16 k=-i/HX*log(old(NX)/old(NX-1));
17 if real(k)<0
18 k=i*imag(k);
19 end;
20 tbc=exp(i*k*HX)*u/HX^2;
21 FF(NX,NX)=FF(NX,NX)+tbc;

```

```

22     BB(NX,NX)=BB(NX,NX)-tbc;
23     end;
24     new=BB\FF*old;

```

Figure 7 shows the improvement. Energy may pass the boundaries of the computational window, and no reflections show up.

4.4 Propagation in a slab waveguide

In subsection 2.4 we have discussed a slab waveguide. A thin film on a glass substrate is covered by air. We found, for helium-neon laser light, that there are two guided TE and two guided TM modes. Let us now study how a Gaussian beam entering at $z = 0$ will propagate. We choose the substrate refractive index as reference index.

The previous programs have to be changed only slightly. Here is the listing:

```

1  % this file is prp_swg.m
2  function [x,z,hist]=prp_swg(BC);
3  % BC is the center of the Gaussian beam
4  CW=10.0; % computational window
5  LAMBDA=0.633; k0=2*pi/LAMBDA; % helium neon laser
6  NX=128; % points on x axis
7  HX=CW/(NX-1); % x axis spacing
8  HZ=10*HX; % propagation step
9  FW=1.8; % film width
10 EC=1.00; % cover permittivity
11 ES=1.49; % substrate permittivity
12 EF=1.52; % film permittivity
13 BW=1.0; % Gaussian beam width
14 x=linspace(-0.5*CW,0.5*CW,NX)';
15 prm=ES*(x<0)+EF*((x>=0)&(x<=FW))+EC*(x>FW);
16 nref=sqrt(ES);
17 deps=prm-nref^2;
18 E=exp(-((x-BC)/BW).^2); % initial field
19 u=0.5i*HZ/(2*nref*k0);
20 main=u*(-2*ones(NX,1)/HX^2+k0^2*deps);
21 next=u*ones(NX-1,1)/HX^2;
22 FW=eye(NX)+diag(next,-1)+diag(main,0)+diag(next,1); % forward
23 BW=eye(NX)-diag(next,-1)-diag(main,0)-diag(next,1); % backward
24 NZ=250; % number of propagation steps
25 hist=zeros(NX,NZ); % storage for history
26 for r=1:NZ
27     hist(:,r)=abs(E).^2;
28     E=one_step(HX,u,FW,BW,1e-4,E);
29 end;
30 z=[0:HZ:(NZ-1)*HZ];

```

We present the intensity as a contour plot in Figure 8. Note that power is radiated mainly into the substrate the permittivity of which is close to that of the film.

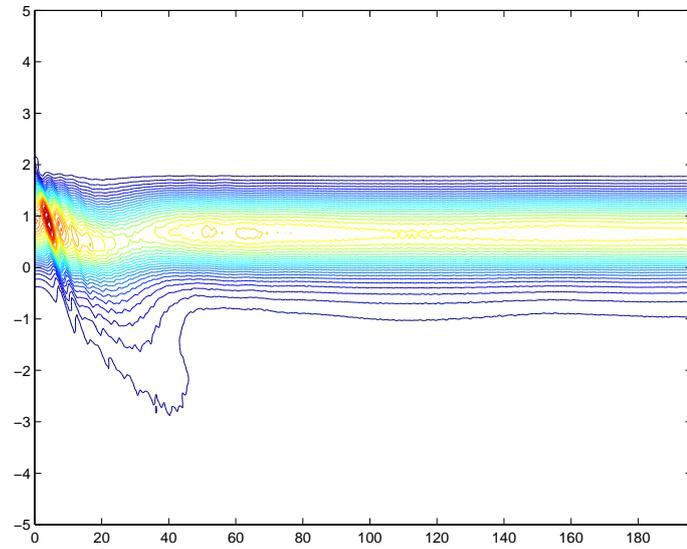


Figure 8: A Gaussian beam is inserted into a slab waveguide. The beam center is at the middle of the film. Propagation is from left to right.

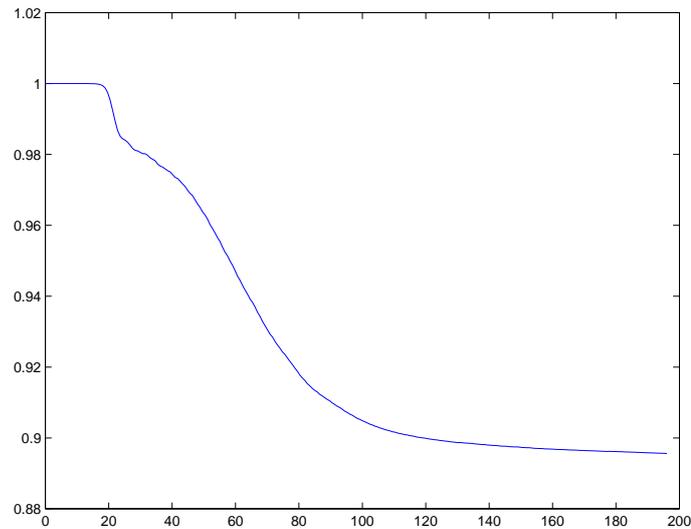


Figure 9: Power within the computational window vs. propagation distance for the previous propagation calculation.

It is interesting to study the power within the computational window which we have depicted in Figure 9. For a short distance it remains constant because the beam has not yet reached the boundaries of the computational window. It then falls off, and after a certain length it becomes constant because the mode is well guided within the computational window. About 11% of the original power are radiated off.

If the center of the Gaussian beam is at the interface between film and substrate, the situation looks different, as shown in Figure 10. Now 27% of the incident power are radiated off.

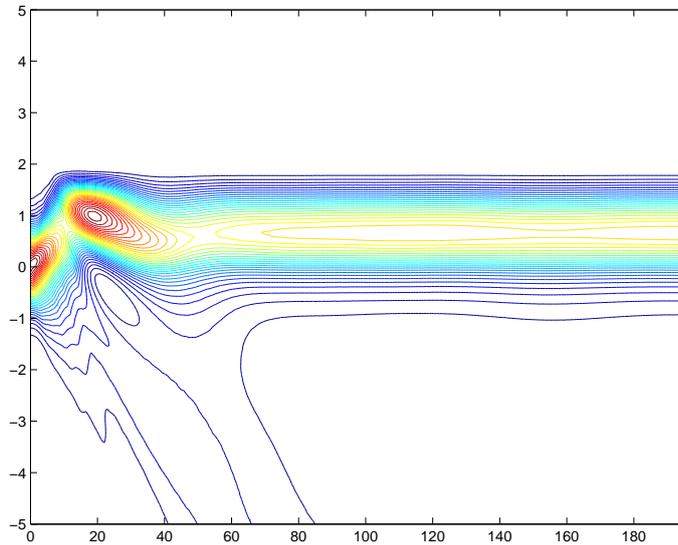


Figure 10: The center of the Gaussian beam is at the interface between film and substrate

In order to minimize insertion losses, a smooth transition from one waveguide (a glass fiber, say) to another waveguide (a rib waveguide, for example) has to be provided for. How to design such a taper is outside the scope of these lectures.

4.5 Other propagation methods

Although we have demonstrated the finite difference propagation method for planar waveguides, all interesting structures are made of strip waveguides with a two-dimensional cross section. In section 3 we discussed that the linear operators to be dealt with can be approximated by sparse matrices. A cross section mesh of 100×100 points corresponds to a $10^4 \times 10^4$ matrix and, if stored in full, requires 800 MByte of RAM. Since there is more than one such matrix involved, even two-dimensional problems would be out of scope of ordinary computers, not to mention three-dimensional structures. Note that the Crank-Nicholson propagation scheme requires to multiply $E(z)$ by $I - i h_z P/2$ to obtain $E(z + h_z/2)$ and to solve the system of linear equations $(I + i h_z P/2)E(z + h_z) = E(z + h_z/2)$. Fortunately there are algorithms for solving linear equations within the framework of sparse matrices. In the following we describe two rather popular alternatives to the finite difference propagation method.

4.5.1 Method of lines

The method of lines represents the x, y, z continuum by lines $x_i(z), y_j(z)$. The field $E = E(z)$ is represented by functions $E_a(z)$ where $a = (i, j)$. The Fresnel

equation now reads

$$-i E'_a = \sum_b P_{ab} E_b(z) . \quad (4.13)$$

This is a set of coupled ordinary differential equations, although a mighty set. The standard procedure to solve it is by diagonalizing the P matrix which represents the Fresnel operator on the waveguide cross section. We may write

$$P_{ab} = \sum_c U_{ac}^\dagger p_c U_{cb} , \quad (4.14)$$

where U is a unitary matrix because P is Hermitian. $\bar{E}_a = \sum_b U_{ab} E_b$ now is subject to a set of uncoupled differential equations

$$-i \bar{E}'_a = p_a \bar{E}_a \quad (4.15)$$

the solution of which is

$$\bar{E}_a(z) = e^{i p_a z} \bar{E}_a(0) \quad (4.16)$$

or

$$E_a(z) = \sum_b U_{ab}^\dagger e^{i p_b z} U_{ba} E_a(0) . \quad (4.17)$$

As said above, working out the exponential of $i P z$ by diagonalization is prohibitive for two-dimensional cross section. A few of the eigenvalues correspond to guided modes, the remaining thousands of eigenvalues refer to radiation modes. Incoming radiation has to be suppressed, outgoing radiation may be represented by a few lossy modes. By a tricky balance between simplicity, storage requirement, run-time and accuracy the method of lines has proven to be a serious alternative to the finite difference method.

4.5.2 Operator splitting

This is the oldest beam propagation scheme¹¹. It is tailored to the computational facilities of more than 30 years ago: random access memory was short and had to be replaced by long program run times. There is an intuitive and a mathematical foundation.

The Fresnel propagation operator consists of two parts. The first one is a cross section Laplacian which describes propagation in free space. As we know, any beam of light diverges when propagating in free space. The second contribution characterizes focussing. The guiding structure has a higher refractive index than the surrounding. The optical path through the centre of a lens is shorter than an off-axis optical path which effect leads to focussing. The Fresnel equation describes propagation in free space and focussing in infinitely rapid succession.

¹¹J. A. Fleck, J. R. Morris, M. D. Feit: Appl. Phys. **10** (1976) 129

We approximate it by finite steps of propagation in free space and focussing by a permittivity profile.

Mathematically, the Fresnel propagation operator $P = D + M$ is the sum of a differential operator

$$D = \frac{\partial_x^2 + \partial_y^2}{2nk_0} \quad (4.18)$$

and a multiplication operator

$$M = \frac{k_0 \delta \epsilon}{2n} . \quad (4.19)$$

Recall that n is a reference index and $\delta \epsilon = \epsilon(x, y) - n^2$. Both act on fields depending on the cross section coordinates x, y .

D describes the propagation in a homogeneous medium of refractive index n while M characterizes the focussing effect. These operators do not commute, and the rule $\exp(A + B) = \exp(A) \exp(B)$ is not applicable.

However,

$$e^{i h_z (D + M)} \approx e^{i h_z D / 2} e^{i h_z M} e^{i h_z D / 2} \quad (4.20)$$

is an approximation, the error being proportional to h_z^3 .

The operator splitting expression, as given by the right hand side of (4.20), is applied as follows.

The field is Fourier transformed,

$$\hat{E}_1(k_x, k_y) = \int dx dy e^{i(xk_x + yk_y)} E_1(x, y) . \quad (4.21)$$

Then

$$\hat{E}_2(k_x, k_y) = e^{-i h_z (k_x^2 + k_y^2) / 4nk_0} \hat{E}_1(k_x, k_y) \quad (4.22)$$

is calculated. Fourier back transformation yields

$$E_2(x, y) = \int \frac{dk_x}{2\pi} \frac{dk_y}{2\pi} e^{-i(xk_x + yk_y)} \hat{E}_2(k_x, k_y) . \quad (4.23)$$

We now set

$$E_3(x, y) = e^{i h_z k_0 \delta \epsilon(x, y) / 2n} E_2(x, y) \quad (4.24)$$

and Fourier transform to

$$\hat{E}_3(k_x, k_y) = \int dx dy e^{i(xk_x + yk_y)} E_3(x, y) . \quad (4.25)$$

Finally,

$$\hat{E}_4(k_x, k_y) = e^{-i h_z (k_x^2 + k_y^2) / 2n k_0} \hat{E}_3(k_x, k_y) \quad (4.26)$$

is worked out which is the last part of the first propagation step and the first part of the next. This sequence of operations is continued with possibly modifying $\delta\epsilon$ if the structure changes along the propagation direction.

Note that the method of operator splitting requires no additional storage for matrices but depends on a fast Fourier transform algorithm. Also note that each single propagation step is described by applying a unitary matrix. Hence, the norm $\|E\|$ of the original field remains constant. Without transparent boundary conditions or another means to avoid reflections the operating splitting propagation method runs into the same problems as discussed earlier.

5 Optical Anisotropy

Up to now we always assumed that the electrical field strength \mathbf{E} and the dielectric displacement \mathbf{D} were parallel. This is true for optically isotropic media, like glass, silicon, or other cubic crystals or amorphous substances. Therefore, we always talked about permittivity as a scalar, as in $D_i = \epsilon \epsilon_0 E_i$. In general it is still true that, for sufficiently small electric fields, displacement and field strength are proportional, but they are not necessarily parallel. We therefore have to write¹² $D_i = \epsilon_{ij} \epsilon_0 E_j$, with the permittivity tensor ϵ_{ij} . This section discusses effects of optical anisotropy.

5.1 Permittivity tensor

We talk about a system of particles at \mathbf{x}_a with charge q_a . The polarization at \mathbf{x} is

$$P_i(\mathbf{x}) = \sum_a q_a x_i \delta^3(\mathbf{x} - \mathbf{x}_a). \quad (5.1)$$

Denote by H the Hamiltonian of matter which is perturbed by the interaction with an electric field $\mathbf{E}(t, \mathbf{x})$. The Hamiltonian now is

$$H_t = H - \int d^3x P_i(\mathbf{x}) E_i(t, \mathbf{x}). \quad (5.2)$$

We assume that the system was in a Gibbs state

$$G = e^{(F - H)/k_B T} \quad (5.3)$$

before the perturbation has set in. F is the free energy, k_B denotes Boltzmann's constant, and T is the temperature of the equilibrium state.

Within the framework of linear response theory, the effect of the perturbation on the polarization may be worked out:

$$P_i(t, \mathbf{x}) = \int_0^\infty d\tau \int d^3\xi \Gamma_{ij}(\tau, \boldsymbol{\xi}) E_j(t - \tau, \mathbf{x} - \boldsymbol{\xi}). \quad (5.4)$$

Here $P_i(t, \mathbf{x}) = \langle P_i(\mathbf{x}) \rangle_t$ is the expectation value of the polarization in the time-dependent perturbed state. The Green's function Γ_{ij} is found to be

$$\Gamma_{ij}(\tau, \boldsymbol{\xi}) = \left\langle \frac{i}{\hbar} [U_{-\tau} P_i(\boldsymbol{\xi}) U_\tau, P_j(0)] \right\rangle \quad (5.5)$$

where U_t denotes time translation by the unperturbed Hamiltonian,

$$U_t = e^{-itH/\hbar}. \quad (5.6)$$

¹²Einstein's summation convention: a sum over doubly occurring indices is silently understood.

The expectation value in (5.5) is with the unperturbed Hamiltonian. Put otherwise, the unperturbed system knows how to react on a perturbation.

By Fourier transforming all quantities¹³ we arrive at

$$\hat{P}_i(\omega, \mathbf{q}) = \epsilon_0 \chi_{ij}(\omega, \mathbf{q}) \hat{E}_j(\omega, \mathbf{q}) \quad (5.7)$$

where the susceptibility tensor χ_{ji} is

$$\chi_{ij}(\omega, \mathbf{q}) = \frac{1}{\epsilon_0} \int_0^\infty d\tau e^{i\omega\tau} \int d^3\xi e^{-i\mathbf{q}\cdot\xi} \Gamma_{ij}(\tau, \xi). \quad (5.8)$$

The permittivity $\epsilon_{ij} = \delta_{ij} + \chi_{ij}$ can thus be calculated, in principle. It depends on angular frequency ω , on a wave vector \mathbf{q} , and on all parameters which enter the Gibbs state, such as temperature or quasi-static external fields.

Matter in its solid state has a time scale which is governed by the speed of sound. Since this is so much smaller than the speed of light, we may write

$$\chi_{ij}(\omega, \mathbf{q}) = \chi_{ij}^{(0)}(\omega) + \chi_{ijk}^{(1)} q_k + \dots \quad (5.9)$$

The first contribution is the ordinary susceptibility, it depends on angular frequency only. The second term describes optical activity. It is responsible for rotating the polarization vector of a linearly polarized wave by a few degrees per centimeter which is a tiny effect. We will not discuss optical activity here and denote the susceptibility by $\chi_{ij}(\omega) = \chi_{ij}^{(0)}(\omega)$ henceforth.

The permittivity should be split into a Hermitian and an anti-Hermitian contribution, $\epsilon_{ij} = \epsilon'_{ij} + i\epsilon''_{ij}$. The Hermitian, or refractive part ϵ' is responsible for refraction, for the bending of light rays. The anti-Hermitian, or absorptive contribution ϵ'' causes absorption. The dissipation-fluctuation theorem assures that ϵ'' is non-negative in accordance with the second law of thermodynamics: light energy is transformed into heat, the opposite is impossible.

The refractive and absorptive parts of the permittivity tensor are not independent, they obey the Kramer-Kronigs dispersion relation

$$\epsilon'_{ij}(\omega) = \delta_{ij} + \int \frac{du}{\pi} \frac{\epsilon''_{ij}(u)}{u - \omega}, \quad (5.10)$$

where the principal value integral is understood. There is no refraction without absorption. It is however possible that the absorptive part of the permittivity tensor almost vanishes in an entire frequency range. We then speak of a transparency window.

Let us mention another useful relation which can be derived by studying (5.5) with respect to time reversal. It turns out that the permittivity is a symmetric tensor provided that an external magnetic field is reversed. Denoting external, or quasi-static fields by \vec{E} and \vec{B} , Onsager's relations amount to

$$\epsilon_{ij}(\omega, \vec{E}, \vec{B}) = \epsilon_{ji}(\omega, \vec{E}, -\vec{B}). \quad (5.11)$$

¹³The Fourier transform of a convolution is the product of the Fourier transforms.

5.2 Anisotropic waveguides

We assume a material without magneto-optic effect. By Onsager's relation the permittivity tensor is symmetric and can be diagonalized by an orthogonal coordinate transformation. We assume the axes of the coordinate system to coincide with the optical axes. In this case the permittivity is a diagonal matrix with entries $\epsilon_x, \epsilon_y, \epsilon_z$.

Let us comment first on slab waveguides.

Our previous analysis, that there are TE and TM modes, remains valid.

$$\mathbf{E} = \begin{pmatrix} 0 \\ E \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \frac{1}{i\omega\mu_0} \begin{pmatrix} -i\beta E \\ 0 \\ E' \end{pmatrix} \quad (5.12)$$

solves Maxwell's equations provided the TE mode equation

$$E'' + k_0^2 \epsilon_y E = \beta^2 E \quad (5.13)$$

is satisfied. E and E' have to be continuous.

Likewise,

$$\mathbf{E} = \frac{1}{-i\omega\epsilon_0} \begin{pmatrix} -i\beta H/\epsilon_x \\ 0 \\ H'/\epsilon_z \end{pmatrix} \quad \text{and} \quad \mathbf{H} = \begin{pmatrix} 0 \\ H \\ 0 \end{pmatrix} \quad (5.14)$$

solve all Maxwell equations if the TM mode equation

$$\epsilon_x \frac{d}{dx} \epsilon_z^{-1} \frac{d}{dx} H + k_0^2 \epsilon_x H = \beta^2 H \quad (5.15)$$

holds true. H and H'/ϵ_z must be continuous.

For strip waveguides the mode equations become very complicated, and we refrain from discussing the various approximation schemes which have been put forward. The deviations from isotropy are usually rather small and may be dealt with as a perturbation.

So let us discuss a waveguide with permittivity

$$\epsilon_{ij} = \bar{\epsilon} \delta_{ij} + \Delta \epsilon_{ij} . \quad (5.16)$$

For the following discussion the inverse tensor

$$\eta_{ij} = \bar{\eta} + \Delta \eta_{ij} \quad (5.17)$$

is the more suitable quantity, where $\bar{\eta} = 1/\bar{\epsilon}$ and $\Delta \eta_{ij} = -\Delta \epsilon_{ij}/\bar{\epsilon}^2$.

The wave equation to be solved is (1.24), namely

$$\mathbf{curl} \epsilon^{-1} \mathbf{curl} \mathbf{H} = L \mathbf{H} = k_0^2 \mathbf{H} , \quad (5.18)$$

where

$$\mathbf{curl} = \begin{pmatrix} 0 & -i\beta & \partial_y \\ i\beta & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \end{pmatrix} \quad (5.19)$$

is the curl operator for a strip waveguide. Note that η in (5.18) is a tensor.

We define the following scalar product for vector fields:

$$(\mathbf{g}, \mathbf{f}) = \int dx dy g_i^*(x, y) f_i(x, y). \quad (5.20)$$

With this scalar product, the curl operator is symmetric. Since we assume a transparent medium, the permittivity tensor is Hermitian, and so is η . It follows that the mode operator L is also symmetric. Its eigenvalues k_0^2 may be obtained from

$$k_0^2 = \frac{(\mathbf{curl} \mathbf{H}, \eta \mathbf{curl} \mathbf{H})}{(\mathbf{H}, \mathbf{H})} = \Phi(\mathbf{H}). \quad (5.21)$$

We have introduced in (5.21) a functional $\mathbf{f} \rightarrow \Phi(\mathbf{f})$ of vector fields. It is a simple exercise to show that Φ is stationary at mode fields:

$$\left. \frac{d}{ds} \Phi(\mathbf{H} + s \delta \mathbf{H}) \right|_{s=0} = 0. \quad (5.22)$$

Now, the eigenvalues k_0^2 of (5.21) depend twofold on $\Delta\eta$, directly and indirectly via the dependency of the mode fields on the inverse permittivity. The latter effect, however, vanishes because the functional Φ is stationary at solutions of the wave equation. Therefore

$$\Delta k_0^2 = \frac{(\mathbf{curl} \mathbf{H}, \Delta\eta \mathbf{curl} \mathbf{H})}{(\mathbf{H}, \mathbf{H})} \quad (5.23)$$

holds true in first order perturbation theory.

This change in Δk_0^2 must be compensated by a shift $\Delta\beta$ of the propagation constant:

$$\Delta k_0^2 + \frac{dk_0^2}{d\beta^2} \Delta\beta^2 = 0 \quad (5.24)$$

where we again may exploit that the derivative $dk_0^2/d\beta^2$ does not depend implicitly on β . With

$$\frac{dk_0^2}{d\beta^2} = \frac{\int dx dy \bar{\eta}(H_x^2 + H_y^2)}{\int dx dy (H_x^2 + H_y^2 + H_z^2)} \quad (5.25)$$

we finally arrive at

$$\Delta\beta = -\frac{1}{2\beta} \frac{\int dx dy \{ \mathbf{curl} \mathbf{H} \}_i \Delta\eta_{ij} \{ \mathbf{curl} \mathbf{H} \}_j}{\int dx dy \bar{\eta} (H_x^2 + H_y^2)} . \quad (5.26)$$

There are many variations of this formula. The effect of anisotropy on the propagation constant may be expressed in the electric fields, this or that component can be removed by the divergence equation or because it is small, and so forth. (5.26) however is free from unnecessary approximations.

5.3 Non-reciprocal effects

We discuss in this subsection a special, but most important optical anisotropy, namely the Faraday effect. We shall explain why magnetism must be involved to achieve non-reciprocal light propagation.

5.3.1 The Faraday effect

Maxwell's equation $\epsilon_0 \mathbf{div} \mathbf{E} = \rho$, $\mathbf{div} \mathbf{B} = 0$, $\mathbf{curl} \mathbf{B}/\mu_0 = \mathbf{j} + \epsilon_0 \dot{\mathbf{E}}$ and $\mathbf{curl} \mathbf{E} = -\dot{\mathbf{B}}$ are compatible with time reversal. If $\mathbf{E}, \mathbf{B}, \rho$ and \mathbf{j} is a solution, then $\mathbf{E}^* = \mathbf{E}, \mathbf{B}^* = -\mathbf{B}, \rho^* = \rho$ and $\mathbf{j}^* = -\mathbf{j}$ is a solution as well, where $f^*(t, \mathbf{x}) = f(-t, \mathbf{x})$. It can be seen from $\mathbf{S} = \mathbf{E} \times \mathbf{H}$ that time reversal implies the reversal of motion. If there is a wave travelling in forward direction, there is an identical wave travelling backward. Not quite. The magnetic field has an oscillating part and may have a quasi-static contribution. That the oscillating part is reversed is a simple consequence of the wave equation. However, if the quasi-static magnetic field affects the propagation of light, it must be reversed as well. On the other hand, if light passes through a device with a built-in magnetization, the reciprocity between forward and backward propagation may fail.

We say that a material has magneto-optic properties if an externally applied magnetic field or a spontaneous magnetization contributes to the permittivity of the material. For example, some garnets which are ferri-electric and completely transparent at the near infrared are to be described by the following permittivity¹⁴:

$$\epsilon_{ij} = \epsilon \delta_{ij} + i K \epsilon_{ijk} M_k . \quad (5.27)$$

As discovered by Faraday, the polarization vector of a wave rotates by an angle $\alpha = z\Theta_F$ proportional to the propagation distance z . The specific Faraday rotation¹⁵ Θ_F of specially grown garnets may be as large as 100 full revolutions per millimeter propagation, at $\lambda = 1.3 \mu\text{m}$.

By the way, (5.27) is compatible with Onsager's relation $\epsilon_{ij}(\mathbf{M}) = \epsilon_{ji}(-\mathbf{M})$. Since $\epsilon_{ij} = \epsilon_{ji}^*$ characterizes the refractive part of the permittivity, a term linear in a magnetic field must be antisymmetric and purely imaginary.

¹⁴ ϵ_{ijk} is the totally antisymmetric Levi-Civita symbol

¹⁵ $\Theta_F = k_0 K M / 2n$ where n is the refractive index

5.3.2 Waveguide isolators

The dimensionless quantity $\xi = KM$ is very small, therefore magneto-optic non-reciprocal effects are rather subtle. They rely on a small effect to be repeated rather often the consequence of which is that effects of deviations from the design parameters are multiplied as well. Therefore, fabrication tolerances are rather strict which fact has prevented a robust, reliable, temperature insensitive integrated-optical isolator so far. The subject is still in the stage of development. Optical isolators are required for protecting lasers from reflected light of their own, but also for circulators and so forth.

Non-reciprocal mode conversion For longitudinal magnetization the permittivity tensor is

$$\epsilon_{ij} = \begin{pmatrix} \epsilon & iKM & 0 \\ -iKM & \epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix}. \quad (5.28)$$

It couples the E_x and E_y components of the electric field and thereby TE and TM modes. A mode which is TE at $z = 0$ should be propagated until it is half TE and half TM. Upon reflection the conversion continues, and at $z = 0$ it is purely TM and might be absorbed¹⁶

The degree

$$R = \frac{\Theta_F^2}{\Theta_F^2 + (\Delta\beta/2)^2} \sin^2(z \sqrt{\Theta_F^2 + (\Delta\beta/2)^2}) \quad (5.29)$$

of TE/TM conversion after propagating the distance z is limited by the propagation constant mismatch $\Delta\beta = \beta_{TE} - \beta_{TM}$. All efforts to make $\Delta\beta$ vanish have failed so far.

Non-reciprocal interferometry If the magnetization is transversal, the permittivity tensor is

$$\epsilon_{ij} = \begin{pmatrix} \epsilon & 0 & iKM \\ 0 & \epsilon & 0 \\ -iKM & 0 & \epsilon \end{pmatrix}. \quad (5.30)$$

TE modes are not affected, the propagation constants of TM modes are

$$\beta_{\pm} = \bar{\beta} \pm gKM \quad (5.31)$$

where g is a geometry dependent dimensionless factor. \pm stand for forward- and backward propagation, respectively.

An integrated optical interferometer may be devised such that interference in forward direction is constructive, but destructive in backward direction.

¹⁶TM modes are wider than TE modes. Properly positioned absorbers will affect TM modes much stronger than TE modes.

Non-reciprocal couplers As we know, light is not strictly confined within the rib of a rib waveguide. If there are two adjacent waveguides, one may excite the other. This effect is known as coupling.

If at least one of the waveguides is magneto-optic, the coupling lengths¹⁷ in forward and backward direction are different. For lateral magnetization we obtain

$$L_{\pm} = \bar{L} \pm gKM/k_0, \quad (5.32)$$

where g is a geometry dependent dimensionless factor.

One may devise a coupler such that it couples an even number of time in forward and an odd number in backward direction. Such devices are the most promising candidates for a robust and working integrated optical isolator (which, in fact, is a circulator). Since there are sufficiently many geometric parameters, even a polarization independent isolator is feasible.

¹⁷Length after which the field has been transferred *in toto* from one to the other waveguide.

A Program Listings

`listing_all.m` extracts picture source code embedded within the documentation. This guarantees that code and documentation coincide (*literate programming*). The following program produces all pictures.

```

1  % this file is dwg_fig.m
2  listing_all; % extract ML and MetaPost source code
3  gi_wg;
4  print -depssc 'gi_wg.eps';
5  ! epstopdf gi_wg.eps
6  clear all;
7  slab_wg;
8  print -depssc 'slab_wg.eps';
9  ! epstopdf slab_wg.eps
10 clear all;
11 rib_wg;
12 mesh(mode1);
13 print -depssc 'mode1.eps';
14 ! epstopdf mode1.eps
15 mesh(mode2);
16 print -depssc 'mode2.eps';
17 ! epstopdf mode2.eps
18 mesh(prm);
19 view(-60,60);
20 print -depssc 'rib_wg.eps'
21 ! epstopdf rib_wg.eps
22 clear all
23 refl_bc;
24 mesh(hist);
25 print -depssc 'refl_bc.eps';
26 ! epstopdf refl_bc.eps
27 clear all;
28 transp_bc;
29 mesh(hist);
30 print -depssc 'transp_bc.eps';
31 ! epstopdf transp_bc.eps
32 clear all
33 [x,z,hist]=prp_swg(0.9);
34 contour(z,x,hist,32);
35 print -depssc 'prp_swg1.eps';
36 ! epstopdf prp_swg1.eps
37 power=sum(hist,1);
38 power=power./power(1);
39 plot(z,power);
40 print -depssc 'prp_swg2.eps';
41 ! epstopdf prp_swg2.eps
42 [x,z,hist]=prp_swg(0.0);
43 contour(z,x,hist,32);
44 print -depssc 'prp_swg3.eps';

```

```

45 ! epstopdf prp_swg3.eps
46 clear all
47 ! del *.eps
48 ! mp --tex=latex slabwg
49 ! epstopdf slabwg.1
50 ! del slabwg.1
51 ! del slabwg.log
52 ! del slabwg.mpx

```

For example, the following MATLAB code produces Figure 3.

```

1 % this file is slab_wg.m
2 lambda=0.6328; k0=2*pi/lambda;
3 d=1.8;
4 LARGE=100;
5 flm_n=1.52; flm_eps=flm_n^2;
6 sub_n=1.49; sub_eps=sub_n^2;
7 cov_n=1.00; cov_eps=cov_n^2;
8 h=0.0001;
9 eff_n=(sub_n+h:h:flm_n-h);
10 eff_eps=eff_n.^ 2;
11 flm_k_te=sqrt(flm_eps-eff_eps);
12 flm_k_tm=flm_k_te/flm_eps;
13 sub_k_te=sqrt(eff_eps-sub_eps);
14 sub_k_tm=sub_k_te/sub_eps;
15 cov_k_te=sqrt(eff_eps-cov_eps);
16 cov_k_tm=cov_k_te/cov_eps;
17 cot_fd=cot(k0*flm_k_te*d);
18 rhs_te=(flm_k_te.^2-cov_k_te.*sub_k_te)...
19 ./flm_k_te./(cov_k_te+sub_k_te);
20 rhs_tm=(flm_k_tm.^2-cov_k_tm.*sub_k_tm)...
21 ./flm_k_tm./(cov_k_tm+sub_k_tm);
22 b1=(cot_fd<LARGE).*(eff_n<1.509);
23 c1=cot_fd.*b1+LARGE*~b1;
24 b2=(cot_fd>-LARGE).*(eff_n>1.510);
25 c2=cot_fd.*b2-LARGE*~b2;
26 plot(eff_n,c1,eff_n,c2,eff_n,rhs_te,eff_n,rhs_tm);
27 axis([sub_n,flm_n,-4,4]);

```